# Integration of Orthographic, Conceptual, and Episodic Information on Implicit and Explicit Tests

MARY SUSAN WELDON and DOMINIC W. MASSARO
*University of California, Santa Cruz*

**Abstract** An experiment was conducted to determine how orthographic and conceptual information are integrated during incidental and intentional retrieval. Subjects studied word lists with either a shallow (counting vowels) or deep (rating pleasantness) processing task, then received either an implicit or explicit word fragment completion (WFC) test. At test, word fragments contained 0, 1, 2, or 4 letters, and were accompanied by 0, 1, 2, or 3 semantically related words. On both the implicit and explicit tests, performance improved with increases in the numbers of letters and words. When semantic cues were presented with the word fragments, the implicit test became more conceptually driven. Still, conceptual processing had a larger effect in intentional than in incidental retrieval. The Fuzzy Logical Model of Perception (FLMP) provided a good description of how orthographic, semantic, and episodic information were combined during retrieval.

The distinction between implicit and explicit memory tests has been an important development in memory research, enabling investigators to uncover a variety of new memory phenomena, and to develop richer models of the architecture and processes of memory. But while much effort has been devoted to demonstrating empirical differences between implicit and explicit tests, relatively less is understood about the underlying mechanisms that distinguish intentional retrieval on explicit tests from incidental retrieval on implicit tests.

Several models suggest that dissociations between implicit and explicit tests indicate the presence of different memory systems (e.g., Tulving & Schacter, 1990; Squire, 1994). These models focus on describing the architecture and organization of the various systems, and the type of information that is processed in each. Process models of memory dissociations (e.g., the transfer appropriate processing models of Graf & Ryan, 1991; Roediger, 1990; Roediger, Weldon, & Challis, 1989) tend to focus on characterizing the mechanisms that underlie retrieval on different types of tasks. Simply labeling tests as implicit or explicit does not reveal much about them, so to understand why they dissociate one needs to understand their underlying mechanisms, which is the goal the of the present research.

In this work, we used two methods to try to characterize intentional and incidental retrieval processes. First, we examined whether conceptual information plays a larger role in intentional than in incidental retrieval by comparing level-of-processing effects (Craik & Lockhart, 1972) in an implicit and explicit test. Second, we conducted a model-fitting exercise in which we used the Fuzzy Logical Model of Perception (FLMP; Massaro, 1987a,b; Massaro, Weldon, & Kitzis, 1991) to test different assumptions about how perceptual, conceptual, and episodic information are integrated during intentional and incidental retrieval.

**Perceptual and Conceptual Processing In Retrieval**

According to the transfer appropriate processing view (Roediger, 1990; Roediger et al. 1989), implicit and explicit memory tests dissociate when they engage different kinds of processes. As an example, one test may be relatively more dependent on perceptual processing (e.g., implicit WFC), whereas another may be relatively more dependent on conceptual processing (e.g., explicit free recall). A perceptual test will benefit from prior encoding that engages perceptual analysis of the target information, whereas a conceptual test benefits from meaningful encoding. Thus, dissociations are easy to produce when one encoding task provides an opportunity to perform the processing most critical to one test, but the other encoding task provides a better processing match with the other test. Historically, most implicit tests have been perceptually based and most explicit tests conceptually based, but this is not a necessary relation. It is possible to have an explicit perceptual test (e.g., graphemic cued recall; see Blaxton, 1989), and an implicit conceptual test (e.g., category production; see Srinivas & Roediger, 1990). Thus, an explicit and implicit test may behave similarly when they engage similar processes (e.g., Blaxton, 1989), or two implicit tests may dissociate when they engage different processes (e.g., Weldon & Roediger, 1987).

The fact that encoding-retrieval processing differences can produce dissociations has important implications for the investigation of intentional and incidental retrieval. To make strong statements about the differences between intentional and incidental retrieval, one must be careful to make sure that the tests differ only in retrieval instructions, and not in the types of test cues. For example, if one obtains a dissociation between an explicit conceptual test like free recall, and an implicit perceptual test like WFC, one cannot determine whether the dissociation is due to differences in the retrieval mode (intentional vs. incidental) or differences in the test cues themselves (Blaxton, 1989; Schacter, Bowers, & Booker, 1989). Therefore, to get a real sense of the effect of intentionality in retrieval, it is important to use the same test cues and vary only the retrieval instructions. Then, by strategically manipulating encoding variables and test conditions, one can examine how the processes that underlie intentional and incidental retrieval differ.

As an example, recent work has demonstrated that explicit tests engage conceptual processes to a greater extent than do implicit tests. Weldon, Roediger, Beitel, and Johnston (1995) used a word fragment completion test and a picture fragment identification test, and showed that when they were administered as implicit tests they did not exhibit either cross-form priming (i.e., words did not prime pictures, and vice versa) or repetition effects (i.e., two spaced presentations produced no more priming than one presentation), findings that are characteristic of perceptually based tests. However, when they were administered as explicit tests, both tests exhibited cross-form retrieval and repetition effects. This showed that conceptual processes were engaged during intentional retrieval, although other aspects of the data indicated that performance was also constrained by the perceptual requirements of the test tasks.

Other investigators have demonstrated that level-of-processing manipulations have little or no effect on implicit WFC and word stem completion tests, but significant effects when people are told to use the fragments or stems as cues to help them remember the studied words (Graf & Mandler, 1984; Roediger, Weldon, Stadler, & Riegler, 1992; but see Challis & Brodbeck, 1992). Specifically, with explicit retrieval instructions more fragments and stems are solved after deep or meaningful encoding than after shallow encoding.

The present work is an extension of these earlier ideas with particular interest in two issues. First, it is often claimed in the literature that one difference between implicit and explicit tests is that implicit tests are insensitive to manipulations of meaningful encoding. This erroneous claim persists despite substantial evidence to the contrary. For example, many investigators have shown that on conceptual implicit tests, priming improves with meaningful encoding (e.g., Blaxton, 1989, 1992; Hamann, 1990; Rappold & Hashtroudi, 1991; Smith & Branscombe, 1988; Srinivas & Roediger, 1990; Weldon & Coyote, in press). One of the novel aspects of our design is that in some test conditions, semantic cues are presented along with the test fragments (e.g., *imagination, original,* _ *re* _ *ti* _ _). Thus, although fragments presented alone result in perceptually-based priming, supplementing them with semantic cues should render the WFC test more conceptually driven because the semantic information can be used in conjunction with the perceptual information to complete the word fragments. According to the principle of transfer appropriate processing, if a test engages conceptual processing it should be sensitive to manipulations of encoded meaning. Therefore, even as an implicit test, WFC should show benefits of deep processing when the test fragments are accompanied by semantic cues.

Second, as noted above there is some evidence that even with all other things equal, intentional retrieval engages conceptual processing more than does incidental retrieval (Graf & Mandler, 1984; Roediger et al., 1992; Weldon et al., 1995). However, past studies have used implicit perceptual tests, and then administered them as explicit tests. Here, we can see whether the finding generalizes when the implicit test already engages both perceptual and conceptual processing. If so, there should be a larger level-of-processing effect on the explicit than implicit test.

### Integration of Multiple Sources of Information in Retrieval: The Fuzzy Logical Model of Perception

Our second major interest was to examine how multiple sources of information are integrated during retrieval. During an encoding phase, subjects saw two lists of eight-letter words (e.g., *blizzard*), and they encoded one list with a shallow processing task and the other with a deep processing task. They then took a word fragment completion (WFC) test in which they were presented either a word fragment alone (e.g., _ *l* _ _ *z* _ _ _), semantic cues alone (e.g. *white, blinding*), or a combination of both (*white, blinding,* _ *l* _ _ *z* _ _ _). Specifically, the test comprised an expanded factorial design, with word fragments containing either 0, 1, 2, or 4 letters in place, and accompanied by either 0, 1, 2, or 3 conceptually related words. Consequently, each subject tried to solve fragments in 15 different test conditions (the 0, 0 condition was not administered). Subjects received either implicit or explicit test instructions.

