

Modeling Forced-Choice Associative Recognition Through a Hybrid of Global Recognition and Cued-Recall

Peter A. Nobel and David E. Huber

Department of Psychology
Indiana University
Bloomington, IN 47405
pnobel@ucs.indiana.edu
dhuber@ucs.indiana.edu

Abstract

Global recognition models usually assume recognition is based on a single number, generally interpreted as 'familiarity'. Clark, Hori, and Callan (in press), tested the adequacy of such models for associative recognition, a paradigm in which subjects study pairs and must distinguish them from the same words rearranged into other pairs. Subjects chose a target pair from a set of three choices. In one condition all three choices contained a common, shared word (OLAP); in the other condition, all words were unique (NOLAP). Subjects performed slightly better in the NOLAP condition, but global recognition models predict an OLAP advantage, due to the correlation among test pairs. Clark et al. (in press) suggested that the subjects may have used cued-recall to supplement their familiarity judgments: the greater number of unique words in the NOLAP case provides extra retrieval chances that can boost performance. We tested this possibility by implementing a retrieval structure that leads to a hybrid of cued-recall and recognition. We did this for several current memory models, including connectionist and neural net models. For all of the models we explored, the observed NOLAP advantage was difficult to impossible to produce. While some researchers propose that there is a cued-recall component to associative recognition, our modeling shows that this component cannot be realized easily in the extant memory models as they are currently formulated.

This research was supported by grant NIMH 12717 to Richard M. Shiffrin.

Introduction

The question of the relation between recognition and recall has always been prominent in the study of memory (e.g. Tulving & Watkins, 1973), with most of the debate focusing on the number and nature of retrieval processes involved. For instance, in an associative recognition task the subject is presented with a list of pairs of words; at test the subject has to discriminate between intact pairs that were on the list and rearranged pairs. This task could be accomplished by a hybrid of cued-recall and global recognition (Clark, Hori, & Callan, in press): 1) The task can be accomplished by obtaining from memory a non-specific 'degree of match' or 'feeling of familiarity', based on a sum across all stored pairs, which we term 'global recognition'; this process does not require retrieval of a specific association; or 2) The task can be accomplished by recall of specific associations, used either to accept the target pair or reject the distractor pairs, which we term 'cued-recall'. Extant memory models, however, almost universally adopt only the global recognition component, not only for single item recognition, but also for associative recognition, the subject of this article.

Evidence for a recall component in associative recognition tasks comes from a variety of studies. For example, word frequency is a variable that shows a dissociation between recognition and recall tasks; recognition performance is better for low frequency (LF) words than for high frequency (HF) words, while recall performance is better for HF words than for LF words. Clark and Shiffrin (in press) showed that associative recognition

performance is better for HF words than for LF words, which is consistent with the recall findings. In a study of the time course of item and associative information Gronlund and Ratcliff (1989) found that when subjects had to discriminate intact from rearranged pairs, decisions that required associative information were about 220 ms slower than when the decision could be based on item information alone. One of their explanations of these data involved a recall process operating in conjunction with a global matching mechanism, where each of the mechanisms use different cues to gain access to item and associative information.

More recently Clark et al. (in press) explored the relationship between recognition and recall by using forced-choice associative recognition. Subjects are presented with a list of pairs of words during study and are tested under two conditions, called OLAP and NOLAP. In the OLAP condition, targets and distractors have overlapping items, while in the NOLAP condition there are no overlapping items between targets and distractors. Figure 1 shows how the two test conditions are constructed from the presented pairs of words (denoted by AB, CD, EF, etc.). During an OLAP test the subject could be asked to discriminate the target AB from distractors AD and AF, while for a NOLAP test the target could be AB, with CF and GJ as distractors.

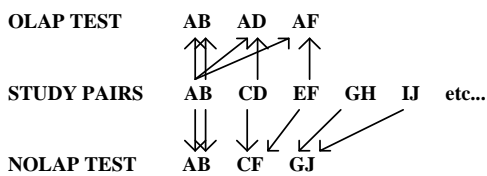


Figure 1. Test conditions OLAP and NOLAP. Adopted from Clark et al. (in press).

