

Intention and awareness in perceptual identification priming

ALAN RICHARDSON-KLAVEHN, MERIDETH G. LEE,
RAOUL JOUBRAN, and ROBERT A. BJORK
University of California, Los Angeles, California

We show that inferences can be made about differences in retrieval intentionality between direct and indirect tests, even when those tests involve different physical cues. When the presence of old items was not mentioned (indirect test condition), we observed a crossed double dissociation between perceptual identification priming and recognition memory as a function of a manipulation of data-driven versus conceptually driven processing at encoding. When subjects were instructed to use their memory to help them identify test items (intentional retrieval condition), priming could be expressed as a monotonically increasing function of recognition memory performance. This *reversed association* (Dunn & Kirsner, 1988) between priming and recognition memory cannot be accommodated by a model that views intentional retrieval as common to the tests and attributes the crossed double dissociation to an intertest difference in physical cues and associated processes. A posttest questionnaire measure indicated that awareness of the presence of previously encountered items was ubiquitous among indirect test subjects. Crossed double dissociations between direct and indirect measures can therefore be ascribed to differences in retrieval intentionality but not necessarily to differences in subjective awareness of the past.

In traditional memory tests, subjects are instructed to recall or recognize information concerning an event or events in their personal history. Success in such *direct* tests of memory (Johnson & Hasher, 1987; Richardson-Klavehn & Bjork, 1988) depends on an intention to retrieve information from the past. By contrast, in *indirect* tests of memory, the subject engages in a task that makes sense in the current context, and the instructions do not make reference to the past. Performance in indirect tests can nonetheless be influenced by past events; for example, prior exposure to a word increases the likelihood that it will be used to complete a word stem and enhances accuracy of identification when that word is rapidly flashed or presented with missing letters.

Such facilitation or *repetition priming* effects are sometimes observed when deliberate efforts to remember the events causing facilitation are fruitless. Priming effects can be intact in memory-disordered patients who show floor-level performance in direct tests (see, e.g., Shimamura & Squire, 1984), and can be observed in normals when direct test performance is at chance (see, e.g., Eich, 1984; Kunst-Wilson & Zajonc, 1980). Furthermore, facil-

itatory priming effects can occur in word-stem completion even when subjects are instructed *not* to emit items previously encountered in the experiment and when they are unable to recognize the old items that they produced (Richardson-Klavehn, Gardiner, & Java, 1994). Taken together, these results provide strong evidence that memory can influence performance independent of an intention to retrieve information from the past. Since the subjects in these studies showed no awareness of the past events whose influence was manifest in their behavior, these findings also suggest that influences of memory on behavior can be unconscious, as well as involuntary.

The cases just cited, however, are exceptions among the numerous studies comparing direct and indirect tests. Most studies aimed at distinguishing competing theoretical accounts of involuntary unconscious memory have been conducted with normal subjects, with standard indirect tests, and under conditions in which direct test performance was well above chance. When direct test performance is above chance, and when a standard indirect test is used, facilitation effects could reflect involuntary memory but could also reflect influences of intentional retrieval strategies. Investigators have often dismissed this possibility because of abundant instances of dissociations between direct and indirect tests (for summaries, see Richardson-Klavehn & Bjork, 1988; Roediger, 1990; Roediger & McDermott, 1993; Schacter, 1987) but have tended to overlook cases in which manipulated and non-manipulated variables have influenced performance in the two classes of test in a similar way.

When memory-disordered subjects have shown deficits (compared to normal controls) in both direct and indirect tests, the deficit in the indirect test has routinely been

We thank James I. Chumbley, Gregory V. Jones, Douglas L. Nelson, Henry L. Roediger III, Daniel L. Schacter, Barry S. Stein, and Mary Sue Weldon for their invaluable comments on earlier drafts of this paper. Experiment 3 was performed while the first author was a Visiting Research Fellow at City University. He is indebted to Paul Williams for expert technical assistance and to the members of the Memory and Cognition Research Group for their generous hospitality and for useful discussions. Correspondence should be addressed to A. Richardson-Klavehn, Department of Psychology, City University, Northampton Square, London EC1V 0HB, U.K.

—Accepted by previous editor, Margaret Jean Intons-Peterson

ascribed to spontaneous use of intentional retrieval strategies by control subjects (see, e.g., Cohen & Squire, 1980; Moscovitch, Winocur, & MacLachlan, 1986; Squire, Shimamura, & Graf, 1987). In normal subjects, variables producing parallel effects on direct and indirect measures include repetition in the study list (see, e.g., Graf & Mandler, 1984; Jacoby & Dallas, 1981), study-list length (Slooman, Hayman, Ohta, Law, & Tulving, 1988), level of attention at encoding (Eich, 1984), and instructions to forget versus remember studied items (MacLeod, 1989). Additionally, priming in tests of conceptual information depends on the processing of semantic information at encoding, in the same way as do recall and recognition (see, e.g., Blaxton, 1989; Hamann, 1990; Srinivas & Roediger, 1990). Most recently, Challis and Brodbeck (1992) have shown that depth of processing at study influences performance in indirect tests involving perceptual information (e.g., stem completion). Depth of processing has been regarded as dissociating direct and indirect tests because the effect on indirect measures is generally nonsignificant at the level of the individual study (see, e.g., Graf & Mandler, 1984; Roediger, Weldon, Stadler, & Riegler, 1992; Schacter & Graf, 1986). Across studies in the literature, however, there is a consistent small advantage of deep over shallow processing.

Parallel effects have sometimes been taken to indicate that certain factors influence involuntary and intentional retrieval in the same way (see, e.g., Jacoby, 1983a). However, according to the *retrieval intentionality criterion* (Schacter, Bowers, & Booker, 1989), when subjects employ intentional retrieval, and direct and indirect tests involve the same physical cues, dissociations between direct and indirect tests should not be observed. Parallel effects are therefore equally compatible with the view that the indirect measure was "contaminated" by intentional retrieval. The fact that the physical cues differed between the direct and indirect tests in most of the studies reporting parallel effects only strengthens this argument: Parallel effects are observed despite differences in test cues.

Recent publications reflect increasing concern that indirect measures may not be "process-pure" measures of involuntary memory (see, e.g., Challis & Brodbeck, 1992; Dunn & Kirsner, 1988, 1989; Merikle & Reingold, 1991; Richardson-Klavehn & Bjork, 1988; Schacter et al., 1989). Jacoby (e.g., 1991; Jacoby, Toth, & Yonelinas, 1993) has responded to this type of concern by proposing that the use of standard indirect measures be abandoned in favor of *opposition methodology* (Jacoby, Woloshyn, & Kelley, 1989), in which intentional and involuntary influences of memory are placed in an antagonistic, rather than a synergistic, relationship. As Richardson-Klavehn et al. (1994) point out, where repetition priming is concerned, opposition tests have disadvantages—as well as advantages—that standard indirect tests do not have. Subjects in opposition tests are instructed not to respond with previously encountered items. In consequence, the likelihood is minimal that priming effects in opposition tests reflect facilitatory influences of intentional retrieval pro-

cesses. However, anecdotal evidence indicates that subjects in indirect tests often respond with items that they did not intend to retrieve but that they consciously recognize (see, e.g., Schacter et al., 1989), a phenomenon we refer to as *involuntary conscious memory*.¹ In an opposition test, by contrast, subjects will suppress items associated with an awareness of the past, even if those items are initially generated as a result of involuntary retrieval processes. This characteristic makes opposition tests particularly suitable for studying influences of memory that are not only involuntary but also unconscious. Nevertheless, until we have a better understanding of the role of involuntary conscious memory, it seems imprudent to abandon the use of indirect tests.

Crossed double dissociations between direct and indirect memory measures have been cited as strong evidence that indirect test performance reflects involuntary retrieval processes (see, e.g., Allen & Jacoby, 1990; Jacoby, 1983b). The term *crossed double dissociation* (Dunn & Kirsner, 1988) refers to cases in which the same variable produces opposite effects on two different memory tests. For example, previous experiments have shown that generating an item at encoding produces better recognition memory than does reading it but that reading an item produces more priming in perceptual identification than does generating it (see, e.g., Jacoby, 1983b). A number of other studies have shown similar crossed double dissociations between direct and indirect tests as a function of generate/read and picture/word manipulations at encoding (see, e.g., Blaxton, 1989; Weldon & Roediger, 1987; Winnick & Daniel, 1970; for reviews, see Richardson-Klavehn & Bjork, 1988; Roediger & McDermott, 1993). The assumption has been that such dissociations are incompatible with the involvement of intentional retrieval processes in the indirect tests, because conditions that improve the success of intentional retrieval attempts (i.e., generating an item, or studying it in pictorial form) lead to a reduction in priming effects.

Unfortunately, the conclusion that crossed double dissociations indicate differences in retrieval intentionality between tests is underdetermined by the data. The great majority of studies showing such patterns of data fall afoul of the retrieval intentionality criterion, because the tests compared differed in terms of physical test cues, as well as in terms of instructions (for an exception see Java, 1994). Intentional retrieval could, therefore, be common to both the direct and indirect tests, the crossed double dissociation being produced by an interaction between the processing demands of the test and the type of information encoded. In consequence, the observed dissociations cannot conclusively be attributed to a difference in retrieval intentionality between tests. Still more contentious is the assumption that such dissociations of task performance imply differences between tests in conscious awareness of past events. Even if indirect test performance reflects only involuntary influences of memory, it is by no means clear that those involuntary influences are unaccompanied by subjective awareness of the past.

Awareness of the past in indirect tests has rarely been a subject of systematic investigation. Bowers and Schacter (1990) classified subjects in their *test-uninformed* condition (indirect test condition) into test-aware and test-unaware groups on the basis of their responses to a structured interview that followed a word-stem-completion test. Subjects who indicated awareness that some of the stems in the test could be completed with previously encountered items showed an effect on priming of depth of processing at encoding similar to that observed in direct tests; subjects who did not indicate awareness of the presence of old items showed no significant effect. Bowers and Schacter also ran a *test-informed* condition, in which subjects were told of the presence of stems corresponding to studied items but asked to use the first item that came to mind, despite their knowledge of the study-test relationship. Like test-unaware subjects, these subjects also showed no significant effect of depth of processing on priming. Taken together, these data suggest that spontaneous test awareness during an indirect test prompts subjects to adopt an intentional retrieval strategy.

This important study, however, leaves several unanswered questions. Bowers and Schacter did not ask their subjects whether or not they had adopted an intentional retrieval strategy; it is, therefore, possible that awareness of the presence of previously encountered items was accompanied by the use of an intentional retrieval strategy only for some subjects in their test-aware group. In addition, they did not ask test-aware subjects whether they suspected prior to the test that it would contain old items, as might happen if subjects tried to "second guess" the purpose of the experiment, or whether they became aware of the presence of old items only after they had started the test, which would represent an instance of involuntary conscious memory. Finally, Bowers and Schacter examined the relationship between test awareness and the effect of a variable that, at best, produces a single dissociation between direct and indirect tests (i.e., depth of processing often exerts a significant effect on direct test performance, in conjunction with no significant effect on indirect test performance). Since greater theoretical resolving power has been attributed to crossed double dissociations between tasks than to single dissociations, it would be useful to determine whether test-aware indirect-test subjects do or do not show such crossed double dissociations.