The word fragments provide orthographic information and the semantic cues conceptual information to guide retrieval, so when they are presented together there are multiple cues that can be combined to produce a solution. In addition, for the studied items, the encoding episode itself serves as another source of information about the solution to the fragment. The encoding instructions (level-of-processing manipulation) and test instructions (implicit vs. explicit) may also modulate performance on the test. Two interesting questions arise. First, how are multiple sources of information integrated during retrieval? And second, does the integration process differ in intentional versus incidental retrieval?

On explicit tests, multiple retrieval cues typically produce better performance than do single cues (although there are exceptions such as part-list cuing inhibition; see Basden, Basden, Church, & Beaupre, 1991; Roediger & Neely, 1982; Slamecka, 1969). The strength of retrieval cues can be increased in at least two ways. First, information can be added along a single dimension, such as adding more letters in a word fragment (e.g. _ *l* _ _ *z* _ _ _, vs. _ *l i* _ *z a* _ _; Tulving & Watkins, 1973), or adding more semantic cues (e.g. *white*, vs. *white, blinding*; Nelson, McEvoy & Friedrich, 1982). A second way to increase cue strength is to combine cues from different dimensions, such as combining orthographic and semantic information (e.g. *white* _ *l* _ _ *z* _ _ _; Massaro et al., 1991; Weldon, Roediger, & Challis, 1989).

To our knowledge, little work has been conducted to examine the effects of multiple cues on implicit tests. Weldon et al. (1989) did not obtain increased priming when strong semantic cues were added to word fragments, but performance was fairly high overall, which may have limited the priming effect. The present design enables us to examine the relation between multiple cues and priming more systematically.

Researchers have investigated 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; Massaro et al., 1991; Rubin & Wallace, 1989). One of the most fundamental questions is whether multiple cues access memory independently, whether their information is combined before retrieval, or whether a hybrid process takes place. To address this issue, Massaro et al. (1991) employed a test of long-term knowledge (i.e., there was no study phase in the experiment) in which they gave people word fragments with either 0, 2, 3, 4, or 5 letters, accompanied by either 0, 1, 2, 3, or 4 weakly related words, and measured the proportion of fragments correctly solved. They then examined how well four different models of information integration fit the data, 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). The results unambiguously favored the FLMP. One of the goals of the present work was to extend this model to see how it can account for the influences of the level of processing at study, and implicit versus explicit retrieval (also see Wenger & Payne, 1995).

The basic structure of the model is presented in Figure 1. The orthographic and semantic sources of information are represented by uppercase letters O and S, respectively. The value $O_i$ would correspond to the *i*th level of the O source, and $S_j$ would correspond to the *j*th 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 $O_iS_j$.

Figure 1 illustrates three operations assumed to be involved in 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 indicates 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.

In the FLMP, feature evaluation gives the degree to which a given source of information supports each test

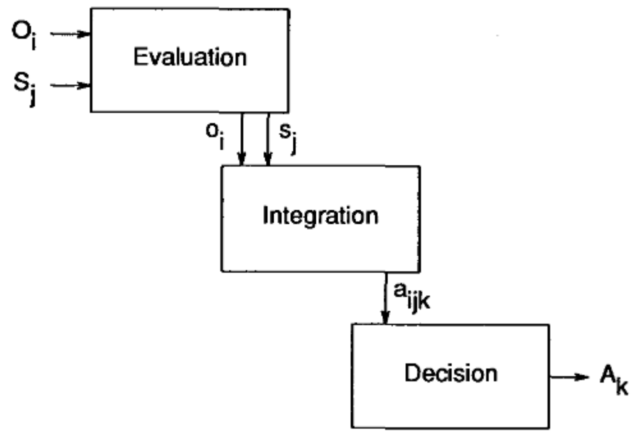


Figure 1. Schematic representation of the three operations involved in retrieval. The three operations are shown as partially overlapping in time to illustrate their necessarily successive but overlapping processing. The model is described in the text.

alternative. For a given response alternative $A_k$, $O_i$ would be transformed to $o_{ik}$, and $S_j$ to $s_{jk}$. 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 $\Sigma$ 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 retrieval is based on 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 based on the degree of support for that word relative to all word candidates in memory.

As noted by Massaro et al. (1991), the memory task requires an extension of the typical application of the FLMP to the situation in which response alternatives are not given, subjects do not have to respond on each trial, and the dependent measure is the percentage of items correctly retrieved. The memory retrieval task can be considered a task with two alternatives: correct or incorrect. Fuzzy truth values represent the degree of support for the correct and incorrect alternatives, and lie between completely true (1) and completely false (0). The neutral truth value in fuzzy logic is .5 when there are two alternatives. A given source supports both the correct and incorrect alternatives. The support for the incorrect alternative would be one minus the support for the correct alternative. 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. When no stimulus information is present, it is assumed that the background degree of support for the correct response is less than for 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 stimulus information is present, because the chance of generating the correct target with no cue is small. To implement this constraint, we assume a background degree of support, $b$, that is less than .5 for a correct response. It follows that the background support would be greater than .5 for an incorrect response. In contrast, the presentation of a positive source of information would support the correct alternative to some degree greater than .5.

In the situation with no priming episode, there are three sources of information supporting correct and incorrect answers: background information, orthographic information from the letter cues, and semantic context from the word 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

$$g(\text{correct}) = b \times o_i \times s_j \tag{2}$$

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

$$g(\text{incorrect}) = b_I \times o_{Ii} \times s_{Ij} \tag{3}$$

where $b_I$ corresponds to the background support for an incorrect response and $o_{Ii}$ and $s_{Ij}$ 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

$$P(C) = \frac{bo_i s_j}{bo_i s_j + b_I o_{Ii} s_{Ij}} \tag{4}$$

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 $b_I = 1 - b$, $o_{Ii} = 1 - o_i$, and $s_{Ij} = 1 - s_j$, and

$$P(C) = \frac{bo_i s_j}{bo_i s_j + (1-b)(1-o_i)(1-s_j)} \tag{5}$$

To fit the model to the present task (with three levels of orthographic information and three levels of semantic information), three values of $o_i$, three values of $s_j$, and one value of $b$ must be estimated as free parameters.

Two additional aspects of the present encoding and test situations need to be addressed. First, some of the test words were presented during an encoding task with either shallow or deep encoding requirements. Second, the subjects were given either explicit or implicit retrieval instructions at test. Conceptualized within the framework of the FLMP, the study task could be viewed as an additional source of information at retrieval (cf. Massaro et al., 1991, in which there was no study phase so the fragment completion test served as a test of long-term knowledge). This source of information would also be fuzzy, and would be integrated with the other available sources of information. The degree of support for the correct alternative from the study task would be modulated by the encoding instructions. For explicit retrieval, and for implicit retrieval with semantic cues, we would expect shallow encoding to provide less helpful information than deep encoding instructions. Similarly, we might expect the retrieval instructions to modulate the background level of support, as well as the level of support given by the study episode. In the Results, we explore the effect of adding parameters to account for these factors.

In sum, we were interested in evaluating how a prior encoding episode affects the way information is integrated during retrieval. In addition, we were interested in seeing whether information integration differs in intentional and incidental retrieval. We examined several different ways of adjusting the parameters in the FLMP to see which best accounts for the effects of level of processing and retrieval orientation. This is a relatively novel approach to examining the processes underlying intentional and incidental retrieval, and this work provides an opportunity to explore its potential value as an empirical and theoretical tool in this field.

## METHOD

### *Subjects and design*

Subjects were 180 undergraduates who participated for credit in lower division courses or for a cash payment. All were native English speakers with normal or corrected vision.