Clark et al. (in press) used this procedure in three different experiments. In Experiment 1 half the subjects received OLAP test trials and the other half NOLAP test trials (between subjects design). The performance measure used was the proportion of hits in each condition, i.e. the number of times the subject correctly identifies the intact pair divided by the total number of test trials. Results showed a large NOLAP advantage of about 12 percent,

but this could be due to differences in storage or retrieval strategies for the two conditions. Experiment 2 was designed to remedy this problem: OLAP and NOLAP trials were mixed together within each test sequence, therefore eliminating the possibility of different study or test strategies. Using the mixed trials, the NOLAP advantage was slight or missing. To see whether the change from Experiment 1 was due to study or test strategies, Experiment 3 used a single study list, to be followed by either an OLAP or NOLAP test list, but subjects did not know which until after study. Results now showed a large NOLAP advantage, which suggested a response strategy accounted for some of the differences between Experiments 1 and 2.

These experimental data are not in agreement with the predictions of the extant memory models. Clark et al. (in press) simulated the OLAP/NOLAP conditions with various global recognition models, including SAM (Gillund & Shiffrin, 1984), Minerva 2 (Hintzman, 1984, 1988), and TODAM (Murdoch, 1982). For these models, for both OLAP and NOLAP, the mean 'familiarity' of intact pairs is greater than the mean familiarity of rearranged pairs. However, the familiarity values are not independent for the OLAP condition: the AB, AD and AF pairs all share the word A, and the familiarities are therefore correlated, whereas the AB, CF, and GJ pairs in the NOLAP condition are independent. This correlation results in a smaller variance of the difference distribution for OLAP pairs, and therefore performance is predicted to be better for the OLAP condition (Clark et al., in press). One exception is TODAM, for which both item information and associative information are stored in the memory vector; when the weights on the item information are set to zero (no information stored), TODAM predicts no difference in performance between OLAP and NOLAP (but cannot predict a NOLAP advantage).

Clark (1992) suggested that associative recognition requires a retrieval process similar to that of cued-recall operating along with the global matching process. We tested how such a combined model does in predicting the results from the OLAP/NOLAP paradigm, within the frameworks of the following models: SAM,

Minerva 2, TODAM, and McClelland and Rumelhart's Auto-associator (McClelland and Rumelhart, 1986).

Theoretical Mechanisms

How does cued-recall help produce a NOLAP advantage? In the NOLAP condition there are extra unique cues; there are 6 unique words in the NOLAP condition, whereas in the OLAP condition there are only 4 unique words. Suppose the subject uses each unique word to try to recall a studied pair. If a retrieved pair is found that matches one of the three test pairs, that choice is made. If a retrieved pair matches only one item from a test pair, that pair is eliminated. If at the end of recall, more than one pair remains viable, global recognition is invoked. When there are more unique items presented in the test pairings, as in NOLAP, there is a greater probability of obtaining useful information (i.e., information allowing elimination of distractor pairs). For example, when tested with distractors (CF, GJ) a retrieval to any of these four items would allow elimination of the pair in question. When tested with distractors (AD, AF) only retrievals to D and F allow elimination of those test pairs (retrieval to cue A or B would allow correct performance for both OLAP and NOLAP).

In reality things are slightly more complicated than this for two reasons: 1) There is a possibility that incorrect retrievals occur; 2) Successive retrieval attempts with the same cue (i.e., the overlapping item, A) might slightly increase the chances of retrieval. Actually, neither of these complications changes the basic argument. To demonstrate this, we carried out the following simulations for each of the four different memory models:

1. "Step" through the set of three test pairs in random sequence
2. Apply cued-recall to each item of the pair and compare the returned item (if any) to the existing paired item.
3. If the pair is labeled as a *mismatch*, it is eliminated as a possible answer.
4. If the pair is labeled as a *match*, it is chosen as the answer and the process stops.
5. If the results of cued-recall are mixed (*match* from one cue and *mismatch* from

the other) or if they were not "strong" enough to be conclusive, the pair is labeled as *unknown*.