With this background in mind, we had two major aims in the current research: (1) to determine whether a crossed double dissociation between a direct and an indirect test could be attributed to a difference in retrieval intentionality between tests, even when those tests involved different physical retrieval cues; and (2) to explore the relationship between awareness of the encoding episode and indirect test performance in the context of a crossed double dissociation between direct and indirect tests, and—if such awareness occurred at all—to examine how it came about and how it was related to the test strategies adopted by subjects.

EXPERIMENT 1

In Experiment 1, we employed both a manipulation designed to affect subject strategy and a posttest questionnaire designed to measure test awareness (Bowers & Schacter, 1990) and subject strategy. We used a test of perceptual identification, because it has been claimed that repetition priming in this test is uninfluenced by intentional retrieval strategies (see, e.g., Jacoby & Dallas, 1981). In one test condition, the subjects were instructed to use their memory for previously encountered items to help them identify test items. In the other test condition, standard indirect test instructions were given; that is, the subjects were simply told to identify the test items, and no mention was made of the presence of previously encountered items.

We were unsure whether an instructional manipulation alone would produce the desired changes in strategy in the perceptual identification test. In consequence, we deliberately confounded the instructional manipulation with a manipulation of the percentage of test items that were previously presented (see Jacoby, 1983a, for a similar confounding). This confounding was designed to make the effect of instructions as powerful as possible and to render the instructions plausible to subjects. Given a significant effect of such a "sledgehammer" manipulation, we could unravel the effect of instructions from any effect of percentage of old test items in a later experiment (see Experiment 3 below). In the intentional retrieval condition, therefore, 90 of the 120 test items had been previously encoded, and 30 were new (75% of test items old); in the standard indirect test condition, 30 of the 120 items had been previously encoded, and 90 were new (25% of test items old). In both instructional conditions, the critical items were 30 old and 30 new low-frequency words. The remaining 60 test items were fillers, consisting of medium- and high-frequency words. In the intentional retrieval condition these items were all old, whereas in the standard indirect test condition, they were all new. Word frequency is positively related to identifiability regardless of old/new status at test (see, e.g., Jacoby, 1983a; Jacoby & Dallas, 1981). When these filler items were old, the subjects were expected to identify a large proportion of them and recognize them as old, reinforcing the effect of the instructions to use a recollective strategy. When they were new, the subjects were expected to identify a large proportion of them and recognize them as new, reinforcing the impression that there was no relation between the encoding and test phases of the experiment. Low-frequency words were used as the critical items because they show larger benefits of prior presentation in tests of perceptual identification than do medium- and high-frequency words (see, e.g., Jacoby & Dallas, 1981).

Following the perceptual identification test, the subjects filled out a questionnaire that asked them if they were aware of the presence of old items on the test list. If they were test-aware, they were asked (1) whether they sus-

pected prior to the test that it would contain old items or whether they became aware of the presence of old items after beginning the test, and (2) whether or not they had attempted to use their memory for previously encountered items to help them identify test items.

The manipulation of test instructions was combined with a manipulation of data-driven versus conceptually driven processing at encoding. This manipulation was expected to produce a crossed double dissociation between performance in the test of recognition memory and performance in the perceptual identification test when standard indirect instructions were given. Given this pattern of data, we could conduct an experimental test of the hypothesis that the crossed double dissociation reflected an interest difference in retrieval intention. If this hypothesis is correct, the crossed double dissociation between recognition memory and perceptual identification priming as a function of type of processing at encoding (data-driven vs. conceptually driven) should disappear when intentional retrieval strategies are used in perceptual identification.

The obvious candidate for a manipulation producing opposite effects on recognition memory and perceptual identification priming was a generate/read manipulation, as used by Jacoby (1983b). However, in conditions requiring subjects to generate low-frequency words, generation failures are endemic unless a fragment corresponding to the word is presented. We believed that presenting a fragment of a to-be-encoded item in the conceptually driven encoding condition would induce some data-driven processing, reducing the difference in encoded information between conditions. To circumvent these difficulties, we used an analog of the more conventional generate/read manipulation. In the conceptually driven encoding condition, items were presented auditorily, and the subject judged how recently the referent of the word had been encountered (auditory-deep condition). In the data-driven encoding condition, items were presented visually and the subjects counted the number of enclosed spaces in the letters of the word (visual-shallow condition). The auditory-deep encoding condition did not present the perceptual "data" that would later appear in the test and involved a task that emphasized the semantic properties of the item; on the other hand, the visual-shallow condition presented the perceptual data that would later appear in the test and involved a task that emphasized the sensory characteristics of the item.

Direct test subjects received a "yes"/"no" recognition memory test instead of a perceptual identification test. As in the perceptual identification test, there were two groups, which differed in the percentage of test items that were old (75% vs. 25%). Both groups were expected to show better *old/new* discrimination in the auditory-deep condition than in the visual-shallow condition. On the basis of signal detection considerations (Green & Swets, 1974), we expected percentage of old test items to influence criterion for a "yes" response, but we did not expect it to influence discrimination. Given this result, we could rule out the hypothesis that the perceptual identification subjects given intentional retrieval instructions (for whom the

majority of test items were old) were differentially able to recollect encoding events in comparison with the subjects given standard indirect instructions (for whom the majority of items were new).

Method

Subjects. The subjects were 134 male and female undergraduate students at the University of California, Los Angeles. They participated in order to fulfill a requirement for an introductory course in psychology and were screened for fluency in English. Six subjects in the perceptual identification groups were replaced because they showed either ceiling or floor effects, leaving a total of 128 subjects.

Design. The experiment was a 2 (recognition memory vs. perceptual identification) \times 2 (75% of test items old vs. 25% old) \times 3 (auditory-deep vs. visual-shallow vs. new items) factorial design. Test type and percentage of test items that were old were both manipulated between subjects, with 32 subjects randomly assigned to each of the four groups. In the perceptual identification groups, the percentage-of-old-test-items variable was accompanied by a manipulation of instructions: When 75% of test items were old, the subjects received intentional retrieval instructions; when 25% of test items were old, standard indirect test instructions were given. In the recognition memory groups, the subjects were informed of the percentage of test items that would be old. Encoding condition was manipulated within subjects. The subjects in all groups encoded a list of 90 items in Part 1 of the experiment, with 45 items in each of the two encoding conditions. Fifteen of the 45 items in each condition were critical low-frequency words.

Materials. Sixty high-frequency, 90 medium-frequency, and 60 low-frequency nouns were selected from Kučera and Francis (1967). Frequency of occurrence per million words of text was 50 or more (high frequency), 10-49 (medium frequency), and 1-5 (low frequency). All words had between four and six letters. They were selected so that it would make sense to the subjects, in the judgment of the experimenters, to perform a judgment of recency of encounter with the referent of the word. This criterion meant that most of the words referred to concrete objects, although some abstract nouns (e.g., *refund*) were included. To preclude ambiguity in the auditory-deep condition, we excluded homophones and words referring to objects in the experimental room.

The high-frequency and low-frequency word sets were each randomly assigned into two 30-word pools (H1, H2; L1, L2). The medium-frequency words were randomly assigned into three 30-word pools (M1, M2, M3). Items in Pool M3 were used as practice items: 2 were used as examples in the instructions prior to the encoding phase, 8 were used in the practice encoding trials, and 20 were used in the practice perceptual identification test. Pools M1 and M2 were used as experimental items.

Four 90-item study lists were constructed, each consisting of 30 high-, 30 medium-, and 30 low-frequency words. A single block-randomized order of frequency categories was used for all four lists. Each of the 30 three-word blocks contained one word from each of the three categories. The four lists differed in the specific pools of words that they contained: H1, M1, L1 (Study List 1); H1, M1, L2 (Study List 2); H2, M2, L1 (Study List 3); or H2, M2, L2 (Study List 4). Two 120-item test lists were created, each containing 30 high-, 30 medium-, and 60 low-frequency words. These lists contained both Pools L1 and L2, and differed only in the high- and medium-frequency words that they contained: H1 and M1 (Test List 1) or H2 and M2 (Test List 2). A single block-randomized order of frequency categories was used for both test lists. The list was divided into 30 four-item blocks, each containing one word from Pool L1 and one word from Pool L2, plus a high- and a medium-frequency word.

The manipulation of percentage of old test items and the counterbalancing of *old/new* status at test for the critical low-frequency items were accomplished by administering all eight possible com-

binations of study and test lists across subjects. For Test List 1, administering Study Lists 1 and 2 resulted in a 75% overlap between study and test items, whereas administering Study Lists 3 and 4 resulted in a 25% overlap. For Test List 2, Study Lists 1 and 2 resulted in a 25% overlap, and Study Lists 3 and 4 resulted in a 75% overlap. For both test lists, when Study Lists 1 and 3 were presented, Pool L1 was old, and Pool L2 was new, and when Study Lists 2 and 4 were presented, Pool L1 was new, and Pool L2 was old. Across subjects, therefore, both pools of critical items appeared under all four possible combinations of study-test overlap percentage and *old/new* item status at test. By contrast, overlap percentage and *old/new* status were deliberately confounded for the noncritical high- and medium-frequency items: When they were old, they were in the 75% overlap condition, and when they were new, they were in the 25% overlap condition. However, the sets of high- and medium-frequency items (H1 and M1 vs. H2 and M2) forming the noncritical items were varied systematically at study and at test.

To counterbalance encoding condition across items, eight unique encoding formats were created from the four study lists (Formats 1 and 5 from Study List 1, Formats 2 and 6 from Study List 2, Formats 3 and 7 from Study List 3, and Formats 4 and 8 from Study List 4). The items in each list were divided into six 15-item blocks. In view of the construction of the study lists, each block therefore contained 5 words from each frequency category. In Formats 1 through 4, items in odd-numbered blocks were in the visual-shallow condition, and items in even-numbered blocks were in the auditory-deep condition. In Formats 5 through 8, items in odd-numbered blocks were in the auditory-deep condition, and items in even-numbered blocks were in the visual-shallow condition.

Four test formats were created from the two test lists. Test List 1 was presented for a perceptual identification test in Format 1 and a recognition memory test in Format 3; Test List 2 was presented for a perceptual identification test in Format 2 and a recognition memory test in Format 4. Combining the eight study formats with the four test formats resulted in 32 unique study/test combinations. Four subjects were assigned to each combination, for a total of 128 subjects.