The basic design of the test conditions was a 4 (orthographic cues; 0, 1, 2, or 4) × 4 (semantic cues; 0, 1, 2, or 3) factorial. However, only 15 of these retrieval conditions were tested since the 0, 0 condition was not actually presented. (We assumed performance would be virtually

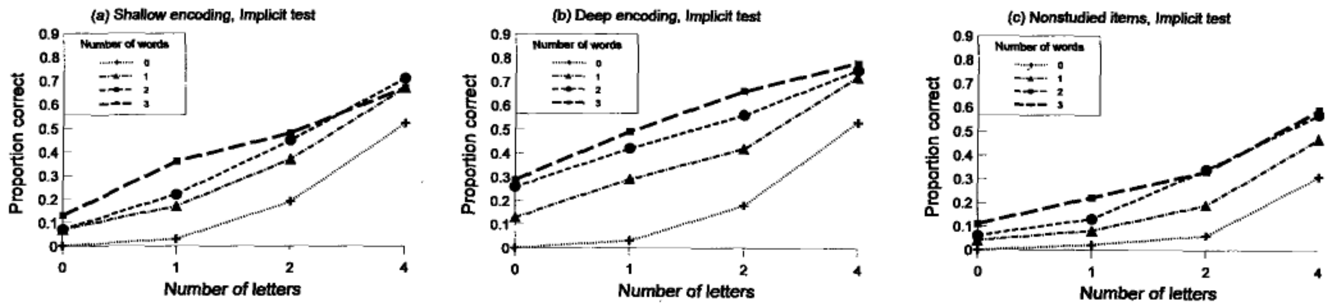


Figure 2. Performance on the implicit test. Panels (a) shallow encoding, (b) deep encoding, (c) nonstudied items.

zero in this condition.) Test instructions were manipulated between subjects with 90 receiving implicit and 90 explicit instructions. Level of processing was manipulated within subjects so that each subject studied one block of items in the deep and one in the shallow encoding condition. Items were counterbalanced so that they appeared in each encoding and test condition an equal number of times across subjects. The order of the encoding tasks was counterbalanced across subjects.

*Stimuli and apparatus*

Target items were 90 eight-letter words taken from Gibson and Watkins (1987). Four fragments were created for each target, with the orthographic information cumulating such that each fragment added new letters to the existing fragment. For example, the fragments for *blizzard* contained no letters (_ _ _ _ _ _ _ _), one letter (_ _ _ _ *z* _ _ _), two letters (_ *l* _ _ *z* _ _ _), or four letters (_ *l i* _ *z a* _ _). The semantic cues were words with nonsynonymous and relatively weak associations to the targets, and minimal orthographic similarity. Across conditions, levels of semantic information were cumulative such that subjects saw either no semantic cue, one cue (e.g., *white*), two cues (e.g., *white, blinding*), or three cues (e.g., *white, blinding, cold*). In the combined conditions, the designated numbers of semantic and letter cues were presented together (e.g., *white, blinding,* _ *li* _ *za* _ _). An additional 30 filler items were created to increase the number of nonstudied items on the test; each filler item was presented with four letters and three semantic cues.

Targets were randomly divided into 15 sets of six words, and then two words from each set were randomly assigned to the deep, shallow, and nonstudied encoding conditions. To achieve the counterbalancing, the words in each encoding condition were first rotated through all 15 test conditions, and then each set of two words was rotated into a new encoding condition. This process was repeated until each word had appeared in every encoding and test condition, creating 45 different lists. Each list was presented to two subjects in each test condition. During encoding, half the subjects received the graphemic task first and half the semantic task first. Three primacy and three recency buffers were added to each study list so that subjects saw a total of 36 items in each encoding condition.

Stimuli were presented on a CRT, with the experiment controlled by an IBM 286AT computer. During the encoding phase each word was presented centered on the screen in a black lower case font against a grey background. During the test phase the fragments were presented in the same font in the center of the screen, and the cue words were presented to the upper left of the fragment. Subjects typed their answers on the keyboard.

*Procedure*

Subjects were told they were participating in an experiment about word reading and comprehension, and were not told they would receive a memory test. Instructions for each phase were presented on the CRT. During the encoding phase subjects saw two blocks of items. For the items presented in the shallow condition subjects counted the number of vowels in each word and entered the number on the keyboard. For the deep encoding condition subjects read each word and rated its pleasantness on a scale of 1 (extremely unpleasant) to 6 (extremely pleasant).

After encoding both blocks of words, subjects received the test instructions. In the implicit test condition they were told they would see fragments with various numbers of letters and related words, and should try to complete as many as possible. When they had a solution, they were instructed to hit the space bar and then type their answer, and then the next fragment would appear. In the explicit test condition, subjects were told they were receiving a memory test for the words they saw earlier, and they were to use the fragments and word cues as clues to help them remember the words. In both conditions, subjects had up to 20 sec to solve each fragment, and if they did not provide an answer the computer advanced to the next item. Before beginning the test, subjects were given eight practice items to make sure they understood the task.

## RESULTS

The proportion of items correct in each encoding condition (deep, shallow, nonstudied) are displayed in Figure 2 (implicit test) and Figure 3 (explicit test). Priming on the implicit test (studied minus nonstudied performance) is shown in Figure 4. First, we examined the data by using analyses of variance (ANOVAs) to test our experimental hypotheses, then we evaluated how the FLMP accounts for the same set of data. For the ANOVAs, $p < .05$.

*Analyses of variance: Priming.* First, an analysis was con-

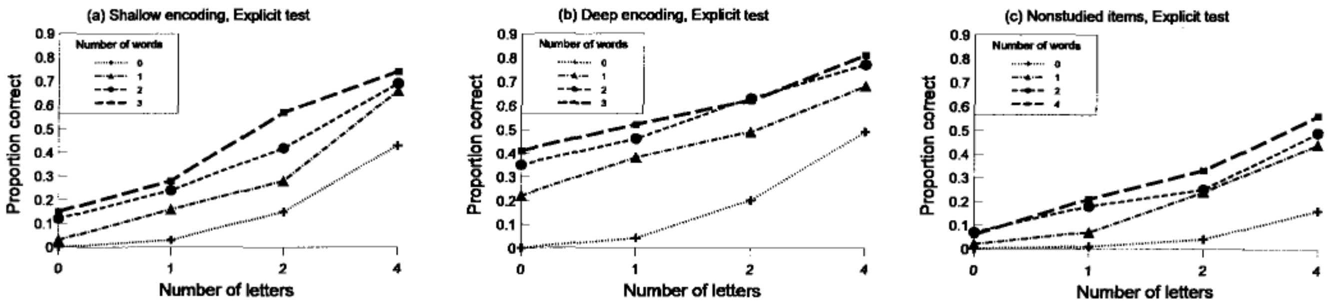


Figure 3. Performance on the explicit test. Panels (a) shallow encoding, (b) deep encoding, (c) nonstudied items.

ducted to determine whether priming was obtained in the implicit test condition. An ANOVA comparing performance in the shallow versus nonstudied conditions revealed a significant main effect of study condition, $F(1, 89) = 92.48$, $MS_e = .076$. An ANOVA comparing the deep to the nonstudied conditions also revealed a significant priming effect, $F(1, 89) = 234.14$, $MS_e = .108$. Priming scores were computed by subtracting the nonstudied baselines from total performance in the studied conditions, and are presented in Figure 4 for the shallow and deep encoding conditions. T-tests revealed that priming was obtained in nearly every condition. The exceptions were in the following shallow encoding conditions (number of words/number of letters): 0/0, 1/0, 2/0, 3/0, and 3/4 (marginally significant).

Figure 4 indicates that there is a tendency for priming to increase as cue information increases. However, this increase flattens out or even decreases as performance attains higher levels, presumably due to a ceiling effect compressing priming at higher levels of performance. An ANOVA conducted on the priming scores in the shallow condition revealed a main effect of the number of letters, $F(3, 267) = 11.59$, $MS_e = .139$, but no main effect of the number of words. This indicates that after shallow encoding, priming increased with increments in orthographic but not semantic information. Interestingly, the deep condition revealed main effects of both types of retrieval information, with a main effect of number of letters, $F(1, 267) = 5.12$, $MS_e = .159$, as well as number of words, $F(3, 267) = 10.35$, $MS_e = .157$. Thus increments in semantic information at test led to increments in priming only when the meaning of the target words had been processed during the encoding phase, a finding consistent with the transfer appropriate processing view.