6. If two pairs have been labeled as *mismatches*, the third is chosen as the answer through *elimination*.
7. If the last pair is reached and the answer has not been reached through *match* or *elimination*, the standard global recognition measure for that model is used to choose between the pairs labeled as *unknown*.

Results and Discussion

The method discussed in the last section was implemented within SAM, Minerva 2, TODAM, and the Auto-associator (the nature of information storage, mechanisms of recognition and cued-recall, and relevant parameters for each of these models are briefly discussed in the Appendix). An individual simulation of the OLAP/NOLAP paradigm involved presentation of a study list of 34 pairs, and testing of either 6 OLAP or 6 NOLAP trials. Total percent correct was calculated as well as the breakdown for the cued-recall and recognition contributions. Every data point reported has been averaged over 1000 such simulations.

A fairly extensive parameter search within Minerva 2, TODAM, and the auto-associator was unable to produce the observed NOLAP advantage. The reason for the failure can be shown through a careful analysis of Figure 2 in which performance is shown as a function of the recall criterion. The cued-recall process in these three models produces a noisy signal which is then compared to items in semantic memory through a dot product. If the dot product does not exceed a minimum criterion, cued-recall fails to return an item. In the hybrid model, when both cues of a test pair fail to produce a recall, global recognition is used to make a decision. Therefore the recall criterion can be used to factor in recognition or cued-recall to various extents.

For the case of a low recall criterion (the cued-recall region of Figure 2), nearly every test pair is labeled as either a *match* or *mismatch*. Since it is very unlikely that cued-recall should accidentally produce the test pair, rearranged pairs are labeled *mismatches* regardless of the

increased accuracy that extra cueing in NOLAP trials provides. In the cued-recall region accurate performance only occurs through the correct labeling of the intact pair as a *match*. Since the intact pair is the same for OLAP and NOLAP trials, there is no difference in performance between them. Alternately, if the recall criterion is set very high, cued-recall produces nothing and the test pairs can only be differentiated through a recognition comparison (the recognition region of Figure 2). In the recognition region the correlation between test pairs in an OLAP trial leads to higher OLAP performance.

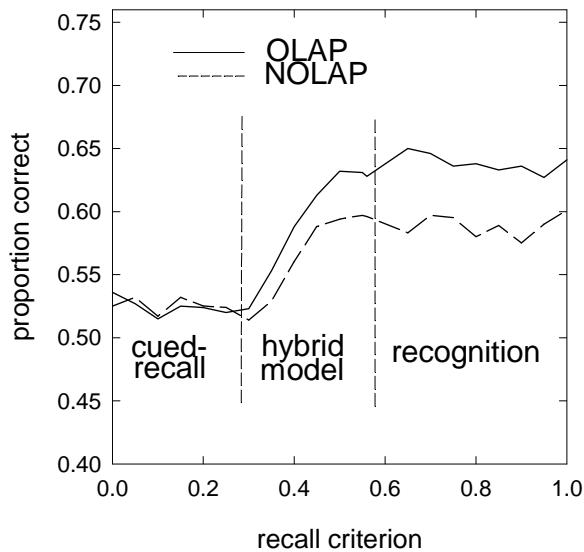


Figure 2. Proportion correct is shown for OLAP and NOLAP test trials as a function of the recall criterion for the Minerva 2 ($l = .7$).