Apparatus and Procedure. An IBM AT personal computer and Sony PVM 1271Q color monitor were used to present stimuli and record the subjects' responses. The subjects were told that the experiment concerned various aspects of their knowledge of English words. Prior to the encoding phase, they were told to make one of two types of judgments; which judgment they made would depend on how a word was presented to them. When a word appeared in uppercase blue letters on the screen, they were to count the number of enclosed spaces in the letters of the word (visual-shallow condition). An enclosed space was defined as a space surrounded by a continuous line, as in the letters A, B, D, O, P, Q, and R. Each word appeared for 3 sec. At the disappearance of the word, the computer prompted the subject for a numerical judgment, which was entered via the keyboard. When a row of green asterisks within red quotation marks appeared on the screen for 3 sec, the experimenter would say the to-be-judged word, and the subjects were to judge how recently they had encountered the thing denoted by the word (auditory-deep condition). At the disappearance of the asterisk stimulus, the subject entered their numerical judgment using a 7-point scale (1 = *minutes ago*, 2 = *hours ago*, 3 = *days ago*, 4 = *weeks ago*, 5 = *months ago*, 6 = *years ago*, and 7 = *never*). The category corresponding to each number was marked at the top of the keyboard. The subjects were told that they should make this judgment by retrieving a specific encounter with the referent of the word, but if they could not, to try and make a reasonably accurate judgment. The subjects never indicated any difficulty in making the judgment. In both encoding conditions, the subjects were allowed a maximum of 20 sec to make their judgment before the computer skipped to the next word. Pushing the "Enter" key before 20 sec had elapsed initiated the next trial. To acquaint subjects with

the tasks, a practice list of 8 medium-frequency items (four visual-shallow, four auditory-deep) was presented prior to the experimental trials. In the experimental trials, the subjects alternated between the two types of judgments, changing judgment every 15 words.

Following the encoding trials, all subjects received a perceptual identification test consisting of 20 new medium-frequency items. This test served as practice for perceptual identification subjects and as a distractor for recognition memory subjects. All subjects were told that the test served to index another aspect of their knowledge of words. On each trial, a fixation point (a white asterisk) appeared for 1 sec above the position in which the first letter of the word would appear. The target word was then flashed in yellow letters against a blue background, a display type that was selected in order to produce sharply defined edges to the letters of the word. After each word was flashed, it was replaced by a mask (details of mask type below), and the computer prompted the subject to enter a response via the keyboard. Pushing the "Enter" key initiated the next trial. In order to "shape" the subject's identification performance, the exposure duration was 148 msec for the first item and was reduced by 10 msec for each successive item until it reached 48 msec on Trial 11. For the remaining trials the exposure duration remained at 48 msec, the duration that was subsequently used for the critical perceptual identification trials.² The subjects were told that they would often not be sure of the identity of the flashed word, but that if they had some idea of its identity, they should enter the word that they believed had been presented. They were allowed up to 20 sec to make a response before the computer skipped to the next item. If they had no idea of its identity, pushing the "Enter" key before 20 sec had elapsed initiated the next trial.

Following the practice trials, different instructions were given depending on between-subject experimental condition. The perceptual identification subjects were told that they would continue to perform the task they had just performed. The subjects in the standard indirect test group were not told that the test would contain words from the encoding phase of the experiment. The subjects in the intentional retrieval group were told that the upcoming test list would consist largely of words from the encoding phase of the experiment and that they should try to use their memory for those items to help them identify test items. The procedure on each of the identification trials was exactly the same as in the practice trials. The subjects in the recognition memory groups pushed either the Y key if they believed that the word had been previously presented or the N key if they believed that it had not. Recognition test items were presented in yellow uppercase letters against a blue background, as in the perceptual identification test, but they remained on the screen until one of the two response keys was pushed, at which time the computer skipped to the next word. The maximum time allowed for a response was 20 sec. The subjects for whom 25% of the test items were old were told that about a quarter of the upcoming test items would be old; the subjects for whom 75% were old were told that about three quarters would be old. These instructions were designed to mirror as far as possible the instructions received by the two groups of perceptual identification subjects.

At the conclusion of the test, the perceptual identification subjects filled out a questionnaire. They were told that the purpose of the questionnaire was to determine their state of awareness and their strategies during the test, and that they should answer the questions as accurately as possible. Each question presented the subjects with two response alternatives; the subjects checked a box indicating their choice. The first question asked them whether they were aware that some of the flashed words had been presented previously in the experiment ("yes"/"no" response). They were required to answer the remaining questions only if they answered this question affirmatively. The second question asked them whether (1) they had suspected prior to the beginning of the test that some of the flashed words would be words that had been previously pre-

sented or (2) they had realized at some time during the course of the test that some of the flashed words had been previously presented. The third question asked them whether they had actively tried to use their memory for words that they had encountered before to help them identify the flashed words ("yes"/"no" response).

After 64 of the 128 subjects had been tested (16 per between-subject group), the mask used in the perceptual identification trials was changed because of an update in the program used to present stimuli. For the first 32 perceptual identification subjects, therefore, the mask consisted of a row of white "at" symbols (i.e., @@@@). For the remaining 32 subjects, it consisted of a rectangular white field with a zigzag arrangement of black lines superimposed on it. The change in the program did not affect recognition memory trials. Since the change was made after the first complete iteration through the 32 study/test formats, mask type constituted an additional factor that was completely crossed with the other factors in the perceptual identification portion of the design.³

Results

Recognition memory data. The mean proportions of "yes" responses to critical low-frequency old and new test items are presented in Table 1. Means for auditory-deep and visual-shallow items are hit proportions and are based on 15 observations per subject per condition; means for new items are false alarm proportions and are based on 30 observations per subject. Because of ceiling effects in the auditory-deep condition and floor effects in new item scores, the distributions of scores in these conditions showed significant deviations from normality (i.e., skewness and kurtosis). In view of these deviations, we eschewed analysis of variance of the hit and false alarm proportions. Instead we transformed the data by using a signal detection model prior to analysis.⁴ Applying the model permitted us to examine two critical questions separately: (1) Did the auditory-deep condition lead to superior *old/new* discrimination than did the visual-shallow condition? and (2) Did the manipulation of percentage of old test items affect *old/new* discrimination, or did it simply affect overall bias toward emitting "yes" responses?

Two d' scores were calculated for each subject by using the distribution of the standard normal variable—one for the auditory-deep condition and one for the visual-shallow condition. An estimate of criterion for emitting a "yes" response was obtained concomitantly. In order to ensure that d' was defined for each subject in each condition, the subjects showing a hit score of 1.0 were arbitrarily assigned a score of .99, and the subjects showing a false alarm score of zero were assigned a score of .01. The resulting mean d' and criterion scores are presented in Table 2. The distributions of these scores showed no significant deviations from normality. A 2×2 analysis of

Table 2
Mean d' and Criterion Scores for Critical Low-Frequency Items in Experiment 1 as a Function of Encoding Condition and Percentage of Test Items That Were Old

Percent Test Items Old	d'		Criterion
	Auditory-Deep	Visual-Shallow	
25	3.37	1.27	1.81
75	3.14	1.25	1.18

variance (ANOVA) of the d' scores revealed a highly significant main effect of encoding condition [$F(1,62) = 359.57, p < .001$], with the auditory-deep condition leading to better *old/new* discrimination than did the visual-shallow condition. Percentage of old test items did not influence discrimination overall ($F < 1$) and did not interact with encoding condition [$F(1,62) = 1.02, p > .25$]. However, a separate one-way analysis of the criterion scores showed a highly significant effect of percentage of old test items [$F(1,62) = 16.94, p < .001$]. The subjects for whom 75% of test items were old used a more liberal criterion for emitting a "yes" response than did subjects for whom 25% of the test items were old.

Perceptual identification data. For the analyses to be reported here, the subjects were scored as correct only when they responded with the exact letter string presented. Including potential typographical errors and grammatical transformations of target items as correct responses did not change the pattern of data. Four subjects in the intentional retrieval group and 2 in the standard indirect group were replaced during the course of the experiment because their performance on critical low-frequency words was above .90 or below .10 for all three item types (auditory-deep, visual-shallow, and new).

The mean proportions of critical low-frequency items correctly identified are presented in Table 3. Means for old items are based on 15 observations per subject per condition, whereas means for new items are based on 30 observations per subject. The data were subjected to a $2 \times 2 \times 3$ ANOVA, with mask type ("at" symbol vs. zigzag) and instructions (standard indirect vs. intentional retrieval) as between-subject factors and item type (auditory-deep vs. visual-shallow vs. new) as a within-subject factor. The main effect of mask type was highly significant [$F(1,60) = 14.14, p < .001$], the zigzag mask ($M = .70$) leading to better performance than the "at" symbol mask ($M = .49$). However, mask type did not enter into significant interactions with any of the other factors (all $ps > .20$), so the data presented in Table 3 are averaged over the levels of mask type. The main effect of item type was highly significant [$F(2,120) = 15.12, p < .001$], but this effect was qualified by a significant interaction between item type and instructions [$F(2,120) = 3.97, p < .025$], reflecting different profiles across the means of the three item types for the two groups. The main effect of instructions was nonsignificant ($F < 1$).

Planned comparisons (averaged over the levels of mask type) indicated different patterns of repetition priming for

Table 1
Mean Proportions of "Yes" Responses to Critical Low-Frequency Items in Experiment 1 as a Function of Encoding Condition and Percentage of Test Items That Were Old

Percent Test Items Old	Old Items		New Items
	Auditory-Deep	Visual-Shallow	
25	.91	.34	.06
75	.96	.52	.16

Table 3
Mean Proportions of Critical Low-Frequency Items
Identified in Experiment 1 as a Function of
Encoding Condition and Test Instructions

Instructions	Old Items		New Items
	Auditory-Deep	Visual-Shallow	
Standard indirect	.58 (.04)	.62 (.08*)	.54
Intentional retrieval	.66 (.13*)	.63 (.10*)	.53

Note—Means in parentheses are priming scores. *Significant priming.

the intentional retrieval and standard indirect test groups.⁵ For the standard indirect group, prior encoding of items in the visual-shallow condition resulted in an improvement in performance when compared with new items [$F(1,60) = 14.40, p < .001$], whereas encoding items in the auditory-deep condition did not [$F(1,60) = 2.17, p > .10$]. By contrast, for the intentional retrieval group, prior encoding of items in both the visual-shallow [$F(1,60) = 19.13, p < .001$] and auditory-deep conditions [$F(1,60) = 27.75, p < .001$] produced significant performance improvements, in comparison with the new items. Baseline (new item) performance did not differ significantly between the groups ($F < 1$).

An additional 2 (mask type) \times 2 (instructions) \times 2 (encoding condition) ANOVA was conducted on priming proportions, obtained by subtracting each subject's new-item score from their scores for previously encountered items. The mean priming proportions are presented parenthetically in Table 3. This analysis revealed a marginally significant main effect of mask type [$F(1,60) = 2.87, .05 < p < .10$], with the "at" sign mask leading to somewhat greater priming than did the zigzag mask. This result would be expected in virtue of the higher overall level of performance obtained with the zigzag mask, which left less headroom for priming effects to be observed. As in the previous analysis, however, mask type did not enter into significant interactions with the other factors in the design (all $ps > .20$). The main effect of instructions was significant [$F(1,60) = 4.41, p < .05$], with the intentional retrieval group showing a greater overall level of priming than did the standard indirect group, whereas the main effect of encoding condition was nonsignificant ($F < 1$). The interaction of instructions and encoding condition just failed to reach significance [$F(1,60) = 3.76, p = .054$].