*Analyses of variance: Implicit versus explicit tests.* Our primary interest was in the level-of-processing effects on the implicit versus explicit tests. However, it is often difficult to compare directly performance on implicit and explicit tests because different levels of performance may introduce scaling problems. That is, priming scores computed by subtracting the nonstudied baselines may be lower than recall scores on the explicit tests, making interactions difficult to interpret. Furthermore, it may not be appropriate to subtract nonstudied performance in the explicit test conditions because these items might mean something quite different from baselines on an implicit test. For example, they may contain many false alarms. Furthermore, in our data, subtracting the baselines would have introduced a problem with flattened or reduced priming effects at high levels of performance, whose influence in the analyses would have been arguably artifactual. It turned out that the baseline levels of performance were very similar in each test group (implicit $M = .22$; explicit $M = .20$) and an ANOVA revealed that they were not significantly different, $F(1, 178) = 2.96$, $p = .09$. Because we were interested in examining level-of-processing effects per se, and the nonstudied performance was not different in the two test conditions, we chose to exclude the nonstudied conditions from the overall ANOVA, rather than subtract them. This seemed to provide the most straightforward way to compare the level-of-processing effects on the explicit and implicit tests.

A 4 (letters) × 4 (words) × 2 (level of processing) × 2 (test instructions) ANOVA was conducted to ascertain overall differences between the implicit and explicit test conditions. The complete ANOVA is presented in Table 1, and we mention the findings of immediate interest here. Not surprisingly, there were main effects of letters (O) and words (W), indicating that performance improved as the number of letters in the fragments or number of word cues increased. Of more interest, there was a significant interaction between the number of words and letters in the test cue (W × O), suggesting that the information from each source was not combined in an additive fashion. Also, there was no significant three-way interaction between words, letters, and test type (W × O × T), suggesting that the manner in which orthographic and semantic information were combined was not appreciably different on the implicit and explicit tests.

There was a significant level-of-processing effect (L), indicating that deep processing ($M = .43$) produced better performance than shallow processing ($M = .32$). The level-of-processing effects, computed as the differences between deep and shallow processing are presented in Table 2. There was a significant interaction between level of processing and test (L × T) such that the level-of-processing effect was larger on the explicit test than on the implicit test, as was predicted.

We also predicted that on the implicit WFC test, when semantic cues were added to the word fragments, the test would become more conceptually driven. In Table 2 one can see that on the implicit test, there was

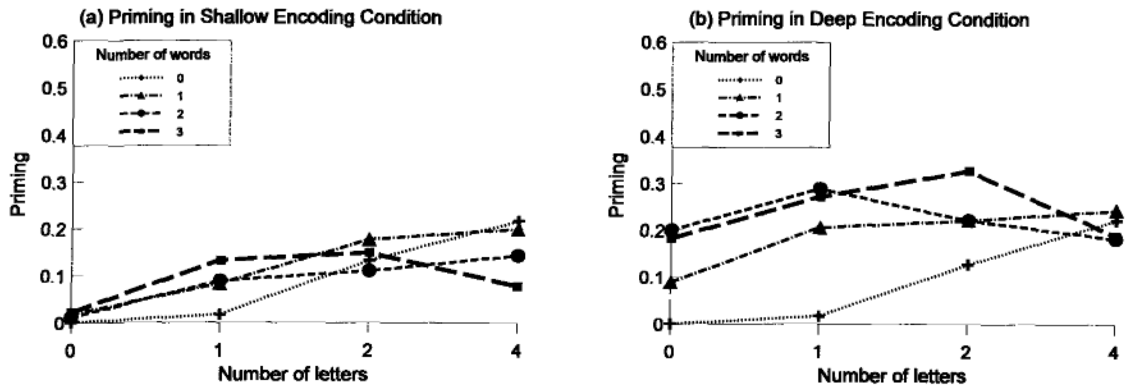


Figure 4. Priming on the implicit test. Priming was computed by subtracting the nonstudied baseline from total performance. Panels (a) shallow encoding, (b) deep encoding.

no level-of-processing effect in the zero-word condition, that is, when no semantic cues accompanied the word fragments, replicating the standard finding (e.g., Roediger et al., 1992). However, when the semantic cues were added (words = 1, 2, and 3), a level-of-processing effect was obtained. To assess this, a 2 × 4 × 4 ANOVA was performed on the implicit test with level of processing, words, and letters as the factors, and it revealed a significant interaction between level of processing and number of words, $F(3, 267) = 9.11$, $MS_e = .090$. We interpret this result as evidence that the semantic cues made the WFC test more conceptually driven, so priming became sensitive to conceptual processing. This outcome was consistent with predictions, and illustrates that implicit tests are not necessarily insensitive to conceptual processing, as is often claimed in the literature.

Note that on both tests, when several letter and word cues were presented (e.g., four letters and/or three words), the level-of-processing effect tended to be small, which may be due to ceiling effects compressing the range.

In summary, the data were consistent with predictions. First, when semantic cues were presented with word fragments, priming became sensitive to conceptual encoding processes. Second, although there was an increased role of conceptual processing on the implicit WFC test, the level-of-processing effect was still larger on the explicit test than on the implicit test, indicating that conceptual information is used to a greater extent during intentional retrieval than during incidental retrieval. Third, multiple sources of information appeared to be combined in a nonadditive fashion during both intentional and incidental retrieval, consistent with the assumption of the FLMP, which we examine in the next section.

*The Fuzzy Logical Model of Perception.* The predictions of the models were fit to the results by using the parameter estimation program STEPIT (Chandler, 1969). Each model was defined as a prediction equation with a set of unknown parameters. STEPIT minimized the deviations between the observed and predicted values of the models by adjusting the parameters in an iterative fashion. 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. Thus RMSD values directly specify the correspondence between the predictions of models and the data. Smaller RMSDs indicate a better fit of the model. Table 3 summarizes the number of parameters and RMSDs for eight of the models we tested, and Table 4 lists the parameters for Models 1, 3, 4, and a simulation explained later

*Models 1 and 2.* Our data are limited with respect to the range of models that can be tested. To establish these limits, we fit two models representing two extremes of the number of parameters employed. These are illustrated in Figures 5 and 6. In the most conservative model, Model 1, we fit the seven-parameter FLMP (three for levels of orthographic information, three for levels of semantic information, and one for background information) to all six conditions (implicit shallow, deep, nonstudied; explicit shallow, deep, nonstudied) with the same set of parameters. Fitting this model requires only seven parameters for the 90 data points (six conditions × 15 levels of letter-cue combinations). For the most liberal form of the model, Model 2, a unique set of seven parameters was estimated for each of the six different conditions for a total of 42 free parameters. The RMSDs were .1099 for the conservative Model 1, and .0324 for the liberal Model 2. Accordingly, these two models provide a window for testing various models. The goal is to find a model that gives an RMSD close to the liberal model without assuming too many free parameters. The conservative model has 35 fewer free parameters with an RMSD increase of about .08. We would be happy with a model that would eliminate about 80 or 90% of this difference and requires only about 4 or 5 additional free parameters. Because there is no formal test of the value of the trade-off between the number of parameters and the RMSD, there is a point at which determining the best model becomes a judgment call.

*Models 3, 4, and 5.* Presently, there are two major viewpoints of memory dissociations. One viewpoint assumes that implicit and explicit tests tap different

TABLE 1
Analysis of Variance: Level of Processing, Test Instructions, Letters, and Words

| Source | *df* | *F* |
|---|---|---|
| | Between subjects | |
| Test instructions (T) | 1 | .64 |
| $S_{within\text{-}group}$ error | 178 | (.332) |
| | Within subjects | |
| Number of words (W) | 3 | 28.19* |
| W × T | 3 | .20 |
| W × $S_{within\text{-}group}$ error | 534 | (.116) |
| Number of letters (O) | 3 | 837.43* |
| O × T | 3 | 1.97 |
| O × $S_{within\text{-}group}$ error | 534 | (.089) |
| Level of processing (L) | 1 | 154.57* |
| L × T | 1 | 6.26* |
| L × $S_{within\text{-}group}$ error | 178 | (.113) |
| W × O | 9 | 6.14* |
| W × O × T | 9 | .96 |
| W × O × $S_{within\text{-}group}$ error | 1602 | (.083) |
| W × L | 3 | 17.54* |
| W × L × T | 3 | 1.33 |
| W × L × $S_{within\ group}$ error | 534 | (.090) |
| O × L | 3 | 7.61* |
| O × L × T | 3 | .66 |
| O × L × $S_{within\ group}$ error | 534 | (.077) |
| W × O × L | 9 | 2.03* |
| W × O × L × T | 9 | 1.77 |
| W × O × L × $S_{within\text{-}group}$ error | 1602 | (.077) |

Note. Values enclosed in parentheses represent mean square errors. $S$ = subjects. *$p < .05$. Performance on the nonstudied items was not included in this analysis.

