Journal of Experimental Psychology:
Learning, Memory, and Cognition
1991, Vol. 17, No. 2, 277–287

Copyright 1991 by the American Psychological Association, Inc.
0278-7393/91/$3.00

Integration of Orthographic and Semantic Information in Memory Retrieval

Dominic W. Massaro, Mary Susan Weldon, and Stephen N. Kitzis
University of California, Santa Cruz

Our central goal was to determine how multiple sources of information are evaluated and integrated during memory retrieval. An expanded factorial design was used to vary two sources of information independently of one another and to present each source alone. Subjects solved crossword-like puzzles with varying numbers of orthographic and semantic cues. The results of the experiment indicated that (a) performance is better given two sources of information relative to just one; (b) evaluation of each source of information provides continuous rather than just categorical (all-or-none) information; and (c) the two sources are integrated multiplicatively rather than simply used independently of one another as claimed by nonintegration models. A fuzzy logical model of perception (FLMP)—taken from the pattern recognition domain—gave a good description of the memory results. A single channel model, an averaging model, and an adding model produced poor descriptions of the results.

The present research is carried out according to the procedures of falsification and strong inference. We follow these guidelines in our experimental tests of alternative hypotheses about memory retrieval. Most studies in experimental psychology represent primarily one of two metatheoretical approaches. The psychophysical approach aims to discover laws relating the physical world to observable behavior. Its value is apparent in the immense data base that has been acquired on the psychophysics of our sensory experience. The information-processing approach aims to discover how the stimulus world is processed to generate some observable behavior. The value of this approach is apparent in the progress that has been made in understanding the component processes involved in pattern recognition, memory, and decision making. The present research combines the goals of both psychophysical and information-processing methods.

We ask how multiple sources of information are evaluated and integrated in memory retrieval. Consider the retrieval of the word *performance*, given the semantic cue *act* and the letters *per----ance*. The retrieval of the critical word is achieved by integrating the bottom-up and the top-down sources of information—that is, the letters and semantic cue. This task is analogous to solving a crossword puzzle, which involves exploiting a variety of semantic and orthographic constraints. Our analysis builds on the assumption that memory retrieval is a form of pattern recognition. Evaluation and integration processes have been shown to be fundamental to all forms of pattern recognition (Massaro, 1987a, 1987b), and we expect to find analogous processes in memory retrieval.

Multiple retrieval cues generally produce better retrieval than do single cues (although there are some exceptions, such as part-list cuing and retrieval inhibition effects, e.g., Blaxton & Neely, 1985; Roediger & Neely, 1982; Slamecka, 1969). How are multiple sources of information utilized during retrieval? Additional cues might constrain the set of plausible candidates that are searched, might increase the amount of information available to retrieve the target, or both. Retrieval cues can be compounded in two ways. First, additional information along the same attribute dimension can be added to the cue. For example, the target *grape* can be cued with an orthographic cue such as a word fragment (*g----*), and this cue can be strengthened by adding more letters (*gr---*; Tulving & Watkins, 1973). Thus, the additional letters provide additional information along the same dimension as the original orthographic cue, which may strengthen this cue. Likewise, multiple rhyme (e.g., *bench*, *wrench*, for *stretch*) or semantic (e.g., *hawk*, *eagle* for *vulture*) cues can be provided to improve retrieval (e.g., Nelson, McEvoy, & Friedrich, 1982). A second way to construct multiple cues is to combine cues from different dimensions, such as combining a rhyme and semantic cue (e.g., "A mythical being that rhymes with post" for *ghost*, Rubin & Wallace, 1989), or an orthographic and semantic cue (e.g., *plane* and *-ar-ch-t-* for *parachute*; Weldon, Roediger, & Challis, 1989).

Debates have arisen regarding how multiple sources of information combine to improve retrieval and whether their combined effects can be predicted from knowledge of their individual strengths (Bruce, 1980; Jones, 1976; Rubin & Wallace, 1989). One long-standing question concerns whether multiple cues independently access memory, whether the cue information is combined prior to the search and acts in

The research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314), the National Science Foundation (BNS 8812728), a James McKeen Cattell Fellowship to DWM, and grants to the three authors from the graduate division and the Biomedical Sciences Award Committee of the University of California, Santa Cruz.

The authors would like to thank Antoinette Gesi, Art Glenberg, Doug Hintzman, and Gregg Oden for their comments on the paper and Steve Clark, Ben Murdock, Doug Hintzman, Mike Humphreys, and Jeroen Raaijmakers for speculating on how their models might handle the results of the experiment.

Correspondence concerning this article should be addressed to Dominic W. Massaro, Program in Experimental Psychology, University of California, Santa Cruz, California 95064.

concert, or whether some hybrid process occurs. In addition, extremely weak individual cues can produce nearly perfect retrieval when they are combined (McLeod, Williams, & Broadbent, 1971; Watkins & Tulving, 1978), stronger individual cues may produce less than perfect retrieval when combined (Rubin & Wallace, 1989), or an additional cue may have no effect (Tulving & Osler, 1968) or even a deleterious effect (Nelson et al., 1982; Roediger & Neely, 1982). It has recently been argued that no single model adequately describes all dual-cue situations (Rubin & Wallace, 1989).

To examine the effects of dual cues, Rubin and Wallace (1989) devised test items with a rhyme cue, a meaning cue, or both. The rhyme cue might be *deal*, the meaning cue *a building material*, and the target *steel*. Rubin and Wallace argue against all composite models—models that predict dual-cue performance solely as a function of performance given each of the single cues. They do this by showing that the facilitation that is obtained, given two cues, cannot be predicted from the absolute level of performance, given the single-cue conditions. Thus, the rhyme *post* and the meaning *a mythical being* produce the correct target *ghost* with probability .16 and .01, respectively, whereas their combination gives perfect (1.0) performance. There are many words that rhyme with *post*, and many words can fit the description of *a mythical being*, but only *ghost* fits both. In contrast, the rhyme *clog* and the meaning *a four-footed animal* produce the correct target *dog*, with probability .38 and .56, whereas their combination gives only .77 correct. The dual cue does not completely constrain the answer because *hog* remains a valid response in this case, whereas there is only a single valid answer in the first case. Thus, the degree of facilitation with two cues relative to the single-cue conditions is a function of both the degree to which the correct alternative is supported, as well as the degree to which incorrect alternatives are supported.