Analysis of questionnaire responses. Of the 32 subjects in the standard indirect test group, 2 (6.3%) were unaware that the perceptual identification test contained old items, 5 (15.6%) suspected prior to the test that it would contain old items, and 25 (78.1%) realized at some time during the test that it contained old items. Of the 30 test-aware subjects in this group, 13 (40.6% of total) indicated that they had attempted to use their memory for previously presented items to help them identify test items, and 17 (53.1% of total) indicated that they had not. One of the 5 subjects who suspected the presence of old items fell into the former group; the remaining 4 subjects fell into the latter group. The subjects in the intentional retrieval group were informed prior to the test that old items

would be present: As expected, all indicated awareness of the presence of old items. Of these 32 subjects, 28 (87.5%) indicated that they had attempted to use their memory for previously presented items, and 4 (12.5%) indicated that they had not (contrary to instructions). The relationship between instructions and subject status at test (test-unaware vs. test-aware/did not use memory vs. test-aware/used memory) was highly significant [$\chi^2(2) = 15.54, p < .001$], confirming the effectiveness of the instructional manipulation.

To examine the relationship between test strategy and perceptual identification performance, the data for the subjects in the intentional retrieval and standard indirect test groups were separated according to whether or not an intentional retrieval strategy had been employed (as indicated by questionnaire responses). For this purpose, the two test-unaware subjects in the standard indirect group were grouped together with test-aware subjects who did not indicate use of an intentional retrieval strategy. The mean proportions identified are presented in Table 4. The data were subjected to a 2 (test strategy) \times 2 (instructions) \times 3 (item type) ANOVA, which revealed a significant interaction between test strategy (as indicated by the questionnaire) and item type [$F(2,120) = 4.00, p < .025$]. The only other significant source of variance in this analysis was the main effect of item type [$F(2,120) = 9.09, p < .001$]. Most critically, instructions (which were confounded with percentage of old test items) did not interact with item type [$F(2,120) = 0.23, p > .50$] when the variance associated with test strategy was extracted separately. These results suggest that the percentage-of-old-test-items variable could not have directly produced the interaction between instructions and item type obtained in the overall analysis reported above.

Planned comparisons (conducted on the unweighted means) indicated significant priming in the visual-shallow condition [$F(1,60) = 9.12, p < .01$], but not in the auditory-deep condition ($F < 1$), for the subjects who did not indicate the use of intentional retrieval. By contrast, the subjects who reported intentional retrieval showed significant priming in both the auditory-deep condition

Table 4
Mean Proportions of Critical Low-Frequency Items
Identified in Experiment 1 as a Function of
Encoding Condition, Subject Strategy, and Test Instructions

Instructions	Old Items		New Items
	Auditory-Deep	Visual-Shallow	
Subjects not indicating an intentional retrieval strategy			
Standard indirect	.63 (.02)	.69 (.08)	.61
Intentional retrieval	.52 (-.01)	.67 (.14)	.53
<i>M</i> (unweighted)	.57 (.00)	.68 (.11*)	.57
Subjects indicating an intentional retrieval strategy			
Standard indirect	.50 (.06)	.53 (.09)	.44
Intentional retrieval	.68 (.15)	.62 (.09)	.53
<i>M</i> (unweighted)	.59 (.11*)	.57 (.09*)	.48

Note—Means in parentheses are priming scores. *Significant priming.

$[F(1,60) = 18.57, p < .001]$ and the visual-shallow condition $[F(1,60) = 21.40, p < .001]$.

Discussion

As expected, performance in the recognition memory test was much better in the auditory-deep condition than in the visual-shallow condition, consistent with the well-known finding that direct-test performance benefits from processing of semantic-elaborative information at encoding. The advantage of auditory-deep over visual-shallow encoding occurred despite the fact that encoding and test modalities differed in the former condition and matched in the latter condition. On the other hand, priming in the indirect test of perceptual identification occurred only in the visual-shallow condition, in which modality matched between prior exposure and test. This difference in the data pattern between tests is particularly impressive because (1) priming did not occur in a condition that led to extremely accurate recognition memory performance, and (2) priming did occur in a condition that led to very poor recognition memory performance. The subjects in the intentional retrieval group showed significant priming in both the auditory-deep and visual-shallow conditions. This result encourages the conclusion that the effects of intentional retrieval strategies in a perceptual identification test depend on the use of semantic-elaborative information in memory traces corresponding to old items. This type of information also supports performance in traditional direct tests of recall and recognition.

The manipulation of percentage of test items that were old combined with a manipulation of conceptually driven versus data-driven processing at encoding bears some similarity to a design used by Allen and Jacoby (1990), who showed that increasing the percentage of old test items resulted in an increase in the priming effect in a perceptual identification task. However, this effect did not interact with the encoding manipulation (solving an anagram vs. reading an item at study). Additionally, the effect of percentage of old test items was a very small one and was significant only when this factor was manipulated within subjects, confounding percentage of old test items with test order. In view of Allen and Jacoby's failure to obtain an interaction between percentage of old test items and encoding condition in the absence of an instructional manipulation, and of the small size of the percentage-of-old-test-items effect, interactions between our encoding variable and our instructional manipulation in perceptual identification were unlikely to result from a direct effect of percentage of old test items. To determine if the percentage-of-old-test-items manipulation could account for the critical interaction, we extracted the variance associated with subject strategy from the data (using questionnaire responses as an index of strategy). Consistent with a strategic account, there was no significant interaction between percentage of old test items and encoding condition when the variance associated with strategy was taken into account.

Our data do not support Allen and Jacoby's (1990) claim that increasing the percentage of old test items in a "yes"/"no" recognition test improves *old/new* discrimination because of a context reinstatement effect. In the current experiment, changing the percentage of old test items influenced only the criterion for emitting a "yes" response, but not discrimination as reflected in d' scores. This result is similar to those obtained in signal detection experiments in which the relative proportions of signal and noise trials are varied (see, e.g., Green & Swets, 1974). The percentage-of-old-test-items variable exerted a larger influence on visual-shallow hit scores than on auditory-deep hit scores or false alarm scores. However, in a signal detection model, this differential effect would be expected because the mean of the distribution of mnemonic information for visual-shallow items fell closer to the criterion than did the means of the distributions for auditory-deep items and new items. Under the assumption of gaussian distributions, shifts in criterion would have the largest effect near the mean of a distribution, where probability density is greatest.

EXPERIMENT 2

Experiment 2 constituted a conceptual replication of the perceptual identification portions of the design used in Experiment 1, with procedural adjustments designed to eliminate ceiling and floor effects, thereby enhancing statistical power. The primary reason for this replication was that overall levels of priming in Experiment 1 were smaller than those observed in many of the experiments in the literature. This discrepancy could have resulted from the fixed exposure duration used in Experiment 1. When a fixed exposure duration is used, subjects display large variations in overall levels of identification performance, and those whose performance is close to ceiling or floor often fail to show differences between old and new test items. A secondary reason for the replication was that we replaced 6 subjects who showed ceiling or floor effects in Experiment 1, leading to the possibility of subject-selection artifacts. A third reason was that the subjects in the intentional retrieval condition did not show the advantage of visual over auditory encoding reported in many other studies of repetition priming in visual tasks (for reviews, see Richardson-Klavehn & Bjork, 1988; Roediger & McDermott, 1993). On the basis of the typical sensitivity of perceptual identification tests to match in surface features between encoding and test, it has often been concluded that such tests are immune to influences of semantic-elaborative information. The present result might therefore be somewhat controversial.

Method

Subjects. Thirty-two male and female undergraduate students participated to fulfill a requirement for an introductory psychology course at the University of California, Los Angeles. They were screened for fluency in English.

Design. The recognition memory test formats were omitted from the design, resulting in 16 unique study/test formats, instead of the 32 in Experiment 1. Two subjects were assigned to each format, for a total of 32 subjects—16 in the intentional retrieval condition and 16 in the standard indirect test condition.

Materials. The length of the list of medium-frequency items used in the practice perceptual identification test was extended from 20 to 40 items. Twenty additional medium-frequency words were selected from Kučera and Francis (1967) according to the criteria used in Experiment 1 and were added to the original 20 practice items. The resulting 40-item list was divided into four blocks of 10 items. Items were assigned to blocks in such a way that the mean frequency of the items in the four blocks was approximately the same.

Apparatus and Procedure. The apparatus and procedure were the same as those in Experiment 1, except in three respects:

1. The “at” symbol mask used for half the perceptual identification subjects in Experiment 1 was used for all the subjects in Experiment 2.

2. Performance on the practice perceptual identification test was monitored in order to estimate an exposure duration for each subject that would produce approximately 30% correct performance on new low-frequency words. The exposure duration for the practice items was reduced after each block of 10 items. It was 64, 48, 40, and 32 msec for the first, second, third, and fourth blocks, respectively. Performance on the first block was not scored, and performance was scored separately for the last three blocks. The exposure duration that produced performance closest to 40% correct on the practice items was selected for the test phase. If two exposure durations resulted in performance equally close to 40%, the shorter of the two was selected. If performance was poorer than 30% correct on the 48-msec practice items, the exposure duration in the test phase was set to 56 msec. In summary, exposure duration in the test proper was either 56, 48, 40, or 32 msec, depending on the subject’s performance level in the practice test.

3. The questionnaire measure administered at the conclusion of the perceptual identification test in Experiment 1 was not included because the purpose of the experiment was to replicate the pattern of data produced by the experimental manipulations in perceptual identification. It was reintroduced in Experiment 3.

Results

Exposure duration was free to vary in Experiment 2 so that there was a potential confounding between exposure duration and between-subject group resulting from sampling error. Although the mean exposure duration for the standard indirect test group ($M = 47.5$ msec) was somewhat longer than the mean for the intentional retrieval group ($M = 44.0$ msec), the difference was not significant [$F(1,30) = 2.71, p > .10$]. Inspecting the correct identification scores of individual subjects revealed that the exposure adjustment procedure was effective: No subject in either group showed ceiling or floor effects.

As in Experiment 1, the subjects were scored as correct only when they responded with the exact letter string presented. Use of a more lenient scoring criterion did not change the data pattern. The mean proportions of critical low-frequency items identified are presented in Table 5. A two-way ANOVA was conducted, with instructions (intentional retrieval vs. standard indirect) and item type (auditory-deep vs. visual-shallow vs. new items) as sources of effect variance. The pattern of significant and nonsignificant effects replicated that in Experiment 1. The main

Table 5
Mean Proportions of Critical Low-Frequency Items
Identified in Experiment 2 as a Function of
Encoding Condition and Test Instructions

Instructions	Old Items		New Items
	Auditory-Deep	Visual-Shallow	
Standard indirect	.33 (.02)	.50 (.19*)	.31
Intentional retrieval	.43 (.11*)	.42 (.10*)	.32

Note—Means in parentheses are priming scores. *Significant priming.

effect of item type was highly significant [$F(2,60) = 14.9, p < .001$], whereas the main effect of instructions was nonsignificant ($F < 1$). Most critically, the interaction of instructions and item type was significant [$F(2,60) = 5.9, p < .01$]. Planned comparisons between old- and new-item means for the standard indirect test group showed significant priming in the visual-shallow condition [$F(1,30) = 34.6, p < .001$] but not in the auditory-deep condition ($F < 1$). By contrast, the intentional retrieval group showed significant priming in both the visual-shallow [$F(1,30) = 9.7, p < .01$] and auditory-deep [$F(1,30) = 8.4, p < .01$] encoding conditions. The difference in baseline (new-item) performance between the groups was nonsignificant ($F < 1$).