TABLE 2
Level-of-Processing Effects on the Implicit and Explicit Tests

| | Number of Letters | | | |
|---|---|---|---|---|
| Number of Words | 0 | 1 | 2 | 4 |
| Implicit Test | | | | |
| 0 | 0 | 0 | −.01 | .01 |
| 1 | .06 | .12 | .05 | .05 |
| 2 | .19 | .20 | .11 | .04 |
| 3 | .16 | .13 | .18 | .11 |
| Explicit Test | | | | |
| 0 | 0 | .01 | .05 | .06 |
| 1 | .19 | .22 | .21 | .02 |
| 2 | .23 | .22 | .21 | .08 |
| 3 | .26 | .24 | .05 | .07 |

*Note.* The level-of-processing effect was computed by subtracting performance in the shallow condition from the deep condition.

memory systems (e.g. Tulving & Schacter, 1990; Squire, 1994). Another viewpoint is that they reflect different processing demands at encoding and retrieval (e.g., Roediger et al., 1989; Weldon et al., 1995). Extending the framework of the FLMP promotes a somewhat different perspective, however. The premise is that retrieval is influenced by multiple sources of information. For the nonstudied items, the two information sources that are varied in our task are letter and semantic cues. However, for studied items the encoding episode provides an additional source of information and the value of this source is modulated by shallow or deep encoding. The issue arises as to how the encoding episode should be conceptualized in extending the FLMP. One possibility is that the encoding episode simply represents a third source of information, which is integrated with the letter and semantic information in the same manner as the latter two sources are integrated with one another. Written in equation form, we have Model 3 simply as:

$$P(C) = \frac{bo_i s_j p_k}{bo_i s_j k + (1-b)(1-o_i)(1-s_j)(1-p_k)} \qquad (6)$$

where $p_k$ represents either shallow or deep encoding of the study words. Furthermore, we here assume that the episodic information is integrated in the same way during both implicit and explicit retrieval, so no additional parameter is added for this factor. This yields a nine-parameter model with an RMSD of .0493, which is illustrated in Figure 7. This model appears to be most similar to a view of memory dissociations which would assert that although different types of information may have relatively more or less importance during implicit or explicit retrieval, retrieval can be described with a single mechanism in which there is no fundamental difference in the retrieval process, at least with respect to how information is integrated.

How might episodic information be combined multiplicatively with the letter and semantic cues as an additional source of information? One possibility is that there is a long-term (semantic memory) representation of the word that receives activation during encoding, and the amount of activation is modulated by the encoding instructions. In addition, letter and semantic cues may activate this representation to various degrees during the memory test, and their activation may combine multiplicatively with the activation of the encoding episode. This mutual, multiplicative activation would underlie both implicit and explicit retrieval.

There is a second way to think about how this mechanism might operate. Most contemporary process models of memory dissociations argue that both implicit and explicit retrieval depend largely on the retrieval of episode-specific information, rather than on activation alone. Thus, it is useful to consider how the multiplicative integration of orthographic, semantic, and episodic information might occur under the assumption that the encoding episode establishes a unique episode-specific representation rather than simply the activation of a preexisting representation. During the test, the orthographic and semantic cues may initiate retrieval of two representations, the episodic representation of the word and a long-term representation of it. The cues may recapitulate orthographic and semantic processing that took place during encoding, initiating access to the episodic representation. Simultaneously, the cues may initiate access to a representation of long-term

TABLE 3
Number of Parameters and RMSD Values for the FLMP Model Fits

| | Number of Parameters | RMSD |
|---|---|---|
| Model 1 | 7 | .1099 |
| Model 2 | 42 | .0324 |
| Model 3 | 9 | .0493 |
| Model 4 | 11 | .0480 |
| Model 5 | 9 | .1099 |
| Model 6 | 11 | .0676 |
| Model 7 | 11 | .0757 |
| Model 8 | 16 | .0437 |

*Note.* Features of models are described in the text.

knowledge. As candidate solutions are generated (i.e., a subset of words that have similar orthographies and meanings), they may provide additional retrieval information, thus multiplying the strength of the retrieval cues themselves. This would result in the multiplicative effect of integration described by the FLMP.

One outcome that would be favorable to the multiple systems viewpoint would be if implicit and explicit retrieval required different parameters to provide the best fits to performance. In Model 4, we tested the idea that the encoding experience will have different consequences for implicit and explicit retrieval. The contribution of the encoding episode should thus be different for the two different types of retrieval instructions. This model would assume that the encoding task as well as the retrieval instructions would modulate the contribution of the encoding experience. One test of this prediction within the framework of the FLMP is to assume four different $p_k$ parameters for the four combinations of shallow versus deep encoding and implicit versus explicit retrieval conditions. If indeed the retrieval instructions are critical, then this model with 11 free parameters should give a much better description than the 9-parameter model which assumed the same $p_k$ parameters for implicit and explicit retrieval (i.e., Model 3). It turns out that adding these two additional parameters gives very little improvement in the description of performance, yielding an RMSD of .0480. Thus, there was not a difference in how information was integrated during implicit and explicit retrieval.

This outcome appears to be at odds with the results of the ANOVA, which revealed a significant interaction between level of processing and test instructions. In other words, the ANOVA indicates that test type modulates the value of deep versus shallow encoding. If there were a systematic relation between level of processing and test type, the FLMP should have been able to account for some of this variance by incorporating parameters for implicit and explicit retrieval. We do not have an immediate explanation for this difference between the ANOVA and FLMP, although we have considered several possibilities. First, in Table 2 one can see that the magnitude of the level-of-processing effect varies greatly across letter/word conditions in the explicit and implicit tests, so that the relation is noisy and unsystematic. Second, level of

TABLE 4
Parameter Values for Models 1, 3, 4, and Simulation

| | Model | | | |
|---|---|---|---|---|
| Parameter Source | Model 1 | Model 3 | Model 4 | Simulation |
| Letters (orthographic) | | | | |
| $o_1$ | .676 | .673 | .673 | .715 |
| $o_2$ | .807 | .812 | .811 | .872 |
| $o_4$ | .926 | .932 | .932 | .964 |
| Words (semantic) | | | | |
| $s_1$ | .731 | .733 | .733 | .784 |
| $s_2$ | .804 | .808 | .808 | .858 |
| $s_3$ | .834 | .845 | .845 | .892 |
| Background ($b$) | .046 | [illegible] | .020 | .009 |
| Level-of-processing ($p_k$) | | | | |
| Shallow | — | .671 | — | — |
| Deep | — | .788 | — | — |
| Level-of-processing and Test Instructions ($p_k$) | | | | |
| Implicit/Shallow | — | — | .678 | .576 |
| Implicit/Deep | — | — | .773 | .657 |
| Explicit/Shallow | — | — | .664 | .764 |
| Explicit/Deep | — | — | .803 | .924 |

*Note.* The simulation is described in the text. A dash indicates that the parameter was not included in the model.

processing produced an interaction but not a dissociation between the two tests, so the distinction between them is theoretically weaker.

Third, in Model 4, note that the parameters accounting for retrieval instructions were added after those accounting for level of processing, that is, they were simply added to Model 3. Given that Model 3 already gave a good description, one might hypothesize that this would weaken the ability of the two test parameters to account for much additional variance. A test of this was Model 5, which was a nine-parameter model, except that rather than having two parameters for level of processing as in Model 3, it had two parameters for the test instructions (implicit vs. explicit). The model provided a relatively poor fit, with an RMSD of .1099, adding nothing to the original conservative seven-parameter model (Model 1). Thus, the parameters for level of processing were not precluding the ability of the implicit/explicit parameters to account for variance. Our models provide no evidence that information is integrated differently during implicit and explicit retrieval.