Our purpose in the research reported here is twofold. First, we intend to provide an enriched data base for modeling the effects of single and multiple retrieval cues. As mentioned above, research has been conducted to investigate either the effects of adding cues along one attribute dimension or of combining cues from two different dimensions, but little or no work has been done to examine the effects of both of these. In order to obtain a better picture of how multiple retrieval cues operate, we used a semantic memory test. Subjects were told they would perform a crossword puzzle task in which they would receive semantic cues related to the meaning of the target word, as well as a fragment of the target word itself, and they were to try to generate the target. On each trial, subjects received 0, 1, 2, 3, or 4 semantic cues and an eight-letter word fragment with 0, 2, 3, 4, or 5 letters in place. Following Massaro and Cohen (1990), these two variables were varied in an expanded factorial design in which the two sources of retrieval information were varied independently of one another, and each of the sources was also presented alone. This arrangement resulted in 24 conditions (the 0,0 condition was excluded). Thus, subjects received retrieval information from two different attribute dimensions (semantic and orthographic), and the amount of information provided along each dimension was varied independently. Note that this task can be classified as a semantic or indirect memory task because subjects were not instructed to refer to a specific prior study episode to perform the task (Richardson-Klavehn & Bjork, 1988; Rubin & Wallace, 1989; Tulving, 1983), but simply to think of the appropriate target word.

The second purpose of this research is to examine models of multiple-cued retrieval in order to make inferences about the manner in which multiple sources of information are combined. Four alternative models will be fit to the new data base: the fuzzy logical model of perception (FLMP; Massaro, 1987a, 1987b, 1989), the adding model (ADM), the single-channel model (SCM), and the weighted averaging model (WAM). To test among the models, it is necessary to determine how the sources of information are evaluated and integrated in memory retrieval. To accomplish this, we use methods developed in information-integration theory (Anderson, 1981, 1982) and mathematical model testing (Townsend, 1984).

We use an expanded factorial design that independently manipulates multiple aspects of the environment jointly and in isolation. The two sources of information for retrieval are varied independently of one another. Each of the sources is also presented alone. The orthographic and semantic sources of information are represented by uppercase letters O and S, respectively. The value O_i would correspond to the ith level of the O source and S_j would correspond to the jth level of the S source. A given stimulus composed of a single source would be labeled O_i or S_j, and a given combination would be represented by O_iS_j. The models will be developed in accordance with the general information processing model illustrated in Figure 1.

Figure 1 illustrates three operations involved in memory retrieval. The evaluation process transforms each source of information (orthographic and semantic in our task) into feature values (indicated by lowercase letters). The outcome of feature evaluation gives the degree to which each source of information supports each item in memory. The integration process combines the feature values to give an overall goodness-of-match between all of the available sources of information and each item in memory. The decision operation maps these values resulting from integration into some response, such as recall of a word or a rating of its familiarity.

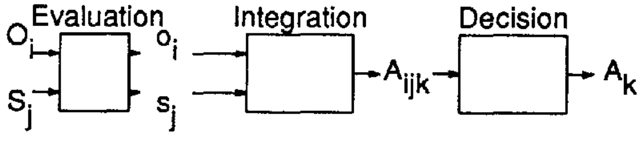

Figure 1. Schematic representation of the three operations involved in recall memory. (The evaluation of an orthographic source of information O_i produces a value o_{ik}, indicating the degree of support for alternative A_k. Evaluation of a semantic source occurs in an analogous way. The integration of the truth values gives an overall goodness-of-match A_{ijk}. The recall given O_iS_j is based on the relative goodness-of-match of the response alternatives. The probability of a response A_k to event O_iS_j is equal to the value A_{ijk} relative to the goodness of match of all response alternatives.)

Fuzzy Logical Model of Perception

One model that accurately describes pattern recognition is a fuzzy logical model of perception (abbreviated FLMP). Its name is derived partly from fuzzy logic—a continuously valued logic that represents the truth of propositions in terms of truth values that range between zero (false) to one (true) (Goguen, 1969; Zadeh, 1965). The model has received support in several domains, including letter and word recognition in reading (Massaro, 1979; Oden, 1984), syllable and word recognition in speech perception (Massaro, 1989; Oden & Massaro, 1978), perceived size (Massaro, 1988), and the categorization of visual objects (Oden, 1981; Thompson & Massaro, 1989). The model assumes three operations in perceptual recognition: feature evaluation, feature integration, and decision. Continuously valued features are evaluated, integrated, and matched against prototype descriptions in memory; then an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions. An attractive feature of the FLMP is that it generates many quantitative experimental predictions. The FLMP can be tested not only against a broad range of results but also against other quantitative models of performance (Massaro, 1987b, 1989; Thompson & Massaro, 1989). The sources of information are evaluated independently of one another, and the outcome of evaluation provides continuous rather than just categorical information. In addition, the multiple sources of information are integrated multiplicatively in such a way that two sources of information can be more informative than just one.

The FLMP describes the three operations between a test presentation of a pattern and some response illustrated in Figure 1. Feature evaluation gives the degree to which a given source of information supports each test alternative. For a given response alternative A_k, O_i would be transformed to o_{ik}, and analogously for dimension S_j. Feature integration consists of a multiplicative combination of feature values supporting each alternative. If o_{ik} and s_{jk} are the values supporting alternative A_k, then the total support for the alternative A_k would be given by the product $o_{ik}s_{jk}$.

The third operation is decision, which gives the relative degree of support for each of the test alternatives. In this case, the probability of an A_k response given O_iS_j is

$$P(A_k \mid O_iS_j) = \frac{O_{ik}\ S_{jk}}{\Sigma}, \qquad (1)$$

where Σ is equal to the sum of the merit of all relevant alternatives, derived in the same manner as illustrated for alternative A_k. Following Massaro and Friedman (1990), this decision operation is called a *relative goodness rule* (RGR).

Applying the model to retrieval is straightforward because we view retrieval as pattern recognition, given several sources of information. In terms of the three processes illustrated in Figure 1, feature evaluation gives the support of each source for each word in memory, integration gives an overall degree of support for each word, and decision consists of retrieving a specific word on the basis of the degree of support for that word relative to all word candidates in memory.

Extension of FLMP to Memory Retrieval

The FLMP has most successfully been applied to situations in which one of a fixed number of responses is made on every trial. The dependent measure is the proportion of each relevant response. In this situation, the relative goodness rule (the decision process of the FLMP) makes sense in terms of a response being equal to the relative strength of that alternative with respect to all alternatives in the task. The present task requires an extension of the typical application of the FLMP to the situation in which the number of response alternatives is not fixed and subjects do not have to respond on each trial. Thus, the dependent measure is the percentage of items correctly retrieved.