An additional 2 (instructions) \times 2 (encoding condition) ANOVA of priming proportions revealed a significant main effect of encoding condition [$F(1,30) = 7.39, p < .01$], but this main effect was qualified by a significant interaction of instructions and encoding condition [$F(1,30) = 10.07, p < .01$]. The main effect of instructions did not approach significance ($F < 1$). Planned comparisons revealed significantly greater priming in the visual-shallow condition than in the auditory-deep condition for the standard indirect group [$F(1,30) = 17.35, p < .001$]. The difference between auditory-deep and visual-shallow encoding was not significant for the intentional retrieval group ($F < 1$).

Pooled analysis of intrusion errors. In view of the different patterns of repetition priming exhibited by the intentional retrieval and standard indirect test groups, we examined the frequency with which the subjects produced studied words as incorrect responses to new low-frequency test items (*intrusion errors*). This analysis was designed to determine whether or not the instructional groups differed in bias to respond with previously encountered items. Many subjects made no intrusions of items encountered prior to the test, so we classified the subjects according to whether or not they made at least one intrusion, and we pooled the error data from Experiments 1 and 2 in order to gain stable estimates. Among the 48 subjects in the standard indirect test condition, 9 made at least one auditory-deep intrusion, and 39 made no intrusions of auditory-deep items. The corresponding frequencies for the 48 subjects in the intentional retrieval condition were 18 and 30. The difference between these frequency distributions was significant [$\chi^2(1) = 4.17, p < .05$]. The proportion of subjects making at least one visual-shallow

intrusion was greater in the intentional retrieval condition than in the standard indirect test condition. However, this trend was nonsignificant.

Data for noncritical items in Experiments 1 and 2. For the noncritical medium- and high-frequency filler items, *old/new* status was of necessity completely confounded with instructions and with the percentage of test items that were old. That is, when these items were old, 75% of test items were old, and intentional retrieval instructions were given; when they were new, 25% of test items were old, and indirect test instructions were given. These data are therefore of limited interest, and we do not report them in full for reasons of brevity. The pattern obtained in perceptual identification was comparable to that reported by Jacoby (1983a) in a similar design and was consistent across experiments: High-frequency words were better identified than were medium-frequency words regardless of *old/new* status, but the advantage of old over new items was larger for medium- than for high-frequency words. As expected, overall identification performance was better for the high- and medium-frequency fillers than for the low-frequency critical items (confirming the rationale for using such items as fillers), but low-frequency critical items showed a larger benefit of prior presentation than did the filler items. In recognition memory (Experiment 1), there was little influence of word frequency on hit rate for filler items, and the expected word-frequency effect was restricted to false alarms, with high-frequency words producing a somewhat higher false-alarm rate than did medium-frequency words. Low-frequency critical items produced a lower false alarm rate than did the high- and medium-frequency filler items, but did not show an advantage over filler items in terms of hit rate.⁶

Discussion

The results of Experiment 2 replicated the perceptual identification results of Experiment 1, despite lower overall performance levels than in Experiment 1. When standard indirect test instructions were given, the subjects showed priming only in the visual-shallow condition and not in the auditory-deep condition. When intentional retrieval instructions were given, the subjects showed priming in both auditory-deep and visual-shallow encoding conditions. In the analysis of priming scores, the interaction between encoding condition and test instructions (marginal in Experiment 1) was significant, probably because of the increased power obtained by calibrating baseline performance for each individual subject. Indirect test instructions resulted in significantly greater priming following visual-shallow encoding than following auditory-deep encoding, whereas intentional retrieval instructions resulted in no significant difference in priming between the encoding conditions. An analysis of intrusion errors pooling the error data from Experiments 1 and 2 showed that the intentional retrieval subjects were significantly more likely than were the indirect test subjects to produce auditory-deep items as incorrect responses to new low-frequency test items. The pattern of intrusion errors

to new items mirrors the pattern found in correct responses to old items.

This intrusion result might suggest that the priming effect shown by intentional retrieval subjects in the auditory-deep condition can be attributed to guessing. We consider this guessing hypothesis implausible because (1) the difference in auditory-deep intrusions between the intentional retrieval and indirect test groups does not seem large enough to account for the difference in priming effects between the groups, and (2) much of the knowledge concerning repetition priming effects in perceptual tasks accumulated in the last 40 years argues against crude guessing interpretations of priming effects (see, e.g., Jacoby & Dallas, 1981). In our view, the intrusion data suggest that intentional retrieval subjects made mental reference to past events in performing the perceptual identification task, using their knowledge of previously encountered items to narrow the range of possible responses to test items. This knowledge would have been more readily available for auditory-deep items than for other items.

EXPERIMENT 3

The validity of the arguments we have made about the role of intentional retrieval depends on the assumption that the differences in the data pattern between our perceptual identification groups reflect differences in the use of intentional retrieval and not a direct influence of the percentage of old test items. The argument that the percentage-of-old-test-items manipulation caused the pattern of effects that we observed in perceptual identification in Experiments 1 and 2 seems unlikely for four reasons:

1. In a design similar to ours, Allen and Jacoby (1990) showed that a manipulation of percentage of old test items did not interact with a manipulation of data-driven versus conceptually driven processing at encoding (reading vs. solving an anagram) when that manipulation was not accompanied by an instructional manipulation. By implication, the interaction obtained in the present studies was probably attributable to the instructional manipulation.⁷

2. The percentage-of-old-test-items variable affected only criterion for a "yes" response in recognition memory. It did not affect *old/new* discrimination overall, nor did it interact with the encoding variable to affect *old/new* discrimination. This result allows us to rule out the hypothesis that memory traces corresponding to old items were differentially available to intentional retrieval processes when many test items were old, as opposed to when few were old.

3. When the variance associated with subject strategy (as indicated by the questionnaire) was extracted from the perceptual identification data in Experiment 1, percentage of old test items did not influence priming. On the other hand, differences in subject strategy were associated with different patterns of priming: The subjects reporting an intentional retrieval strategy showed priming following both visual-shallow and auditory-deep encoding, whereas the subjects who did not report an intentional

retrieval strategy showed priming only following visual-shallow encoding. As our instructional manipulation significantly affected strategy as indexed by the questionnaire, it seems reasonable to argue that the instructional manipulation affected performance via the mediating variable of subject strategy.

4. The perceptual identification subjects given intentional retrieval instructions were significantly more likely than were indirect test subjects to produce auditory-deep items as error responses. This result suggests that the difference between the two groups in the auditory-deep condition reflects an influence of retrieval strategy: Intentional retrieval subjects were more likely than were indirect test subjects to make mental reference to past events.

To confirm the validity of these arguments, in Experiment 3 we held the percentage of old test items constant, and manipulated subject strategy. This experiment was conducted with an English subject population, in contrast to the Californian population used in Experiments 1 and 2. A tachistoscope (as opposed to the computer used in the prior experiments) was used to present stimuli, allowing precise control of exposure duration and viewing distance. There were a number of other procedural variations—most notably, encoding activity was manipulated in a between-list manner, in contrast to the within-list procedure used previously. In view of the procedural differences, we reintroduced the recognition test used in Experiment 1 in order to check the effectiveness of our encoding manipulation. We also reintroduced the questionnaire measure used in Experiment 1 in order to assess the frequency of test awareness and of intentional retrieval strategies in our English subjects. Experiment 3 was therefore expected to provide important information concerning the generalizability of the results across subject populations and experimental procedures.

Method

Subjects. The subjects were 40 male and female student volunteers from City University, London. They were native speakers of English and were paid for their participation.

Design and Materials. The subjects encoded two lists of 25 items each in the first part of the experiment—one in the auditory-deep condition and one in the visual-shallow condition. Each list consisted of 15 critical low-frequency items and 10 medium-frequency fillers. The test list consisted of 100 items. The first 40 of these were medium-frequency fillers which had not been presented at encoding; the remaining 60 items were critical low-frequency words, of which 30 were the critical items presented at encoding and 30 were not previously presented (new items). Test type was manipulated between subjects, two groups receiving a perceptual identification test for all 100 test items, and one group receiving a perceptual identification test for the first 40 test items and a recognition memory test for the remaining 60 items. The perceptual identification groups differed in terms of instructions: One group received standard indirect test instructions, and the other received intentional retrieval instructions. Assignment to testing conditions was random, with 8 subjects assigned to the recognition memory group, and 16 subjects assigned to each of the perceptual identification groups.

The critical low-frequency items were 55 of the 60 low-frequency words (Pools L1 and L2) used in Experiments 1 and 2, plus 5 additional low-frequency items selected from Kučera and Francis (1967) according to the criteria previously described. We replaced 5 of

the words used in the previous experiments (1 in Pool L1 and 4 in Pool L2) because of differences in linguistic usage between England and the United States (e.g., the word *trash* is not commonly used in England), and because one word (*liar*) was discovered to be homophonic (it could be interpreted as *lyre*). The spelling of one item in Pool L1 was changed from the American to the English form (i.e., from *armor* to *armour*). The 40 medium-frequency items used for the practice perceptual identification test in Experiment 2 were used as test items in the current experiment. In addition, we selected 20 items from the medium-frequency experimental items used in Experiments 1 and 2 to serve as study-list fillers. These were randomly assigned into two sets, F1 and F2.

The 30 items in each low-frequency pool (L1 and L2) were randomly assigned into two sets of 15 items, resulting in four pools of critical low-frequency items [L1(A), L1(B), L2(A), and L2(B)]. These four pools were used to create four 25-item study lists. The first 5 and the last 5 items in each list were medium-frequency fillers that were used as buffers to attenuate effects of primacy and recency on critical items. Lists L1(A) and L2(A) used Pool F1 as fillers, whereas lists L1(B) and L2(B) used Pool F2 as fillers.

Order of encoding conditions was balanced across subjects, with half the subjects performing auditory-deep encoding for the first list and visual-shallow encoding for the second list, and half the subjects performing visual-shallow encoding for the first list and auditory-deep encoding for the second list. The subjects either encoded lists L1(A) and L1(B), or lists L2(A) and L2(B), and the order of lists was also varied (A list first vs. B list first). These three binary counterbalancing variables were combined factorially, yielding eight unique encoding formats. Within the recognition test group, 1 subject received each of these formats; within each of the two perceptual identification test groups, 2 subjects received each of these formats. The 100-item test list consisted of the 40 medium-frequency items from the practice perceptual identification test in Experiment 2, followed by the 60 critical low-frequency items from lists L1(A), L1(B), L2(A), and L2(B). A single order of test items was used for all subjects. The order of the critical low-frequency test items was block randomized, with each 4-word block in the last 60 test items containing one item from each of the four study lists. Because of the counterbalancing procedures at encoding, across subjects, each low-frequency test item served as a new item, as an auditory-deep item, and as a visual-shallow item.

Procedure and Apparatus. The subjects were told that they would perform various tasks relating to their knowledge of English words and that the first two tasks involved making a numerical judgment. Appropriate encoding instructions were given prior to the presentation of each of the two study lists. The instructions concerning encoding activity for the auditory-deep and visual-shallow conditions were the same as those in Experiments 1 and 2. In the visual-shallow condition, words were presented on small white cards printed in black uppercase letters. In the auditory-deep condition, the experimenter read the words aloud to the subject. The subjects wrote their numerical ratings on a sheet of paper containing 25 numbered spaces, with a fresh sheet used for each list. The experimenter presented each word, then waited for the subject's written response before presenting the next word in the list. For the auditory-deep condition, the response sheet was headed with a reminder of what the numerical response categories meant (i.e., 1 = *minutes ago*, 2 = *hours ago*, etc.).