Finally, the difference in the magnitude of the overall level-of-processing effect on the implicit and explicit tests was not large (implicit test, $M$ = .09; explicit test, $M$ = .13), and it may not have been necessary to capture such a small effect to give a good description of the data. To explore this hypothesis, we conducted a simulation to assess how big the difference in the level-of-processing effects between the implicit and explicit tests would need to be to produce a sizable difference in the parameter values for the implicit versus explicit tests. We worked backwards by first generating large differences in the parameter values, and then generating hypothetical predictions for the original results, based on these new parameters. Using Model 4, we increased the difference between the parameters for the implicit and explicit tests

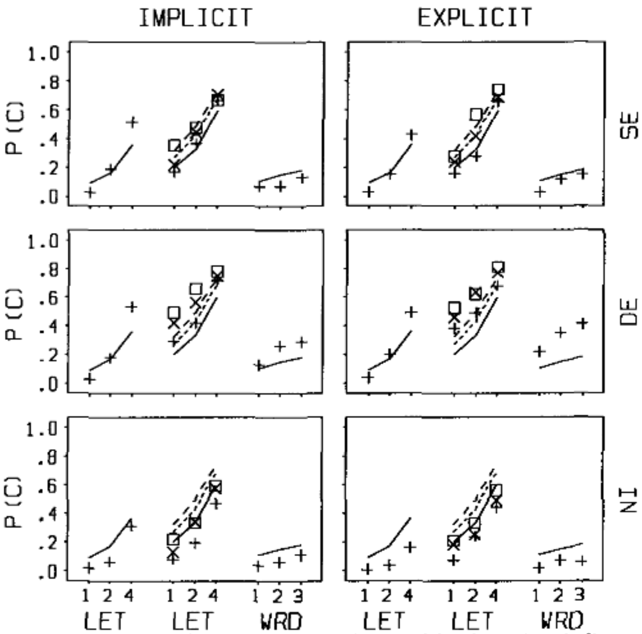


Figure 5. Model 1, the most conservative model. The points indicate the observed data, and the lines the predicted data. Top panels, SE = shallow encoding. Middle panels, DE = deep encoding. Bottom panels, NI = nonstudied items, for the implicit (left panels) and explicit (right panels) tests. In each panel, the left function represents conditions with no words; the middle functions represent conditions with both words and letters; the right function represents conditions with no letters. LET = number of letters; WRD = number of words.

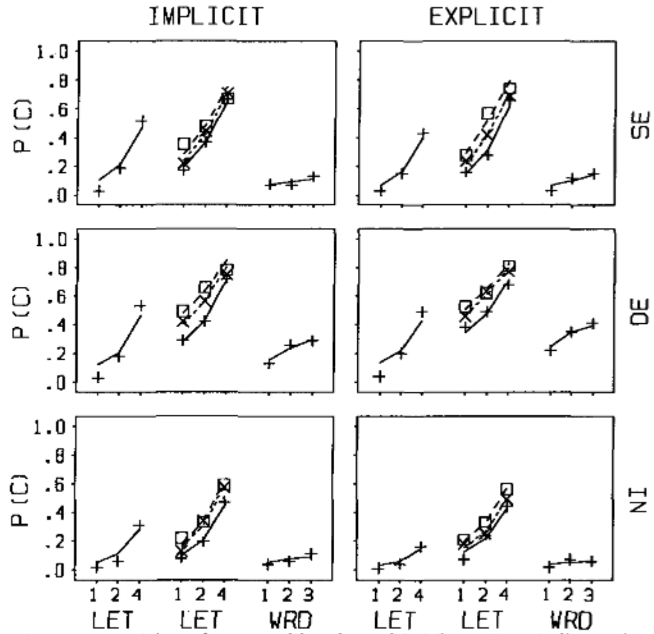


Figure 6. Model 2, the most liberal model. The points indicate the observed data, and the lines the predicted data. Top panels, SE = shallow encoding. Middle panels, DE = deep encoding. Bottom panels, NI = nonstudied items, for the implicit (left panels) and explicit (right panels) tests. In each panel, the left function represents conditions with no words; the middle functions represent conditions with both words and letters; the right function represents conditions with no letters. LET = number of letters; WRD = number of words.

by decreasing the implicit parameter values by 15% and increasing the explicit values by 15%. The original and new values for these four parameters (implicit/shallow, implicit/deep, explicit/shallow, explicit/deep) are shown in Table 4 (Model 4 vs. Simulation). These parameter values were fixed, and then the other parameter values were estimated (see Table 4) to give the best description of the data. The new predicted mean values were implicit/shallow = .24, implicit/deep = .29, explicit/shallow = .37, explicit/deep = .59. Thus, the level-of-processing effect (deep minus shallow) would be .05 on the implicit test, and .22 on the explicit test, larger than that obtained in the present experiment. Interestingly, these hypothetical level-of-processing effects are more in line with what is typically observed on implicit and explicit versions of the WFC test. However, because the implicit WFC test used here was more conceptually driven than traditional versions, due to the addition of the semantic cues, the level-of-processing effect was larger than is usually the case on the implicit test, thereby reducing the difference between the implicit and explicit tests. Also, ceiling effects at the higher range of performance might contribute to an underestimation of the level-of-processing effects.

Thus, any of several different factors might have made it possible for the model to describe the data without assuming an effect of implicit versus explicit retrieval. It remains possible that a different set of variables that produced a dissociation, or a more systematic effect of the encoding variable, or a stronger effect of the encoding variable, could produce evidence for differences in information integration in implicit and explicit retrieval. Also, some other model might be devised to show a significant influence of implicit versus explicit test instructions. For the present, however, our analysis provides some evidence that the nature of retrieval instructions has no effect on how information is integrated during retrieval, at least with respect to the variables we examined here. Of course there is no reason why in principle two independent systems cannot have some similar processes, but our analysis does not provide evidence for independent systems.

*Model 6.* In Model 6 we examined the possibility that episodic information is simply added to the orthographic and semantic information, rather than multiplied with it, yielding

$$P(C) = L_{(i,j)} + E_k \tag{7}$$

where $L_{(ij)}$ is given by Equation 5, and $E_k$ differs for the four the different conditions (deep/implicit; shallow/implicit; deep/explicit; shallow/explicit) This model was fit with 11 parameters and gave a relatively poor description of the results, with an RMSD of .0676. Model 6 does a poorer job describing the results because it assumes that the episodic experience provides only an

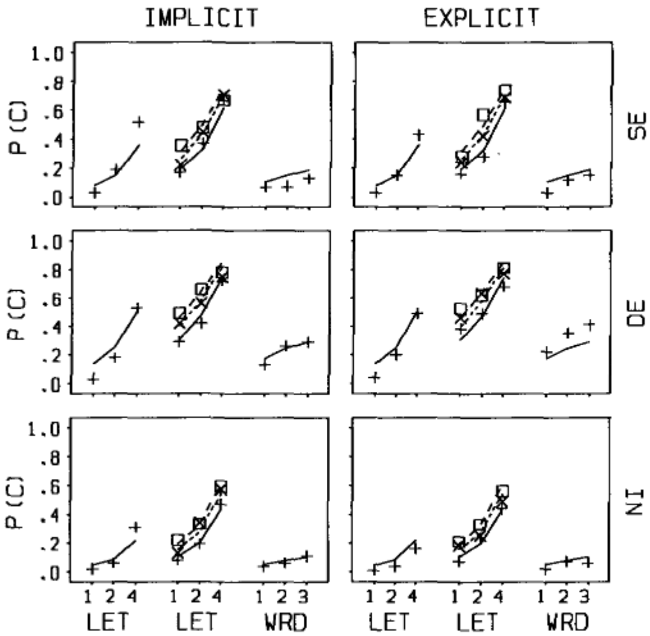


Figure 7. Model 3, which appears to provide the best trade-off between number of parameters and RMSD. The points indicate the observed data, and the lines the predicted data. This model assumes multiplicative integration of information, and does not distinguish between intentional and incidental retrieval. Top panels, SE = shallow encoding. Middle panels, DE = deep encoding. Bottom panels, NI = nonstudied items, for the implicit (left panels) and explicit (right panels) tests. In each panel, the left function represents conditions with no words; the middle functions represent conditions with both words and letters; the right function represents conditions with no letters. LET = number of letters; WRD = number of words.

additive increment to performance, rather than a multiplicative effect as in Model 3.