The neutral truth value in fuzzy logic is .5 when there are two alternatives. In the FLMP, therefore, any support for being correct in a two-alternative task must be greater than .5. Any value greater than .5 is positive support; any value less than .5 is negative support. Negative support means that the source of information actually supports an incorrect response—either a wrong alternative or a failure to respond—more than the correct response alternative. Following this logic, a forced-choice task with a response on every trial must be described differently from an open-ended task with some failures to respond. When no information is present, the two response alternatives in the forced-choice task are supported to degree .5. There is no inherent advantage of one response alternative over the other. In the open-ended task, however, no information must support the correct response less than the incorrect response (which includes trials in which no response is made). Thus, support for a correct response must be less than .5 when no information is present. To implement this constraint, we assume a background degree of support that is less than .5 for a correct response and greater than .5 for an incorrect response. (This assumption is not unreasonable because there was also an additional source of background information in experiments with responses on each trial. However, the background information was neutral [.5] and, thus, could be ignored in the analysis.) The presentation of a positive source of information would support the correct alternative to some degree greater than .5. The integration of multiple sources of support would follow the same multiplicative integration algorithm of the FLMP. Similarly, decision would follow a relative goodness rule in the same manner as in situations with known response alternatives.

In the framework of the FLMP, there would now be three sources of information supporting correct and incorrect answers: background information, orthographic information from the letter cues, and semantic context from the semantic cues. These sources of support for the correct answer are represented by b, o_i, and s_j. The overall degree of support for a correct answer, g(correct), is equal to

$$g(\text{correct}) = b \times o_i \times s_j\,. \qquad (2)$$

The value b is less than .5, whereas o_i and s_j are greater than .5. To predict performance, it is also necessary to determine the degree of support for an incorrect response. Given that we do not know how much the three sources of information—

background, letter cues, and semantic cues—support the incorrect response, additional free parameters appear to be necessary. The overall degree of support for an incorrect answer, g(incorrect), is equal to

$$g(\text{incorrect}) = b_1 \times o_{1i} \times s_{1j} \quad (3)$$

Where b_1 corresponds to the background support for an incorrect response, and o_{1i} and s_{1j} correspond to the support for an incorrect response given by the letter cues and semantic cues.

Given the decision operation, the overall likelihood of a correct response, $P(C)$, would be equal to

$$P(C) = \frac{bo_is_j}{bo_is_j + b_1o_{1i}s_{1j}}. \quad (4)$$

Given Equation 4, it appears the parameter values for the degrees of support for the incorrect response are independent of the degrees of support for the correct response. However, in actual practice the free parameters representing the support for the second alternative in a two-alternative task can be set equal to the additive complements of the parameters for the other response without any loss of predictive power (Massaro, 1989, p. 788). Thus, we assume that $b_1 = 1 - b$, $o_{1i} = 1 - o_i$, and $s_{1j} = 1 - s_j$, and

$$P(C) = \frac{bo_is_j}{bo_is_j + (1-b)(1-o_i)(1-s_j)}. \quad (5)$$

In order to fit the model to the present task (with four levels of orthographic information and four levels of semantic information), four values of o_i, four values of s_j, and one value of b must be estimated as free parameters. We now develop alternative models of performance, based on different assumptions about evaluation and integration.

Adding Model (ADM)

An adding model is analogous to the FLMP in many respects, except that the cues are added rather than multiplied. Therefore, the probability of a correct response is given directly by the sum of the support of the background, orthographic, and semantic cues:

$$P(C) = b + o_i + s_j. \quad (6)$$

If the prediction of Equation 6 is greater than 1, the predicted probability is simply 1. As in the FLMP, one value of b, four values of o_i, and four values of s_j must be estimated as free parameters.

Single-Channel Model (SCM)

This model can be viewed as a psychological representation of a normative independence (nonintegration) model. Fragment theory is one example of this type of model (Jones, 1987; Rubin & Wallace, 1989). Subjects attempt to retrieve the word on the basis of letters and to retrieve the word on the basis of semantic cues. Separate decisions are assumed to be made, given the orthographic and semantic sources or channels. Thus, there are two chances to retrieve the word in the dual-cue condition relative to the single-cue condition. If subjects are correct by retrieving the word given either source, the percentage correct, given both sources, $P(C|O_iS_j)$, would be

$$P(C|O_iS_j) = P(C|O_i) + P(C|S_j) - P(C|O_i)P(C|S_j), \quad (7)$$

where $P(C|O)$ and $P(C|S)$ are the percentages of correct retrieval, given just the orthographic cue and just the semantic cue, respectively. The single-channel model predicts a performance improvement, given both sources relative to just one. Given 20% correct for orthographic and 30% correct for semantic information separately, Equation 7 predicts that the factorial performance based on independent recalls of the test word based on the letters and semantic cues would be 44% correct:

$$P(C|O_iS_j) = .20 + .30 - (.20 \times .30) = .44.$$

This model requires eight free parameters to predict the 24 data points, one for each level of o_i and s_j.

Although multiple sources of information are available, the assumption of the SCM is that they are not integrated. The sources of information are evaluated independently of one another, as in the FLMP. However, the outcome of evaluation provides just categorical information because a given source either produces a correct response or does not. The nature of the integration algorithm is not relevant for this model because integration does not occur.

Weighted Averaging Model

Like the FLMP, the weighted averaging model (WAM) assumes the integration of the multiple sources of information (Anderson, 1981). The sources are evaluated independently, and the outcome of evaluation is continuous rather than just categorical. The only difference between the WAM and the FLMP is the nature of the integration algorithm—assumed to be compromising for the WAM and enhancing for the FLMP. Given continuous and independent evidence from the orthographic and semantic cues, the perceiver is assumed simply to average the two sources of evidence. The amount of support for the correct alternative can be assumed to be a weighted average of the support, given the orthographic source and the semantic source. If o_i and s_j are the values supporting the correct alternative, then the total support for the correct alternative would be given by a weighted averaging of o_i and s_j:

$$g(\text{correct}) = wo_i + (1-w)s_j, \quad (8)$$

where w is the weight given the orthographic source and $(1 - w)$ is the weight given the semantic source (Massaro, 1987b, p. 182). The total support for the incorrect alternative would be given by a weighted averaging of $(1 - o_i)$ and $(1 - s_j)$:

$$g(\text{incorrect}) = w(1-o_i) + (1-w)(1-s_j). \quad (9)$$

The decision operation is based on the support for the correct alternative relative to the sum of support for correct and incorrect alternatives. Given that this sum is equal to 1, the probability of a correct response given O_iS_j is simply equal to g(correct),

$$P(C\,|\,O_iS_j) = wo_i + (1 - w)s_j. \qquad (10)$$

The weighted averaging model requires nine free parameters—four values of o_i, four values of s_j, and a weight w—when applied to the results of the present experiment.