Following the encoding phase, all subjects performed a perceptual identification test. Words were presented via an Electronic Developments three-field tachistoscope with an automatic card changer. On each trial, a fixation point (a cross) was presented for 1 sec. The to-be-identified word was then presented briefly. It was immediately replaced by a mask consisting of a string of "at" symbols that appeared for 1 sec. All stimuli were centered and were printed in black ink on white cards, with the words printed in 5-mm-high uppercase letters. At a viewing distance of 508 mm, the character height subtended approximately 0.5° of visual angle

in the vertical plane. The number of degrees of visual angle subtended by words in the horizontal plane varied from approximately 2° for four-letter words to approximately 3° for six-letter words.

Following the presentation of each test item, the subjects responded in written form on a sheet containing numbered spaces. They were told that the task was a difficult one because they would often not be sure of the identity of the word presented. If they had some idea of the word presented, they should write that word down, but if they had no idea, they should put a dash in the space corresponding to that word. The next word would not be presented until they had responded. The first 10 words were described to the subjects as practice items, with an exposure duration of 70 msec for Items 1-5 and 60 msec for Items 6-10. Following presentation of the 10th word, the experimenter asked the subjects whether they understood the task and answered any questions. The subjects were told that the task would become increasingly difficult over the next 30 items. The next 20 test words were then presented at an exposure duration of 50 msec for Items 11-20 and 40 msec for Items 21-30. The experimenter then paused to insert a new deck of cards into the card changer. At this point, different instructions were given depending on between-subject group.

The recognition memory subjects were told that the next 10 trials would be identification trials such as those just performed. However, after those trials, the words would be presented for a longer duration so that reading the items would be easy. The task would then be to judge whether each item had been encountered in the lists encoded in the first part of the experiment. They were to write Y (for "yes") if they believed they had previously encountered the item and N (for "no") if they believed they had not. The perceptual identification subjects in the intentional retrieval condition were told that the next 10 words would be items that they had not encountered before in the experiment but that after those items there would be a mixture of new items and items from the lists encoded in Part 1 of the experiment. They were instructed to try to use their memory for the items encountered in Part 1 to help them identify the test items. The subjects in the standard indirect test condition were told that they would continue to perform the same task for the remaining test items. No mention was made of the presence of old words among the upcoming items.

Items 31-40 were then presented at a duration of 30 msec. The experimenter monitored the performance of the perceptual identification subjects on these items by comparing their written responses to the words on the cards ejected from the card changer. If they correctly identified eight or more of these new medium-frequency words, the exposure duration for Items 41-100 (the critical items) was set at 25 msec; if they identified three or fewer, it was set at 40 msec. Otherwise the exposure duration remained at 30 msec. For the recognition test subjects, the exposure duration for Items 41-100 was set at 1 sec. The intertrial interval for all subjects was approximately 5 sec throughout the test because of the time taken by the card changer. There was a further brief pause after Item 70 had been presented in order for the experimenter to install the remaining cards. Following Item 100, the recognition test subjects were debriefed and paid, whereas the perceptual identification subjects were asked to complete a questionnaire concerning test awareness and strategy similar to the one used in Experiment 1. After they had filled out the questionnaire, they were also debriefed and paid.

Results

Recognition memory data. The mean proportions of hits (based on 15 items per subject per condition) were .97 for the auditory-deep condition and .52 for the visual-shallow condition. The mean proportion of false alarms (based on 30 observations per subject) was .17. As in Experiment 1, the hit and false alarm scores were

used to calculate d' scores on the basis of the distribution of the standard normal variable. The subjects showing a hit proportion of 1.0 for auditory-deep items were assigned a score of .99 in order to allow d' to be defined for that condition. There were no zero scores for new items (in contrast to Experiment 1). A one-way ANOVA of the d' scores showed that *old/new* discrimination was significantly better in the auditory-deep condition ($M = 3.09$) than in the visual-shallow condition [$M = 1.07$, $F(1,7) = 53.67$, $p < .001$]. The mean criterion score was 1.02.

Perceptual identification data. The mean exposure duration for critical low-frequency items was 30.9 msec for the standard indirect test group and 30.0 msec for the intentional retrieval group. These means did not differ significantly in a one-way ANOVA ($F < 1$). The subjects were scored correct only when they wrote down the exact letter string presented; use of a more lenient criterion would not change the data pattern to be reported. The mean proportions of old and new items identified are presented in Table 6. Means for old items are based on 15 items per subject per condition, whereas means for new items are based on 30 items per subject.

The data were initially submitted to a $2 \times 2 \times 3$ ANOVA with instructions (standard indirect vs. intentional retrieval) and order of encoding activity (enclosure counting first vs. recency judgment first) as between-subject factors, and item type (auditory-deep vs. visual-shallow vs. new items) as a within-subject factor. This analysis revealed no significant main effect of order of encoding activity and no significant interactions involving order of encoding activity (all F s < 1). The data presented in Table 6 are therefore averaged over order of encoding activity, and the inferential analyses to be presented next disregard this variable.

A 2 (standard indirect vs. intentional retrieval) $\times 3$ (auditory-deep vs. visual-shallow vs. new items) ANOVA revealed a significant main effect of item type [$F(2,60) = 14.90$, $p < .001$], but this main effect was qualified by a significant instructions \times item type interaction [$F(2,60) = 3.42$, $p < .05$]. Planned comparisons on the means for the standard indirect test group revealed significant facilitation in performance for visual-shallow items compared with new items [$F(1,30) = 13.32$, $p < .001$]. Encoding items in the auditory-deep condition, however, did not result in significant performance facilitation in comparison with new items [$F(1,30) = 2.06$, $p > .15$].

Table 6
Mean Proportions of Critical Low-Frequency Items Identified in Experiment 3 as a Function of Encoding Condition and Test Instructions

Instructions	Old Items		New Items
	Auditory-Deep	Visual-Shallow	
Standard indirect	.40 (.07)	.50 (.17*)	.33
Intentional retrieval	.57 (.20*)	.52 (.15*)	.37

Note—Means in parentheses are priming scores. *Significant priming.

The data pattern across item type for the intentional retrieval group differed from that for the standard indirect group, with significant differences between auditory-deep and new-item means [$F(1,30) = 15.99, p < .001$], and between visual-shallow and new-item means [$F(1,30) = 9.25, p < .01$]. The small difference in baseline (new-item) identification scores between the standard indirect and intentional retrieval groups was nonsignificant ($F < 1$).

An additional 2×2 ANOVA of priming proportions showed a significant interaction of instructions and encoding activity [$F(1,30) = 9.69, p < .01$] but no significant main effects of instructions ($F < 1$) or of encoding activity [$F(1,30) = 1.20, p > .25$]. Planned comparisons revealed that visual-shallow encoding produced significantly greater priming than did auditory-deep encoding for the indirect test group [$F(1,30) = 8.86, p < .01$]. The small advantage of auditory-deep over visual-shallow encoding for the intentional retrieval group was nonsignificant [$F(1,30) = 2.03, p > .10$].

Questionnaire data. In the questionnaire administered after the perceptual identification test, all 16 subjects in the standard indirect test group responded affirmatively to the question asking if they were aware of the presence of old items on the test list. Two of these subjects (12.5%) suspected prior to the perceptual identification test that it would contain old items, and 14 (87.5%) realized at some point during the test that old items were present. Six subjects (37.5%) responded affirmatively to the question asking if they had actively used their memory for old items to help them identify test items; the remaining 10 (62.5%) responded negatively. Both the subjects who suspected the presence of old items fell into the latter group. The subjects in the intentional retrieval group were all test-aware by virtue of the instructions. Of these 16 subjects, 15 (93.7%) reported that they had used their memory for old items in performing the test, and 1 (6.3%) reported that he/she had not (contrary to instructions). The relationship between instructions and reported test strategy (used memory vs. did not use memory) was significant [$\chi^2(1) = 11.22, p < .01$].

Nonparametric analysis of ordinal arrangement of population means. The results of the parametric analyses presented thus far strongly suggest that our encoding variable produced opposite effects on recognition memory and on priming in an indirect test of perceptual identification. The situation with respect to perceptual identification with

intentional retrieval instructions is more difficult to evaluate. The analyses of priming effects in the present experiment and in Experiment 2 showed that the influence of our encoding variable on priming varied significantly depending on instructions on perceptual identification; furthermore, in all three experiments, the intentional retrieval subjects showed priming in both auditory-deep and visual-shallow encoding conditions, whereas the standard indirect test subjects showed significant priming only in the visual-shallow condition. Despite these differences in the data pattern between perceptual identification groups, the ordinal arrangement of population means for the two encoding conditions could conceivably be the same for both groups. That is, one could claim that the disadvantage of auditory-deep encoding relative to visual-shallow encoding was simply smaller (in the population) in the intentional retrieval condition than in the standard indirect condition, leading to the observed encoding/retrieval interaction. The small observed advantage of auditory-deep over visual-shallow encoding for the intentional retrieval condition could then be attributed to sampling error, because that advantage—although consistent across experiments—was not significant in any one experiment. The claim that giving intentional retrieval instructions in perceptual identification removed a crossed double dissociation between recognition memory and perceptual identification might, therefore, be false in terms of the underlying population means.

We evaluated this skeptical claim statistically by classifying each subject as to whether she/he showed an advantage of auditory-deep over visual-shallow encoding, an advantage of visual-shallow over auditory-deep encoding, or no difference between encoding conditions. We pooled the data across experiments for this analysis, in view of the similarity of the designs and of the data patterns. The resulting frequencies are shown in Table 7. Frequencies for the recognition memory subjects are based on d' scores from Experiments 1 and 3 (a total of 72 subjects), and disregarded the percentage-of-old-test-items manipulation in Experiment 1, because that manipulation did not affect d' . Frequencies for the perceptual identification subjects are based on priming proportions from all three experiments (a total of 64 subjects per instructional group). We conducted three sign tests, one on the data from each test condition. In view of the large number of cases in each test, we used the normal approximation to

Table 7
Frequency of Ordinal Arrangements of Scores for the Two Encoding Conditions in Experiments 1-3 as a Function of Test Type and Instructions

Test Type (Instructions)	Ordinal Arrangement of Scores		
	Auditory-Deep > Visual-Shallow	Visual-Shallow > Auditory-Deep	Auditory-Deep = Visual-Shallow
Recognition memory	72	00	00
Perceptual identification (Standard indirect)	11	39	14
Perceptual identification (Intentional retrieval)	34	17	13

the binomial distribution with the standard continuity correction.