*Models 7 and 8.* Another reasonable model of memory performance would assume that a person can come up with the correct word in two independent ways. First, the letter and semantic cues could lead to access from long-term memory, as presumably is the case with nonstudied items but also could be the route used to solve studied items. Second, the test stimulus could be used to retrieve the specific encoding episode for studied items. This model makes specific predictions about performance in our tasks. A liberal form of the model would assume that in the episodic route, the letter and semantic cues each provide an extra independent route to retrieval for the studied items relative to the nonstudied items. These letter and semantic cues would be evaluated and integrated for episodic retrieval in the same way they are for the retrieval of long-term knowledge. However, the parameter values would be different in the two retrieval routes, because each route would be accessing a different representation (i.e., long-term knowledge vs. episode-specific memory), and thus retrieval would be constrained differently. For example, in each route the search sets might be constructed differently, or different search strategies might be employed (e.g., parallel vs. serial). This model would require an exorbitant number of free parameters, but it would still be valuable to formulate a model that captures this two-route idea with a fairly small number of free parameters.

In our first instantiation of this model, Model 7, there is some probability of accessing the studied word from the episodic representation, and this probability can differ for the shallow and deep encoding instructions and for the implicit and explicit retrieval instructions, because the level of processing may affect the probability of retrieval differently on implicit and explicit tests. Each of these four conditions (shallow/implicit, deep/implicit, shallow/explicit, deep/explicit) would have some unique probability $E_k$ for accessing the word via the episodic route. Next, item retrieval from long-term knowledge would be described by the basic seven-parameter FLMP. We also have to take into account the probability of accessing the word via both routes. This joint probability has to be subtracted from the sum of the two independent probabilities. Thus, the predictions of Model 7 are

$$P(C) = L_{(i,j)} + E_k - L_{(i,j)}E_k \tag{8}$$

This dual-route model was fit with 11 free parameters and gave a relatively poor description of the results, with an RMSD of .0757.

However, note that Model 7 assumes that the probability of retrieval on the episodic route is independent of the number of letter and semantic cues, which is unlikely to be true. The model can be modified to enable the probability of retrieval to vary with the number of letters and semantic cues. In addition, we retained parameters to account for the level of processing at encoding, which also modulates the probability of episodic retrieval. However, to keep the model from having excessive parameters, we did not include different parameter sets for implicit and explicit retrieval. This seemed reasonable since Models 4 and 5 suggested that distinguishing between implicit and explicit retrieval did not result in a substantial reduction in the RMSD. This results in Model 8 as

$$P(C) = L_{(ij)} + E_{(ij)} - L_{(ij)}E_{(ij)} \tag{9}$$

where $L_{(ij)}$ corresponds to Equation 5 and $E_{(ij)}$ corresponds to Equation 6. Model 8 requires 16 parameters. The long-term memory route incorporates the seven basic parameters, while the episodic route incorporates the seven basic parameters plus two for shallow and deep encoding. This dual-route model yielded an RMSD of .0437. Thus, although it fits the data well, it does require quite a large number of parameters, more than needed to fit the data about equally well with only nine parameters in Model 3.

Note that Models 7 and 8 both assume that the episodic memory and long-term memory routes simply feed forward and add their retrieval information at the decision stage, but this does not appear to be a good account of the retrieval process.

In sum, Model 3 appears to provide the best fit with the fewest parameters. This model assumes that orthographic, semantic, and episodic information are integrated multiplicatively during implicit and explicit retrieval, and that information integration does not differ in implicit and explicit retrieval.

## General Discussion

The goal of this research was to examine the mechanisms that underlie retrieval on implicit and explicit memory tests. During encoding, subjects studied two sets of words, one with a shallow task and one with a deep task intended to increase conceptual processing. Subjects then received either an implicit or explicit WFC test on which they saw varying numbers of letters and semantic cues. Two different types of analyses were used to evaluate the data, ANOVA and model-fitting with the FLMP.

The ANOVA revealed several important outcomes. First, up to a limit, priming increased as the number of retrieval cues increased. In this respect intentional and incidental retrieval are similar. Second, when word fragments were accompanied by semantic cues, the implicit WFC test became more conceptually driven, as evidenced by the fact that a significant level-of-processing effect was obtained when semantic cues were present, but not when they were absent. In the literature, it is often claimed that one of the important differences between implicit and explicit tests is that implicit tests are not sensitive to manipulations of conceptual processing. Clearly, this is an overgeneralization. Other investigators have reported a variety of implicit tests on which priming improves with meaningful encoding, such as category production. Here, we have taken a traditionally data-driven test, WFC, and rendered it more conceptually driven by adding meaningful cues (also see Weldon, Roediger, & Challis, 1989). Thus, implicit tests do show effects of conceptual manipulations when the tests themselves engage conceptual processing. As argued in the transfer appropriate processing framework, it is not simply whether a test is implicit or explicit that determines whether two tests will dissociate. Rather, the match between encoding and retrieval processes must be taken into account.

A third important outcome of the experiment was that even though the two tests provided identical cues, the explicit test engaged conceptual processing to a greater degree than did the implicit test. This was evidenced by the fact that the level-of-processing effect was larger on the explicit test than on the implicit test. Thus, whereas it is inaccurate to say that implicit tests are insensitive to conceptual manipulations, it appears to be true that explicit tests are *more* sensitive than implicit tests given that the same test cues are used on each (also see Weldon et al., 1995).

There are many difficulties in trying to compare directly the levels of performance on an implicit and explicit test because the baselines are usually so discrepant. Typically, the nonstudied baseline is much lower on the explicit test, so that scaling differences hinder any attempt to compare performance on the two tests. One of the advantages of our testing method is that the baselines were virtually identical, enabling us to ignore them for the purpose of comparing the magnitude of the level-of-processing effect on the implicit and explicit test. In addition, performance on the studied items was within the same range on both tests, further mitigating potential scaling problems. However, we cannot claim that this comparison is trouble free because the baselines might still mean something quite different on the explicit and implicit tests. Thus, work by Weldon et al. (1995) is also helpful in addressing the issue of whether explicit tests engage conceptual processing more so than implicit tests. They observed that word fragment completion and picture fragment identification were insensitive to conceptual manipulations when administered as implicit tests, but became sensitive to these manipulations when administered as explicit tests, thus producing a dissociation. These results are consistent with the argument that, all other things being equal, explicit tests do rely on conceptual processing to a greater degree than do implicit tests.

A fourth important outcome of our results was that orthographic and semantic information were combined nonadditively, as indicated by a significant interaction between the number of orthographic and semantic cues. Thus we employed a second analytic technique to examine how multiple sources of information are integrated during explicit and implicit retrieval by testing eight variants of the FLMP. We chose to test several different versions of the FLMP, rather than compare the FLMP to a variety of different models, because previous work has shown that FLMP provides a good description of how orthographic and semantic information are combined in unprimed WFC (a long-term memory test; Massaro et al., 1991; also see Wenger & Payne, 1995).

There is no formal test that can balance the number of parameters and the RMSD to determine which model provides the optimal fit. However, our analyses offered several insights. First, the FLMP provides a good description of how multiple sources of information are combined on both an implicit and explicit episodic memory test. Second, we observed that adding parameters to distinguish between implicit and explicit retrieval did not improve the fit of the model. Overall, it appears that Model 3 provided the best trade-off between number of parameters and RMSD. This suggests that multiple sources of information, in this case orthographic, semantic, and episodic information, are combined in a multiplicative fashion, with the level of processing at encoding modulating the value of the episodic information. Furthermore, these sources of information appear to be integrated in a similar fashion during implicit and explicit retrieval. As discussed previously, this latter outcome represents a difference from the ANOVA. The ANOVA revealed an interaction between level of processing and test type, whereas the FLMP did not need parameters to distinguish between the implicit and explicit tests. The reason for this difference is not entirely clear at this time, although several possibilities were addressed, and more work is needed to understand it. In addition, further work is needed to assess the generalizability of our findings across other implicit and explicit tests.