We now describe the experiment and the fits of the four models to the results.

Method

Subjects

Subjects were 24 University of California, Santa Cruz, undergraduates who participated for credit in an introductory psychology course. All were native English speakers and had normal or corrected vision.

Stimuli

Target items were 72 eight-letter words taken from Gibson and Watkins (1987). These words were uniquely determined by two-letter word fragments (e.g., *FLEXIBLE* is the only solution for ---X--L-). These two letters formed the first nonzero level of orthographic cues; higher levels included additional letters. Specific levels were cumulative because higher levels always included the lower. For example, successive levels for *FLEXIBLE* were ---X--L-,--XI-L,--EXI-L-, and --EXI-LE.

Semantic cues were words selected from a thesaurus with nonsynonymous associations with the target item. They also had minimal orthographic similarity to the target: The words *flexing* or *stretchable* would be too close to *flexible* to be used as cues, particularly if presented together. As was the case for orthographic cues, specific semantic cue levels were cumulative. Thus, successive levels for *FLEXIBLE* were *STRETCH*, *STRETCH* plus *MANIPULATE*, both of those plus *VERSATILE*, and then all three plus *RESILIENT*.

Design

The basic design was a 5 (orthographic cues) × 5 (semantic cues) expanded factorial. Both factors were manipulated within subjects. Orthographic cue levels were 0, 2, 3, 4, or 5 letters presented; semantic levels were 0, 1, 2, 3, or 4 words. There were only 24 possible retrieval conditions for each item, rather than 25, because the combination of zero cues for both factors was not presented. Data analysis was based on the assumption of a 5 × 5 complete factorial, with results for the missing cell set to zero. This is justified on the grounds that even if the missing condition had been presented to subjects, a correct response to no cues at all would have been an extremely fortuitous event (i.e., subjects guessing with no information at all as to what the word might have been).

Each subject was tested with the same set of 72 target items, 3 items in each of the 24 retrieval conditions. Items were completely counterbalanced across conditions and subjects, each rotated through all of the retrieval conditions. Item presentation order was randomized for each subject.

Apparatus

Orthographic and semantic cues were presented visually on IBM personal computers, with a 40 character per line black/white screen mode. Orthographic cues consisted of uppercase letters, with underline characters indicating missing letters (e.g., EL_P__NT for *elephant*). Semantic cues appeared one above the other, directly above the orthographic cues. Cues were presented in the center of the screen, with instructions at the bottom. Subjects sat at an unconstrained, comfortable viewing distance from the screen and typed their responses on the keyboard. Subjects' answers were saved on disk, as were the time from cue presentation to initial response and the time from initial response to response completion, both to an accuracy of 1 ms.

Procedure

Subjects were asked to determine words when given some combination of letter and word clues for each. Instructions stressed that all words would be eight letters long, the letters given would be in the correct positions within the word, but that word clues could be suggestive of the target word without necessarily being synonyms. For example, *white* or *oval* could be clues to the target word *eggshell*. Subjects were told that the number of clues would vary from word to word and that sometimes letter clues might be given without word clues, or word clues without letters. The task was described as being most similar to working a crossword puzzle in varying states of completion. However, subjects had only 30 s to initiate a response for a word, and then up to 20 s to type their answer. Subjects were tested individually at separate work stations. Each completed a practice session of 12 words, identical for all subjects, with a chance to ask questions before proceeding with the actual test session. The entire procedure lasted about 1 hr, including a short rest break midway through the test.

For the data analysis, subjects' responses were scored as being either fully correct or incorrect with respect to each target item. Partial hits were not allowed, but in some cases an incorrect spelling was scored as being a correct response if it was obvious which word was misspelled. For example, the incorrect spelling of *flexable* would be accepted as a correct response for the word *flexible*.

Results

Accuracy

Figure 2 shows the average proportion of times that subjects were able to determine correctly the target item for each of the 24 different retrieval conditions. Results were averaged together for all subjects and all items for each condition. Both panels show the same data, but plotted with either the number of orthographic cues or the number of semantic cues on the abscissa. The proportion of correct responses generally increased with the number of cues presented, for either type of cue, with the effect due to increasing the number of letters being stronger than that due to increasing the number of semantic cues. The stronger effect of orthographic cuing is shown best by the steeper slopes when the data are plotted against the orthographic cue level than when plotted against the semantic level.

An analysis of variance was performed on the proportion of correct responses for each of the retrieval conditions—once with subjects as the source of variance, three items per condition for the 24 subjects, and once with items as the source, 1 subject per condition for the 72 items. In both cases, the analysis assumed a 5 × 5 complete factorial design, with the

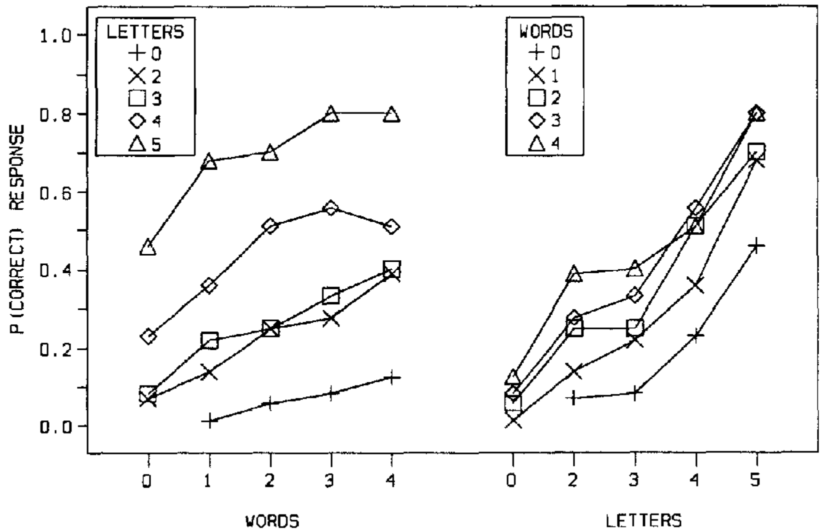

Figure 2. Observed probability of correct recall as a function of the number of letters in the orthographic cue and the number of words in the semantic cue. (The left panel gives the results with the orthographic cue as the curve parameter, whereas the right panel gives the results with the semantic cue as the curve parameter.)