It is only with the recall criterion set to some mid-level that the hybrid model can come to fruition. As the recall criterion is raised from the cued-recall region, more trials are solved through performing recognition on two or three of the test pairs. Due to the extra cueing of the NOLAP trials, the onset of this change occurs sooner for the OLAP trials. Since recognition is a more effective way of differentiating test pairs than cued-recall, this leads the OLAP curve to rise above the NOLAP curve. If the situation were reversed and cued-recall alone was more effective than recognition alone, the curves would be downward sloping in the hybrid region. Then extra cueing would allow the

NOLAP curve to maintain its high level longer than the OLAP curve. This would lead to the desired NOLAP advantage. For Minerva 2, TODAM, and the auto-associator, the parameters equally affect cued-recall and recognition which means that there is no possibility to create a situation in which cued-recall is more effective than recognition.

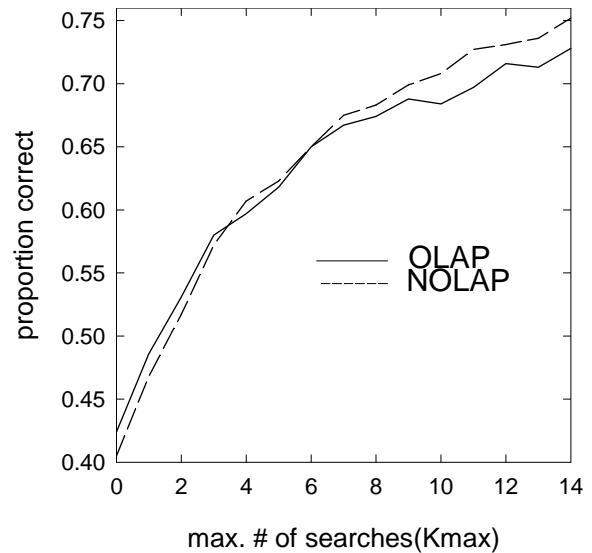


Figure 3. Proportion correct is shown for OLAP and NOLAP test trials as a function of the maximum number of searches for SAM. $a = .1, b = .5, c = .2, d = .1$.

It is known that for the SAM model cued-recall performance and recognition performance can be separated out through the strength parameters (Gillund & Shiffrin, 1984). Lowering the self strength parameter (c) will improve cued-recall, but harm recognition performance. Raising the associative strength parameter (b) will improve both cued-recall and recognition performance. By setting the associative strength significantly higher than the self strength, SAM can operate in a region where cued-recall is more effective than recognition. In light of the psychological relevance of these parameters this might not be desirable, but it does produce the correct pattern of results.

SAM does not include a recall criterion for scaling between cued-recall and recognition, but the maximum number of searches (K_{max}) performs a similar function. With K_{max} equal to zero, test pairs are only differentiated through recognition and therefore SAM produces an

OLAP advantage (see Figure 3). As the number of searches increases, the more effective cued-recall comes into play and both curves rise. Now the extra cueing involved in a NOLAP trial provides an advantage over recognition and NOLAP performance becomes clearly better than OLAP performance.

Conclusions

Extant memory models assume associative recognition is based on a global measure of familiarity for a given set of test probes. The experimental data, however, suggest that a cued-recall process might be involved. The OLAP/NOLAP paradigm developed by Clark et al. (in press) provides a good example of such data. We have simulated four models in which a cued-recall process is used in addition. In these simulations the cues are optimally used, either for recalling the correct pair, or for eliminating the incorrect pairs. With the exception of specific parameter settings within SAM, this hybrid retrieval structure could not produce the experimentally observed NOLAP advantage. The difficulty lies in the inability of the models to disassociate cued-recall and recognition performance for the same set of items. The higher level of performance using recognition alone means that cued-recall is not a more effective strategy than global recognition. Experimental evidence suggests that a cued-recall like process is involved in associative recognition, but our modeling demonstrates that the present formulation of cued-recall within the extant memory models is inadequate.