We would like to note an aspect of our data that has been both valuable and troubling, which is the fact that performance levels were so similar on the implicit and explicit tests. On the one hand this enabled us to compare performance without worrying too much about scaling issues. On the other hand one might be concerned that subjects did not engage very different retrieval modes on the two tests, and may have treated the explicit test much like the implicit test. The fact that fit of the FLMP did not improve when parameters were added for test instructions might partly reflect this possibility. Frequently in this type of work, subjects in the explicit condition are given very strong instructions not to write down an answer unless they are very certain it was presented during the study phase. Thus, the number of nonstudied items completed in the explicit test condition tends to be small (about 10%), so it is easy to see that subjects are performing the implicit and explicit tests differently (e.g., Roediger et al., 1992; Weldon et al., 1989). We did not give our subjects strong instructions in this regard so that we would not add a recognition decision to the initial retrieval process, which might have complicated the types of analyses we wanted to perform. However, one indication that subjects were performing the tests differently is the fact that the level-of-processing effect was larger on the explicit test. Still, it would be valuable to employ a test condition in which subjects are given strong instructions not to complete the nonstudied items, and see if this results in important differences in the ANOVA and the model fits for the explicit test data. This issue touches on the interesting question of how much of the difference between implicit and explicit retrieval is due to retrieval processes versus recognition decision processes (see Jacoby & Hollingshead, 1990; Weldon & Colston, in press), and would be useful to address in future work.

To summarize, an experiment was conducted to evaluate how multiple sources of information are integrated during implicit and explicit retrieval. The analyses revealed that (a) priming generally increases as cue information increases; (b) implicit tests are sensitive to manipulations of conceptual variables when the test engages conceptual processing; (c) explicit retrieval benefits from conceptual processing more than does implicit retrieval; (d) the FLMP does a good job of describing how multiple sources of information are combined multiplicatively during implicit and explicit retrieval; and (e) multiple sources of information appear to be combined in a similar fashion during implicit and explicit retrieval.

This work was supported in part by grants from the Public Health Service (PHS R01 NS 20314), the National Institute of Mental Health (1 R03 MH53857-01), a Faculty Research Grant, and grants from the Division of Social Sciences and the Biomedical Sciences Award Committee of the University of California, Santa Cruz. The authors would like to thank Kyra Kissam, Teddy Johnston, Brian Huebsch, Ann Berton, Jonathan Corbett, and Julie Holloway for their assistance testing subjects. Correspondence can be addressed to either author at the Department of Psychology, University of California, Santa Cruz, CA, 95064, or through e-mail to weldon@cats.ucsc.edu or massaro@fuzzy.ucsc.edu.

**References**

Basden, B.H., Basden, D.R., Church, B.A., & Beaupre, P. (1991). Setting boundary conditions on the part-set cuing effect. *Bulletin of the Psychonomic Society*, *29*, 213-216.

Blaxton, T.A. (1992). Dissociations among memory measures in memory-impaired subjects: Evidence for a processing account of memory. *Memory & Cognition*, *20*, 549-562.

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

Challis, B.H., & Brodbeck, D.R. (1992). Level of processing affects priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 595-607.

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

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.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, *23*, 553-568.

Graf, P., & Ryan, L. (1991). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 978-992.

Jacoby, L.L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language*, *29*, 433-454.

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.

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). London: Academic Press.

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

Massaro, D.W. (1990). The psychology of connectionism. *Behavioral & Brain Sciences*, *13*, 403-406.

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

Massaro, D.W., Weldon, M.S., & Kitzis, S.N. (1991). Integration of orthographic and semantic information in memory retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 277-287.

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

Rappold, V.A., & Hashtroudi, S. (1991). Does organization improve priming? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 103-114.

Roediger, H.L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*, 1043-1056.

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

Roediger, H.L., III, Weldon, M.S., & Challis, B.A. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H.L. Roediger, III & F.I.M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.

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

Schacter, D.L., Bowers, J., & Booker, J. (1989). Intention, awareness and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J.C. Dunn & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 47-65). Hillsdale, NJ: Erlbaum.

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

Squire, L.S. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D.L. Schacter & E. Tulving (Eds.), *Memory Systems 1994* (pp. 204-231). Cambridge, MA: MIT Press.

Tulving, E., & Schacter, D.L. (1990). Priming and human memory systems. *Science, 247*, 301-306.

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

Weldon, M.S., & Colston, H.L. (in press). Dissociating the generation stage in implicit and explicit memory tests: Incidental production can differ from strategic access. *Psychonomic Bulletin and Review.*

Weldon, M.S., & Coyote, K.C. (in press). The failure to find the picture superiority effect in implicit conceptual memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

Weldon, M.S., & Roediger, H.L, III. (1987) Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 21*, 519-528.

Weldon, M.S., Roediger, H.L., III, Beitel, D.A., & Johnston, T.R. (1995). Perceptual and conceptual processes in implicit and explicit tests with picture fragment and word fragment cues. *Journal of Memory and Language, 34*, 268-285.

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

Wenger, M.J., & Payne, D.G. (1995). *Cue integration across encoding and retrieval conditions: Implications for the study of retrieval processes.* Manuscript submitted for publication.

## Sommaire

### L'intégration de l'information lors de la récupération implicite et de la récupération explicite

On a réalisé une expérience pour déterminer (a) si l'information conceptuelle est utilisée davantage dans la récupération intentionnelle que dans la récupération incidente, et (b) comment l'information orthographique et l'information conceptuelle sont intégrées au cours de la récupération incidente et de la récupération intentionnelle. Les sujets ont étudié des listes de mots qui comportaient soit une épreuve de traitement superficielle (compter les voyelles) ou une épreuve de traitement en profondeur (classer par goût). Au cours du test, les sujets ont effectué un test de complètement de fragments de mots (CFM). Les fragments de mots ne contenaient pas de lettre (- - - - - - - - -), contenaient une lettre (- - - - z - - - -), deux lettres, ( - l - - z - - - ) ou quatre lettres ( - l i - z a - - ). De plus, la quantité d'information sémantique qui accompagnait chaque fragment de mot variait. Chaque fragment de mot était présenté sans aucun indice sémantique, avec un indice (blanc, par exemple) deux indices (blanc, aveuglant) ou trois indices (blanc, aveuglant, froid). On a demandé à la moitié des participants de faire de leur mieux pour compléter chacun des fragments de mots (consignes de test implicite) et à l'autre moitié de se rappeler la première liste pour y trouver un mot pour compléter les fragments de mots (consignes de test explicite).

Une analyse de la variance (ANOVA) a révélé plusieurs résultats importants. Premièrement, jusqu'à une certaine limite, l'amorçage augmentait à mesure que le nombre d'indices de récupération augmentait. Dans ce sens, la récupération intentionnelle et la récupération incidente sont pareilles. Deuxièmement, quand les fragments de mots étaient accompagnés d'indices sémantiques, le test implicite de CFM était davantage dirigé par la conceptualisation. Cela est mis en évidence par le fait qu'un effet important de niveau de traitement a été obtenu quand les indices sémantiques étaient présents, mais pas quand ils étaient absents. Troisièmement, l'effet de niveau de traitement était plus important dans le test explicite que dans le test implicite. Ce qui indique que le test explicite exige un traitement conceptuel plus important que le test implicite.

Nous avons ensuite testé plusieurs variantes du modèle de perception inspiré de la logique des ensembles flous pour évaluer comment les informations orthographique, sémantique et épisodique étaient combinées pendant la récupération. Le modèle de perception inspiré de la logique des ensembles flous a fourni une bonne description de la façon dont de multiples sources d'information sont combinées dans les deux tests, l'implicite et l'explicite. Toutefois, l'ajout de paramètres pour faire la distinction entre la récupération implicite et la récupération explicite