results of the double-zero combination cell set to zero. (Setting this cell to a value of zero with zero variance was not unreasonable, given that the mean and variance of the condition with no letters and one semantic cue were .014 and .005, respectively.) With subjects as the variance source, the analysis indicated that the main effect of orthographic cue level was significant, $F(4, 92) = 114.0$, $MS_e = .063$, $p < .001$, the main effect of semantic cue level was significant, $F(4, 92) = 28.3$, $MS_e = .051$, $p < .001$, but that the Orthographic × Semantic Cue Level interaction was not significant ($MS_e = .065$, $p > .3$). With items as the variance source, the analysis indicated that the main effect of orthographic cue level was significant, $F(4, 284) = 112.3$, $MS_e = .193$, $p < .001$, the main effect of semantic cue level was significant, $F(4, 284) = 24.8$, $MS_e = .176$, $p < .001$, and the Orthographic × Semantic Cue Level interaction was marginally significant, $F(16, 1136) = 1.62$, $MS_e = .131$, $p = .057$. Thus, the main effects are essentially the same with either subjects or items as the source of variance.

Response Times

Table 1 contains the mean time from cue presentation to the initial key typing response, and the mean time from initial response to response completion (i.e, typing time), for each level of both types of cue. Averages were calculated only for correct responses. The initial response time decreases with increasing orthographic cue levels, but it is relatively flat with increasing semantic cue levels. Typing time is essentially constant for all levels of either cue type. Given the unequal number of correct responses in each retrieval condition and the resulting mix of within- and between-subject sampling, an analysis of variance was not performed on these data. However, Table 1 contains the standard error for each condition, which would be roughly equivalent to half the standard criterion used for statistical significance. Differences between the initial response times for orthographic levels 0, 4, and 5 are greater than three or four standard errors, implying that the decrease in initial response time with increasing orthographic cue levels is unlikely to be due to chance.

Frequency distributions of initial response times were examined to determine whether the 30-s response limit was sufficient for most subjects. Histograms of the times for correct responses were produced with 5-s intervals from 0 to 30 s. In general, distributions were unimodal, with the most frequent time almost always occurring in the 5–10-s interval, for all levels of either cue type. Even though potential responses were obviously being lost by the 30-s limitation, the location of the distribution peaks indicated that 30 s was sufficient to capture most of the responses.

The initial response time distributions were also examined to determine whether there were systemic shifts in the distribution peaks with respect to different retrieval conditions. The only conditions for which the peak did not occur between 5 and 10 s were the two extreme levels of orthographic cuing. For the condition of no letters presented, the peak occurred between 10 and 15 s; for five letters presented, between 0 and 5 s. These shifts parallel the response time decrease with increasing orthographic cue levels described earlier and are further evidence that it is unlikely that this decrease is due only to chance.

Model Tests

The average percentage of correct responses in each experimental condition was used to test the different models. The predictions of the models were fit to these results by using the parameter-estimation program STEPIT (Chandler, 1969). A

Table 1
Response Time Means (Ms) and Standard Errors (SE) (in Seconds)

Cue level	No. correct responses	Initial response time *M*	Initial response time *SE*	Typing time *M*	Typing time *SE*
Orthographic					
0	16	15.09	1.33	6.38	0.56
2	83	11.76	0.71	5.57	0.23
3	92	12.30	0.76	6.28	0.37
4	156	10.50	0.52	6.52	0.27
5	250	8.22	0.38	5.88	0.18
Semantic					
0	61	11.25	0.96	6.32	0.43
1	102	9.99	0.64	6.31	0.31
2	126	10.04	0.58	5.97	0.23
3	146	10.13	0.57	6.02	0.25
4	162	9.84	0.49	5.98	0.25

model was defined in STEPIT as prediction equations with a set of unknown parameters. STEPIT minimized the deviations between the observed and predicted values of the models by iteratively adjusting the parameters of the equations. Root mean square deviation (RMSD) values index the overall goodness of fit of the model. This value is the square root of the average squared deviation between the predicted and observed values. RMSD values are used because these specify directly the correspondence between a model and data or the correspondence between the predictions of two models. The smaller the RMSD value, the better the fit of the model.

The FLMP provided a good description of the results, with an RMSD of .0283. The parameter values are given in Table 2. The predicted values are shown as lines in Figure 3.

Figure 4 gives the fit of the adding model (ADM). The fit to the ADM was about a factor of two poorer than the fit of the FLMP. The RMSD for the ADM was .0545.

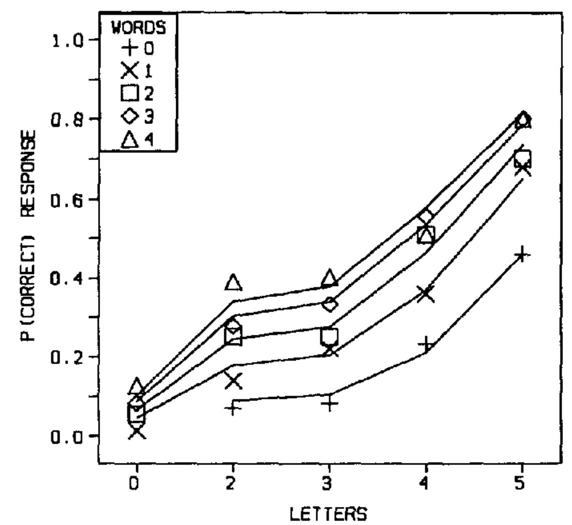

Figure 3. Probability of correct recall as a function of the number of letters in the orthographic cue and the number of words in the semantic cue. (Points give the observed results, and lines give the predictions of the fuzzy logical model of perception [FLMP].)

As can be seen in Figure 5, the SCM gives a poor description of the results, with an RMSD of .0707. The parameter values are given in Table 2. The SCM underpredicts the accuracy of performance, given two sources of information relative to just one. Another way to observe this fact is to use Equation 7 to predict performance directly in the factorial conditions from performance observed in the single-factor conditions. In every case, observed performance in the factorial conditions was better than that predicted by Equation 7, given the single-factor conditions.