In recognition memory, there was a highly significant advantage of auditory-deep over visual-shallow encoding, with all subjects showing the same ordinal arrangement of scores ($z = 8.37$, $p < .001$, two-tailed). In perceptual identification, the subjects showing equal performance in the two encoding conditions were dropped from the analysis, as prescribed by Siegel (1956) and Hays (1973). The percentage of subjects showing no difference between conditions was similar in the standard indirect (21.9%) and intentional retrieval (20.3%) test groups, so that any difference in the results of the sign tests between test groups could not be an artifact of dropping these subjects. Consistent with the results of the parametric analyses in Experiments 2 and 3, indirect test instructions in perceptual identification resulted in a highly significant advantage of visual-shallow over auditory-deep encoding ($z = 3.81$, $p < .001$, two-tailed). Most critically, however, across experiments intentional retrieval instructions in perceptual identification resulted in a significant advantage of auditory-deep over visual-shallow encoding ($z = 2.24$, $p = .025$, two-tailed). This latter result permits us to rule out the hypothesis that the ordinal arrangement of the population means for the two encoding conditions was the same regardless of whether indirect or intentional retrieval instructions were given in perceptual identification.

Discussion

Experiment 3 produced a data pattern closely comparable to that in Experiments 1 and 2. The auditory recency judgment task at encoding led to excellent recognition memory performance, whereas the visual enclosure counting task led to rather poor recognition memory performance. Despite the high quality of the mnemonic information resulting from the recency judgment task, encoding items in the auditory-deep condition did not result in significant priming when a standard indirect test of perceptual identification was given. Under these test conditions, only visual-shallow encoding resulted in significant facilitation of later perceptual identification. When the subjects were given instructions to use their memory for previously encountered items to help them identify test items, a different pattern was observed. Significant priming resulted from encoding items in both the auditory-deep and visual-shallow conditions. In the analysis of priming proportions, this difference in the pattern of priming between instructional groups emerged as a significant interaction between instructions and encoding condition (as in Experiment 2).

As in Experiment 1, awareness that the perceptual identification test contained old items was common among indirect test subjects. However, according to the posttest questionnaire, test awareness was not necessarily accompanied by spontaneous adoption of an intentional retrieval strategy. The percentage of indirect test subjects reporting an intentional retrieval strategy in the current experiment (37.5%) was similar to that in Experiment 1 (40.6%).

The perceptual identification portions of Experiment 3 meet the strictest form of the retrieval intentionality cri-

terion proposed by Schacter et al. (1989). The physical test cues presented to the two groups were identical (i.e., rapidly flashed items), as was the percentage of old test items. The current experiment replicated the pattern of effects observed in Experiments 1 and 2 despite the fact that the percentage of old test items was held constant across the two instructional conditions in perceptual identification. This result permits us to rule out conclusively the hypothesis that the percentage-of-old-test-items manipulation in Experiments 1 and 2 (which was deliberately confounded with instructions) caused the differences in the pattern of priming between the groups. The consistency of the data pattern across experiments is particularly impressive in view of the considerable procedural differences between the current experiment and Experiments 1 and 2, and the difference in subject populations.

Sign tests, pooling data across experiments, confirmed that the ordinal arrangement of population means for the two encoding conditions differed between recognition memory and a standard indirect test of perceptual identification: Although auditory-deep encoding was superior to visual-shallow encoding in recognition memory, visual-shallow encoding produced better perceptual identification performance than did auditory-deep encoding. This reversal in the ordinal arrangement of population means across tests (as a function of the same encoding variable) represents a crossed double dissociation between tests. A sign test on the pooled data for the subjects given intentional retrieval instructions in perceptual identification revealed that auditory-deep encoding produced better performance than did visual-shallow encoding. This difference between encoding conditions, although quite consistent across subjects, was not large enough to be significant at the level of the individual experiment.

GENERAL DISCUSSION

Retrieval Intention and Crossed Double Dissociations

Crossed double dissociations between direct and indirect measures of memory (see, e.g., Jacoby, 1983b; Winnick & Daniel, 1970) have been taken as strong evidence that priming in indirect tests does not rely on intentional retrieval (see, e.g., Allen & Jacoby, 1990; Jacoby, 1983b). This conclusion is based on two assumptions: (1) that direct test performance relies on intentional retrieval (and that manipulations that influence direct test performance do so by influencing the success of intentional retrieval attempts), and (2) that a negative relationship between tests as a function of a variable implies that one of the tests relies on intentional retrieval, whereas the other does not. Although the former assumption seems reasonable, the latter conclusion is not required by the data, because the direct and indirect tests usually involve different physical retrieval cues. Intentional retrieval could be common to both direct and indirect tests, with the negative relationship between tests being attributable to an interaction between the processing demands of the test and the type of information encoded. Negative associations

of this type have frequently been demonstrated between direct tests. For example, (1) word frequency can influence recall and recognition in opposite ways (for a summary, see Gillund & Shiffrin, 1984), (2) meaning- and appearance-based recognition tests are negatively related as a function of depth of processing at encoding (Stein, 1978), (3) free recall and recall with cues that visually resemble studied items are negatively related as a function of a generate/read manipulation at encoding (Blaxton, 1989), and (4) free recall and cued recall with word-stem and word-fragment cues are negatively related as a function of a picture/word manipulation at encoding (Weldon, Roediger, & Challis, 1989). However, few would argue that performance in one of the pair of tests compared does not rely on intentional retrieval. Negative associations have also been found between pairs of indirect tests (for a summary, see Roediger, Srinivas, & Weldon, 1989), yet these dissociations are not used to argue that performance in one of the indirect tests relies on intentional retrieval. In sum, crossed double dissociations between direct and indirect tests that present different physical cues do not, taken in isolation, show conclusively that indirect test performance does not depend on intentional retrieval.

Researchers may have accepted that such dissociations reflect differences in retrieval intention because they have implicitly made a *transparency assumption* concerning the relationship between processes and tasks (Dunn & Kirsner, 1989). In the case of indirect tests, the transparency assumption is that intentional retrieval is not involved in facilitation effects because the instructions do not refer to a prior event. This assumption is incorporated in the common practice of referring to indirect tests as *implicit memory tests* (Graf & Schacter, 1985), because the term *implicit memory* is also used to refer to a hypothetical involuntary memory process underlying task performance. When processes are equated with tasks in this manner, however, dissociations between direct and indirect tests are redundant as evidence for involuntary influences of memory: The hypothesis that the tests reveal different forms of memory has been accepted a priori. In addition, the transparency assumption with respect to indirect tests has been increasingly called into question in recent years because of evidence that indirect measures sometimes show influences of intentional retrieval (see, e.g., Challis & Brodbeck, 1992; Jacoby, 1991; Jacoby et al., 1993; Merickle & Reingold, 1991; Richardson-Klavehn & Bjork, 1988; Schacter et al., 1989).

Richardson-Klavehn and Bjork (1988) adopted Johnson and Hasher's (1987) use of the term *indirect test* in order to avoid the type of transparency assumption involved in the term *implicit test*. Their intention was that tests should be classified solely with respect to observables (i.e., instructions and measurement criteria), with no a priori assumptions concerning the processes and states of awareness involved in a particular test. Under this task classification, it is not a contradiction to suggest that performance in an indirect test reflects intentional retrieval of past events, because subjects may spontaneously adopt

intentional retrieval strategies. In consequence, whether or not indirect test performance reflects intentional retrieval must be determined from the data.

In the current experiments, we examined whether a crossed double dissociation between a direct and an indirect test could be attributed to a difference in retrieval intentionality between the tests, even though the tests involved different physical retrieval cues. We found that a crossed double dissociation between recognition memory and perceptual identification as a function of an encoding variable disappeared when the perceptual identification subjects were given intentional retrieval instructions. At a qualitative level of analysis, this result strongly suggests that the observed crossed double dissociation reflected a difference in retrieval intentionality, because it could not have occurred if a majority of the indirect test subjects had spontaneously employed an intentional retrieval strategy.

At a formal level of analysis, the pattern of data across the three experiments reported here demonstrates a *reversed association* (Dunn & Kirsner, 1988, 1989) between recognition memory and perceptual identification performance. When intentional retrieval instructions were given in perceptual identification, perceptual identification priming could be expressed as a monotonically increasing function of discrimination performance in recognition memory. Such a relationship could be accommodated by a model assuming that performance in both tasks was monotonically related to the efficiency of a single underlying process—in this case, intentional retrieval—together with the assumption that perceptual identification is a less sensitive measure than is recognition memory. On the other hand, when indirect test instructions were given in perceptual identification, a negative association between the tests was observed. This negative relationship could not be predicted by the intentional retrieval model: If recognition memory and perceptual identification priming were both monotonic functions of an underlying intentional retrieval process, the rates of change of these functions with respect to the underlying process could never change sign, so a negative association could not be observed. The logic of reversed association requires that both monotonically positive and monotonically negative relationships be observed between the same pair of tests. The data presented here show both types of relationships and therefore permit formal rejection of the hypothesis that recognition memory and perceptual identification priming depend on a single underlying intentional retrieval process.⁸

The reversed-association logic permits an inference to be made about differences in retrieval intentionality between direct and indirect tests, even when those tests present different physical cues to the subject (as in the present study). This inference follows because the difference between the cues in recognition memory and the cues in perceptual identification was held constant, whereas the association between the tests (monotonically positive vs. monotonically negative) varied as a function of the instructional manipulation. The presence or absence of a double dissociation between the tests cannot, therefore, be

attributed to a difference in test cues. In this sense, the reversed-association method of making inferences about retrieval intentionality is more powerful than the method prescribed by the *retrieval intentionality criterion* (Schacter et al., 1989), which requires that test cues be held constant. The current method permits one to verify that dissociations between direct and indirect tests that present different physical cues are not artifactual results of a confounding of physical test cues with direct/indirect status.

The reasoning presented here has considerable similarities in the domain of manipulated variables to the reasoning behind the *method of triangulation* (Hayman & Tulving, 1989a, 1989b), which is used to make inferences about stochastic relationships between tests. In the method of triangulation, performance in a direct and an indirect memory test is compared to performance in a previously administered reference test. If the stochastic relationship to the reference test varies as a function of direct/indirect status, this difference cannot be accounted for by interest biases, because those biases will be constant for the direct and indirect tests. In the current experiments, recognition memory can be seen as the "reference" test. Perceptual identification was compared to this reference under two different instructional conditions. The difference in the test cues between recognition memory and perceptual identification can be seen as creating a bias toward functional independence between the tests. However, this bias would have been constant for the two instructional conditions in perceptual identification and could not account for the way in which the relationship of perceptual identification to recognition memory varied as a function of instructions.

Awareness and Retrieval Intentionality in Indirect Tests

The present questionnaire data suggest that a complete absence of awareness of the presence of old items in an indirect test is rare among normal subjects. Among the 32 indirect test subjects in Experiment 1, only 2 were unaware of the relationship between the encoding and test phases of the experiment, even though only 25% of the test items were old. In Experiment 3, in which 50% of the critical test items were old, none of our 16 subjects were test-unaware, even though the 60 critical items were preceded by 40 items that had not been previously encountered. These results differ from those of Bowers and Schacter (1990), who found higher proportions of test-unaware subjects. In their Experiment 1, for example, 20 of 40 indirect test subjects were test-unaware. However, these researchers went to great lengths to disguise the relationship between the encoding and test phases of their experiment by administering the stem-completion test as one of a number of other tasks. The procedures used in the current experiment are more typical of studies in the literature using indirect tests.