References

- Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory and Cognition*, 20, 231-243.
- Clark, S. E., Hori, A., & Callan, D. E. (in press). Forced-choice associative recognition: implications for global memory models.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recall and recognition. *Psychological Review*, 91, 1-67.
- 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, 846-858.
- Hintzman, D. L. (1984). MINERVA 2: A simulation of human memory. *Behavioral Research Methods, Instrumentation, and Computers*, 26, 96-101.
- Hintzman, D. L. (1988). Judgments of frequency and recognition in a multiple-trace memory model. *Psychological Review*, 95, 528-551.
- McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland & D. E. Rumelhart (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, pp. 170-215). Cambridge, MA: MIT Press.
- Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609-626.
- Tulving, E., & Watkins, M. J. (1973). Continuity between recall and recognition. *American Journal of Psychology*, 86, 739-748.

Appendix

SAM

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. 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. Memory is probed with two or more cues: the cue provided by context as well as the item(s) being tested. The familiarity of the probe is defined as the activation caused by the probe cues, which is the sum of the activations of all the memory images. As for all the models, the answer in a forced-choice associative recognition test is the pair with the highest familiarity. Recall is carried out by a two stage process: 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 must be recovered for a response to

take place. This sampling followed by attempted recovery continues over and over until a response is found or the subject gives up.

The following parameters are used in SAM: a , the context to item strength; b , the strength of association of items that were rehearsed together; c , the self-strength of an item to its own image; d , the residual strength between items that are not rehearsed together; K_{max} , the maximum number of sampling operations for a particular probe.

Minerva 2

Memory traces within the Minerva 2 framework consist of vectors of features. Each vector is stored separately and the possible feature values are 1, -1, or 0. Associative information is stored by placing both studied items in the same double length vector. The encoding of items is probabilistic; there is a probability that each feature is encoded with the correct value, otherwise that feature is encoded with a zero. In both the recognition and cued-recall process, an activation value is computed for each memory trace by taking the cube of the dot product between the cues and the stored features. For recognition, these activations are summed up to provide a familiarity measure. In recall, "echo" vectors are summed instead of activations. Each echo is found by multiplying the activation of a trace across each feature value of the trace. The resultant vector from this summation is then compared to a list of possible items (semantic memory) and the item with the highest dot product is produced. We have implemented a recall criterion such that no item is produced if the highest correlation does not exceed a minimum threshold.

The following parameters are used in Minerva 2: l , the probability of encoding a feature; c_r , the criterion for recall.

TODAM

In TODAM, items are represented as real-valued feature vectors. When the item pair AB is presented, the vectors A and B, as well as the convolution, $A*B$, are added to a single composite memory vector which contains all episodic information. Recognition is carried out by taking the dot product of the probe vector (in the case of associative recognition, the convolution of the items) and the memory vector; this results in a measure of familiarity.

In recall, the probe vector is correlated with the memory vector. The resulting noisy vector is compared to semantic memory (usually consisting of all the list items), and the item with the highest dot product above a criterion is chosen as the answer.

The following parameters are used in TODAM: α , the forgetting parameter of the memory vector; γ_i , the weight for item information; γ_a , the weight for associative information; c_r , the criterion for recall.

Auto-associator

The Auto-associator is a highly composite/distributed model in which memory is represented by real-valued connection weights between real-valued feature nodes. Items consist of vectors of 1 and -1 valued features. Whenever an item is presented to the system, the activation at each node cycles up to asymptote through an activation difference equation that reflects the amount of external activation (the item itself) as well as the internal activation (the sum of the activations coming over the weighted connections from the other nodes). In the learning of a study list, connection weights are changed by an amount equal to the multiplication of the activation of the "sender" node and the error at the "receiver" node (error is the difference between the external and internal activations). This connection change is weighted by a learning parameter. A familiarity measure for recognition is found through the dot product between a presented item and the internal activations at asymptote. In cued-recall, the cue is presented to the system and the remaining external activations are set to zero. After activations have reached asymptote, the internal activations of the missing pair are compared to a list of possible items in the same manner as in Minerva 2 and TODAM. Likewise there is a threshold criterion.

The following parameters are used in the Auto-associator: l , the learning parameter; e , i , and d , weights for the external, internal, and decay terms of the activation difference equation; c_r , the criterion for recall.