One might argue that the FLMP provided a better fit than did the SCM because the latter model had one less parameter. To test for this possibility, a background source of information was also assumed for the SCM—as it was for the FLMP—and estimated as the ninth parameter of the model. In this case, the subject has some probability of being correct, given just the background—providing a third potential channel of information. This model did not improve the fit of the SCM and shows that the advantage of the FLMP is not due to simply having an extra free parameter.

The weighted averaging model also gave a poor description of the results, with an RMSD of .1064. Figure 6 gives the predictions, and Table 2 gives the parameter values. Analogous to the evaluation of the SCM, we can ask how well Equation 10 predicts the factorial conditions, given the single factor conditions. This test reveals that the averaging model—even though an integration model—underpredicts the factorial conditions even more than does the SCM.

Discussion

The results are consistent with the hypothesis that memory retrieval is supported by multiple sources of information. Both orthographic and semantic sources are evaluated and integrated, and this results in a continuous degree of support for the test word. Enhancing integration is apparent in the impressive accuracy, given two sources of information relative to just one. In all cases, performance was much better with two sources than that predicted by a single-channel or a weighted averaging model.

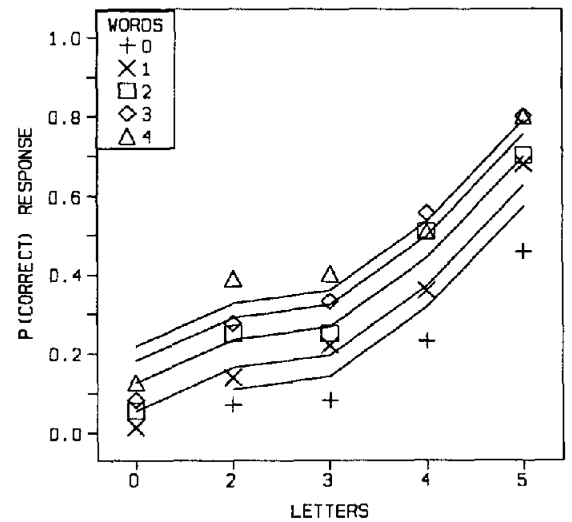

Figure 4. Probability of correct recall as a function of the number of letters in the orthographic cue and the number of words in the semantic cue. (Points give the observed results, and lines give the predictions of the adding model [ADM].)

The use of multiple cues in retrieving a target item can pose an important constraint on several classes of cued-recall memory models. If in a particular model multiple cues are not integrated, but each acts independently, which is the basic assumption of the single-channel model (SCM), then multiple cues will increase the probability of successful retrieval over that of each acting alone, but only to the extent that the joint probability of retrieval is equal to the union of the individual cue retrieval probabilities, as given in Equation 7. In contrast, if multiple cues are integrated according to a compromising integration algorithm given by the WAM, then the joint probability of retrieval will be a weighted average of the probabilities of each cue acting separately, as given by Equation 10. Finally, when multiple cues are integrated according to the enhancing integration algorithm of the FLMP, then the joint probability can exceed the union, or even the sum, of the individual single-cue probabilities, as given by Equation 5.

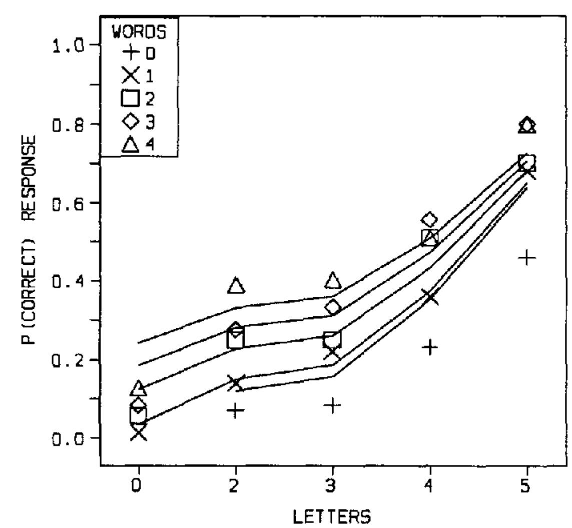

Figure 5. Probability of correct recall as a function of the number of letters in the orthographic cue and the number of words in the semantic cue. (Points give the observed results, and lines give the predictions of the single channel model [SCM].)

Our experiment acquired results for both single-cue and dual-cue conditions. There were 16 joint-cue conditions, in which both orthographic and semantic cues were presented, and 8 single-cue conditions, four of each type. In every case, the dual-cue results were greater than the sum of the two corresponding single-cue conditions. The empirical results of this experiment support only enhancing integration models similar to the FLMP and are in conflict with models based upon the assumptions of the SCM or WAM.

It is informative to ask how various contemporary retrieval models account for the effects of multiple cues. We attempt to compare these results against a number of specific memory models by first mapping them onto the more general SCM, WAM, and enhancing integration models. All of the models under consideration here store and utilize associations between cues and the target information to be retrieved from memory. A common characteristic of many of these models is the explicit assumption that memory is organized as vectors of featural elements, such as in Flexser and Tulving's (1978) episodic memory model, Hintzman's (1986) Minerva II, or Humphreys, Bain, and Pike's (1989) matrix model. Other models are less concerned with how items are stored, however, and specify more details of the recall and recognition processes, such as Gillund and Shiffrin's (1984) search of associative memory model (SAM) or Ratcliff's (1978) diffusion model. With the sole exception of SAM, none of these models is specifically oriented toward integrating different sources of external information in the retrieval process, though a number of them treat the original encoding context as a second source of information. Given the complexity of these models and the fact that a model often can serve as a general framework for a series of more specific variations, our mapping must be treated as a tentative extension of each model's basic assumptions toward handling multiple retrieval cues. In addition, it should be noted that all of these models are concerned primarily with episodic retrieval mechanisms rather than semantic retrieval as required by our experiment.

Flexser and Tulving's (1978) episodic memory model was developed to deal with the phenomenon of recognition failure of recallable words in paired-associate learning. In their model, episodic traces are stored as vectors of features, and the probability of a cue retrieving an item from memory is based on the featural overlap between the cue and memory trace. Given probabilistic featural encoding of the original trace and separate recall and recognition cues, recall and recognition results can be largely independent of each other. The simplest extension of the model to handle multiple retrieval cues in semantic memory would be to assume that each cue operates independently in attempting a probabilistic match against the target item, which, of course, is the basic assumption of the SCM. Another model based on similar assumptions is Jones (1976) fragment theory and is also

Table 2
Parameter Values for the Model Fits

Model	No. of parameters	RMSD	Orthographic source O_2	O_3	O_4	O_5	Semantic source S_1	S_2	S_3	S_4	B	W
FLMP	9	.0283	.8186	.8424	.9241	.9749	.6873	.7658	.8141	.8389	.0214	—
SCM	8	.0707	.1193	.1566	.3532	.6388	.0345	.1238	.1848	.2415	—	—
WAM	9	.1064	.2254	.2580	.4338	.6888	.0130	.0570	.0830	.1271	—	.9999
ADM	9	.0545	.1096	.1422	.3180	.5730	.0542	.1254	.1816	.2176	.0001	—

Note. The parameter *B* refers to the background information, and *W* is the weight. FLMP = fuzzy logical model of perception; SCM = single-channel model; WAM = weighted averaging model; ADM = adding model.

functionally equivalent to the SCM. To account for the enhancing integration results of our experiment, these models would have to be modified beyond this simplest extension to allow some sort of nonlinear interaction between the retrieval cues and the stored traces or to allow the generation of a compound retrieval cue on the basis of an enhancing integration of the multiple cues prior to memory access.

Ratcliff's (1978) diffusion model does not specify how items are stored in memory but is more concerned with specifying details of the retrieval process itself, particularly with regard to the quantitative modeling of latency of retrieval. The model assumes that memory access occurs by means of a large number of individual processes running in parallel, one per item in memory, in response to a memory probe. Each process independently proceeds as a random walk on the basis of similarity between the probe and the memory item represented by that process. Matching features drive an item toward a match decision, whereas nonmatching features drive it toward a nonmatch decision. Processes reaching a nonmatch decision terminate and have no further effect, but the first item to reach its match point becomes the item to be retrieved

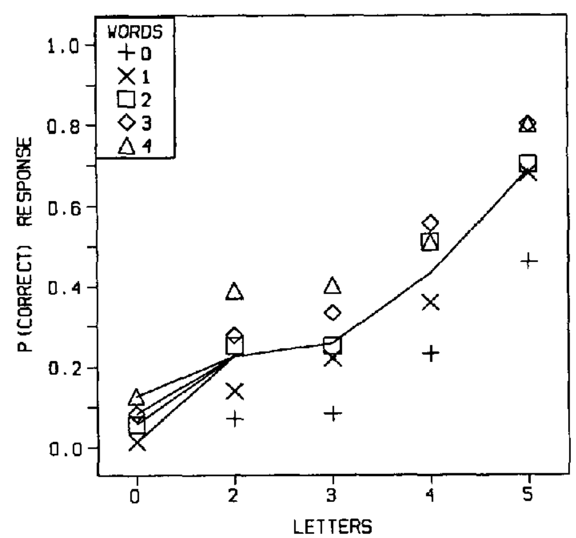

Figure 6. Probability of correct recall as a function of the number of letters in the orthographic cue and the number of words in the semantic cue. (Points give the observed results, and lines give the predictions of the weighted averaging model [WAM].)

from memory in response to the probe. Assuming that letter and semantic cues operate as memory probes, the simplest extension of the model to handle multiple-cue retrieval would be to assume that each cue operates independently in driving the parallel decision processes (i.e., parallel inputs driving parallel decision processes), which is equivalent to the SCM. Once again, to account for the enhancing integration results of our experiment, the model would have to be modified beyond this simplest extension to allow for a more complicated matching function, either some sort of nonlinear interaction between the retrieval cues and the stored traces or the generation of a compound retrieval cue on the basis of an enhancing integration of the multiple cues prior to memory access.

In Humphreys et al.'s (1989) matrix model, memories of episodes and episodic associations are summed together into a common distributed associative memory. Specific episodes are separable because of the explicit storage of episodic context. All items and context are stored as arrays of feature strengths. The decision process is described only as a process that operates on the output of context and cue matches against the associative memory to produce a scalar or single vector output. The matrix model has provisions for a number of different types of retrieval, including recognition, cued recall, free recall, and word fragment completion. In a cued but context-free recall task, such as employed in our experiment, the response to a memory cue would be a single featural vector consisting of a weighted average of all items previously associated with the cue. Given the linear nature of adding items into the common distributed memory, item weights in the blended output vector would be proportional to the number of past item–cue associations. The simplest extension of the model to handle multiple-cue retrievals would be to produce an output vector for each cue, all of which then would be added together. This would be the equivalent of averaging all associations with any of the cues into a single common output vector. Even though the decision process is not well defined by Humphreys et al. (1989), blending vectors of different noise levels will not improve the quality of the less noisy vector, which appears to put the matrix model in the same class as the WAM.

Other linear summation distributed memory models, such as Murdock's (1982, 1983) TODAM or Knapp and Anderson's (1984) distributed memory model, appear to operate similarly and make the same predictions given by the WAM. There are two methods by which this class of models could be modified to account for the enhancing integration results of our experiment. The first is to generate a compound retrieval cue on the basis of an enhancing integration of the multiple cues prior to memory access. The second is to replace context as a retrieval cue in the models with the second external retrieval cue. This would allow nonlinear interactions to occur between the two retrieval cues and episodes stored in memory.

Hintzman's (1986) Minerva II model stores individual memory traces as separate vectors of features, but its output in response to a retrieval cue is a blended vector very similar to that produced by the Humphreys et al. (1989) matrix model. Each memory trace is activated in proportion to the cube of its similarity to the cue, and the resulting output is a weighted average of the activated traces, weights being equal to the amount of activation. Given that the activation of a trace is a nonlinear function, the model could potentially account for an enhancing integration of the retrieval cues, though a detailed simulation would be necessary to determine whether the unmodified model could account for the results presented here.

Of the set of specific models considered here, only Gillund and Shiffrin's (1984) SAM model had been designed with the intent to handle the integration of multiple retrieval cues, and, not coincidentally, it is the only model to use explicitly what we are calling an enhancing integration algorithm. In SAM, recall is a sequential search through individual items stored in memory, where the search is guided by the match between multiple cues and each item stored. The total match strength for an item is the product of individual cue-to-item strengths, rather than a sum, and this nonlinearity is explicitly introduced into the retrieval process. The probability of an item's being selected next for possible recall is the ratio of its match strength to the sum of all items' match strengths. The multiplication of strengths and the ratio of match strengths are mathematically equivalent to the integration and decision operations of the FLMP. Once an item has been selected for possible retrieval, the cues operate independently in eliciting recall of the item. If recall is successful, then the sequential search is terminated; otherwise, the selected item remains in the pool of items, and the search continues with the selection of another item on the basis of all the cue-to-item match strengths. In one sense, the SAM model is a hybrid in that the actual retrieval process is equivalent to the SCM, but the process of selecting the next item for possible recall is equivalent to an enhancing integration algorithm. However, because item selection initiates and guides the retrieval process, the total effect is that of an enhancing integration model, similar in many ways to the FLMP.

In conclusion, all of the memory models could probably be extended to capture the results of the present experiment. In the case of the Flexser and Tulving (1978) or Ratcliff (1978) models, the extension would probably take the form of combining the multiple cues into a compound cue prior to the access process. The extension would be more natural in the Humphreys et al. (1989) or Murdock (1982, 1983) models because context would be replaced by the second external retrieval cue. The Hintzman (1986) and Gillund and Shiffrin (1984) models could probably account for an enhancing integration unchanged. The important conclusion here is not whether any specific model can or can not be modified but that only models allowing for a nonlinear enhancing integration algorithm, such as used in the FLMP, would not be in conflict with the empirical results of this experiment.

Rubin and Wallace (1989) claim that no single composite model can account for the effects of dual cues because dual-cue strength is not a monotonic function of individual cue strengths. That is, extremely weak individual cues can produce perfect retrieval when they are combined, whereas stronger individual cues may produce less than perfect retrieval when combined. Rubin and Wallace did not vary their cues along a continuum as we did in the present experiment. In the Rubin and Wallace study, the rhyme *post* and the meaning *a mythical being* produce the correct target *ghost*, with a probability of .16 and .01, respectively, whereas their combination gives perfect (1.0) performance. In a second condition, the rhyme *clog* and the meaning *a four-footed animal* produce the correct target *dog*, with a probability of .38 and .56, whereas their combination gives .77 correct. The single cues in the first condition give poorer performance than those in the second, whereas the two cues presented in combination give better performance in the first condition than in the second. This nonmonotonic result was taken as evidence against models that describe performance given two cues as a function of their individual cue strengths. However, the FLMP would describe the item differences by assuming different levels of background information for each problem—a fact that reflects different degrees of support for incorrect alternatives. We observed similar nonmonotonic results between different items in the present study, but the FLMP was able to describe the average results.

Although the FLMP can predict item differences, we do not take this predictive capability as evidence for the model because the model has to assume as much as it predicts. The evidence for the FLMP comes from the results in which more independent observations were predicted than free parameters. We varied the strength of the individual cues by increasing the number of letters in the orthographic cue and the number of associates in the semantic cue. We did not have a case in which a combination of stronger individual cues resulted in a poorer retrieval than did a combination of weaker cues. If this result had occurred, it would have falsified the FLMP. That is, accuracy of performance, given two cues, was a direct function of accuracy of performance, given each of the separate cues. To the extent performance was good, given each of the separate cues, it was good, given both cues simultaneously.

References

Anderson, N. H. (1981). *Foundations of information integration theory.* San Diego, CA: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory.* San Diego, CA: Academic Press.

Blaxton, T. A., & Neely, J. H. (1985). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory & Cognition, 11,* 500–510.

Bruce, D. (1980). Single probes, double probes, and the structure of memory traces. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 276–292.

Chandler, J. P. (1969). Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science, 14,* 81–82.

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review, 85,* 153–171.

Gibson, J. M., & Watkins, M. J. (1987). A pool of 1086 words with unique two letter fragments. *Behavioral Research Methods, Instruments, & Computers, 20,* 390–397.

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

Goguen, J. A. (1969). The logic of inexact concepts. *Synthese, 19,* 325–373.

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

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233.

Jones, G. V. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. *Journal of Experimental Psychology: General, 105,* 277–293.

Jones, G. V. (1987). Independence and exclusivity among psychological processes: Implications for the structure of recall. *Psychological Review, 94,* 229–235.

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 616–637.

Massaro, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 595–609.

Massaro, D. W. (1987a). Integrating multiple sources of information in listening and reading. In D. A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Shared mechanisms in listening, speaking, reading, and writing* (pp. 111–129). San Diego, CA: Academic Press.

Massaro, D. W. (1987b). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1988). Ambiguity in perception and experimentation. *Journal of Experimental Psychology: General, 117,* 417–421.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology, 21,* 398–421.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science, 1,* 55–63.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review, 97,* 225–252.

McLeod, P. D., Williams, C. E., & Broadbent, D. E. (1971). Free recall with assistance from one and from two retrieval cues. *British Journal of Psychology, 62,* 59–65.

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.

Nelson, D. L., McEvoy, C. L., & Friedrich, M. A. (1982). Extralist cuing and retrieval inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 89–105.

Oden, G. C. (1981). A fuzzy propositional model of concept structure and use: A case study in object identification. In G. W. Lasker (Ed.), *Applied systems and cybernetics: Vol. VI* (pp. 2890–2897). Elmsford, NY: Pergamon Press.

Oden, G. C. (1984). Dependence, independence, and the emergence of word features. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 394–405.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172–191.

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

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475–543.

Roediger, H. L., & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology, 36,* 213–242.

Rubin, D. C., & Wallace, W. T. (1989). Rhyme and reason: Analyses of dual retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 698–709.

Slamecka, N. J. (1969). Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology, 81,* 557–560.

Thompson, L. A., & Massaro, D. W. (1989). Before you see it, you see its parts: Evidence for feature encoding and integration in preschool children and adults. *Cognitive Psychology, 21,* 334–362.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments, *Journal of Mathematical Psychology, 28,* 363–400.

Tulving, E. (1983). *Elements of episodic memory.* Hillsdale, NJ: Erlbaum.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77,* 593–601.

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

Watkins, M. J., & Tulving, E. (1978). When retrieval cueing fails. *British Journal of Psychology, 69,* 443–450.

Weldon, M. S., Roediger, H. L., & Challis, B. H. (1989). The properties of retrieval cues constrain the picture superiority effect. *Memory & Cognition, 17,* 95–105.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8,* 338–353.

Received November 21, 1989
Revision received July 19, 1990
Accepted July 24, 1990 ■