Although awareness that the test contained old items was ubiquitous among the indirect test subjects in the

present experiments, it was not necessarily accompanied by the use of intentional retrieval. When the perceptual identification test instructions made no reference to prior events, a crossed double dissociation between perceptual identification priming and recognition memory was observed, despite the fact that the vast majority of subjects were test-aware. Given the absence of such a dissociation for the intentional retrieval condition, the dissociation from recognition memory in the indirect test condition clearly could not have occurred unless the majority of subjects in that condition failed to adopt an intentional retrieval strategy. The subjects' questionnaire responses in Experiments 1 and 3 are consistent with this reasoning: Less than half of the indirect test subjects reported adopting an intentional retrieval strategy to help them identify test items.

These considerations suggest that dissociations between direct and indirect measures can be ascribed to differences in retrieval intention across tests but not necessarily to differences in subjective awareness of past events. Such a conclusion stands in contrast to the one that might initially be drawn from Bowers and Schacter's (1990) study (data from test-uninformed subjects). They found that test-aware subjects showed a significant depth-of-processing effect in stem completion priming similar to that found in direct tests, whereas test-unaware subjects did not show an effect. An uncritical examination of these data would suggest that test-awareness/test-unawareness is the critical factor determining whether dissociations will be observed. However, Bowers and Schacter did not ask their test-aware subjects whether or not they had adopted an intentional retrieval strategy. It is therefore possible that a subset of their test-aware subjects—having either predicted or noticed the presence of old items—spontaneously employed an intentional retrieval strategy, leading to a depth-of-processing effect. In accordance with this hypothesis, Richardson-Klavehn et al. (1994) found no significant effect of depth of processing on priming in an indirect stem completion test, even though all 24 subjects in that test condition were test-aware. None of these subjects indicated adopting an intentional retrieval strategy in the posttest interview.

Also consistent with this hypothesis are the data from Bowers and Schacter's (1990) test-informed subjects. These subjects were told prior to the test that stems corresponding to studied items would be present (making them test-aware in virtue of the instructions) but were told to complete the stems with the first item coming to mind despite that knowledge. These subjects also showed no depth-of-processing effect. In sum, the data from the current study, from Bowers and Schacter (1990), and from Richardson-Klavehn et al. (1994), yield an essentially consistent picture: Dissociations between direct and indirect tests can be created by the absence of an intentional retrieval strategy in the indirect test. The absence of an intentional retrieval strategy in the indirect test, however, should not be assumed to imply absence of subjective awareness of past events.

The distinction between retrieval intention and subjective awareness of past events has not been widely made in discussions of the relationship between direct and indirect memory measures (for an exception, see Schacter et al., 1989). A single dichotomy between forms of memory underlying test performance is generally assumed. For example, Jacoby and Witherspoon (1982) distinguished between *memory with awareness* and *memory without awareness*, Graf and Schacter (1985) distinguished between *explicit memory* and *implicit memory*, and Hintzman (1990) and Jacoby (1991; see also Jacoby et al., 1993) distinguished between *conscious memory* and *unconscious memory*. These dichotomies share the disadvantage that they conflate differences in retrieval intentionality with differences in subjective awareness of past events. They therefore overlook the possibility that intentional retrieval, although often accompanied by subjective awareness of a past event, may sometimes occur in the absence of such awareness (e.g., when subjects rely on reconstructive processes to retrieve information about a prior event). More important in the current context, these dichotomies also obscure the possibility that subjective awareness of a prior event can occur without an intention to retrieve that event. This type of reminding process occurs commonly in everyday life when reinstatement of certain environmental cues triggers the reexperiencing of past events associated with those cues (see, e.g., Ebbinghaus, 1885/1964). To accommodate these possibilities, Richardson-Klavehn et al. (1994) proposed that the terms *intentional retrieval* and *involuntary retrieval* be used to capture differences in retrieval intentionality and that the terms *conscious memory* and *unconscious memory* be reserved to distinguish between influences of memory that occur with and without a subjective awareness of the past.

The current questionnaire data suggest that involuntary conscious memory occurs frequently in the performance of indirect tests. Of the 46 test-aware subjects in the indirect test groups in Experiments 1 and 3, only 7 indicated that they suspected prior to the test that it would contain old items. These subjects presumably "second-guessed" the purpose of the experiment (although surprisingly only 1 of these 7 reported using intentional retrieval). The remaining 39 indicated that they had realized that old items were present at some time during the test. Test-awareness for these subjects must have occurred spontaneously because no mention was made in the instructions of the presence of old items. Some of these subjects indicated adopting an intentional retrieval strategy, but it must be noted that spontaneous awareness of the presence of previously presented items would have had to occur *prior* to the adoption of such a strategy.

The role of involuntary conscious memory in indirect test performance is unclear. It apparently leads some subjects to adopt an intentional retrieval strategy. However, even if it does not do so, it could conceivably play a causal role in performance, operating as a type of reminding process. This possibility was entertained by Schacter et al.

(1989) in attempting to explain why enhancement of priming effects due to reinstatement of a new associate occurred in test-aware but not in test-unaware subjects in Bowers and Schacter's (1990) studies. They argued that

the phenomenon referred to as "implicit memory for new associations" might be more properly characterized as unintentional or involuntary explicit memory for new associations—that is, associative effects on stem completion may be observed only when subjects are explicitly (though unintentionally) reminded of the occurrence of a target pair on the study list. (p. 62)

On the other hand, involuntary conscious memory may simply be a conscious epiphenomenon, correlated with—but not causally involved in—the tendency to produce old items in an indirect test. Distinguishing between these possibilities is an issue for future research.

Acknowledging that involuntary conscious memory may accompany priming effects in normal subjects, Richardson-Klavehn et al. (1994) argued that the *method of opposition* developed by Jacoby et al. (1989) is better suited than is indirect test methodology for studying influences of memory that are not only involuntary but also unconscious. Applying the method of opposition to repetition priming measures involves instructing subjects to omit from their responses items that are associated with an awareness of the past. Richardson-Klavehn et al. showed that normal subjects can display a facilitatory priming effect in a stem-completion task under these conditions, although they are completely unable to recognize the old items that they produced. Comparing opposition test performance with the performance of indirect test subjects—under conditions in which a contribution of intentional retrieval to indirect test performance can be ruled out—permits a quantitative estimate of the extent to which indirect test performance is associated with involuntary conscious memory.

Making such a comparison between opposition and indirect tests assumes that it is possible to determine whether indirect test performance reflects intentional or involuntary retrieval processes. Researchers now have a number of techniques at their disposal in this regard. The retrieval intentionality criterion (Schacter et al., 1989) represents a powerful tool when retrieval cues can be held constant across tests, whereas the method put forward in the current report, as well as the method of triangulation (Hayman & Tulving, 1989a, 1989b), are useful when test cues cannot be held constant. In addition, the data from Experiment 1 indicate that questionnaire measures of subject strategy, despite their obvious potential shortcomings, may have some promise as a diagnostic tool. In that experiment, responses to the questionnaire item concerning intentional retrieval were correlated with the pattern of priming observed. When the variance associated with self-reported strategy was extracted from the perceptual identification data, there was no effect of the instructional

manipulation on the pattern of priming. This result suggests that strategy was a variable mediating the effect of instructions on the pattern of priming and that strategy can be indexed, albeit crudely, by a self-report measure.

Conclusions and Comments

On the basis of the data from the present experiments, we make two main arguments:

1. *Inferences can be made about differences in retrieval intentionality between direct and indirect tests even when those tests present different physical cues to the subject.* Schacter et al. (1989) argued that direct/indirect dissociations only represent convincing evidence for differences in retrieval intentionality between tests when retrieval cues are held constant, and a dissociation between tests is observed as a function of instructions alone (the retrieval intentionality criterion). We argue here that the retrieval intentionality criterion, although an extremely useful tool, is unnecessarily restrictive in requiring that test cues be held constant. We found that instructions to use an intentional retrieval strategy in perceptual identification removed a crossed double dissociation from recognition memory that was present under standard indirect test conditions. By implication, the observed dissociation must reflect a difference in retrieval intentionality between recognition and perceptual identification, and could not be an artifact of the confounding of instructions with the difference in test cues.

In the present report, we have emphasized the replicability of these results and the generalizability of the results across procedures, subject populations, and overall performance levels, rather than the generalizability of the results to different tests and different encoding variables. It would not, therefore, be appropriate to assume that all of the superficially similar crossed double dissociations between direct and indirect tests in the literature reflect differences in retrieval intentionality between tests. In fact, the logic that we have used here suggests that some direct/indirect dissociations reported in the literature *cannot* conclusively be attributed to a difference in retrieval intentionality. For example, Weldon et al. (1989) compared free recall (direct test) with word-fragment and word-stem completion in a design that parallels that of the current experiments. In the completion tasks, subjects either received standard indirect test instructions or were instructed to use the word stems and fragments as cues to retrieve studied items. The encoding variable consisted of presenting items either as pictures or as words. Pictures produced better performance than did words in the free-recall test, but words produced greater priming than did pictures in the completion tasks when indirect test instructions were given. The current logic suggests that this crossed double dissociation between free recall and indirect completion performance can be attributed to a difference in retrieval intentionality only if this dissociation was eliminated when intentional retrieval instructions were given in the completion tasks. However, in this study, words produced better performance in the completion

tasks than did pictures, even when intentional retrieval instructions were given. Weldon et al. interpreted their data as showing that the dissociation between the free-recall and the indirect completion tests was attributable to the difference in the retrieval cues provided by the two types of tests. We suggest, in addition, that the data pattern reported by Weldon et al. does not permit one to reject the hypothesis that priming in their indirect tests depended on intentional retrieval processes. The observed dissociation, therefore, provides useful information about the distinction between conceptually driven and data-driven tests but not necessarily about the distinction between intentional and involuntary memory.

2. *Dissociations between direct and indirect tests in normal subjects can be ascribed to intertest differences in retrieval intentionality, but not necessarily to intertest differences in awareness of the encoding event.* This argument is based on the finding that a crossed double dissociation between a direct and an indirect test was obtained, but that the vast majority of indirect-test subjects reported awareness that the test contained previously encountered items. Despite this awareness, the majority of indirect test subjects could not have spontaneously adopted a strategy of using their memory for studied items—or, if they did, they failed to apply such a strategy to a majority of test items—because the adoption of such a strategy by most of the subjects on most of the test items would have led to the “undoing” of the dissociation (by the argument given in 1. above).

The current study used a self-report method to assess test awareness, and such methods are inevitably subject to question with respect to their reliability and validity (for more extensive discussion, see Bowers & Schacter, 1990). In particular, it might be claimed that subjects who were test-unaware while performing the test became aware of the presence of previously encountered items at the time that they answered the questionnaire item concerning the presence of previously encountered items. This hypothesis cannot be conclusively ruled out. However, we find it implausible, given that indirect test subjects in Experiment 3 (while being debriefed) had no difficulty in indicating on their test protocol at what point in the test they first noticed that previously encountered items were present. More pressing is the problem that subjects will classify themselves as test-aware, even if only *one* test item is noticed as being previously encountered (Bowers & Schacter, 1990). It then remains possible that many instances in which performance is facilitated by prior exposure are not associated with conscious awareness of the encoding event, although the subject qualifies as test-aware. It is, therefore, important to distinguish global awareness that the test bears a relationship to prior events from awareness of prior encoding events on an item-by-item basis.

The first author and colleagues have attempted to address the latter item-by-item kind of awareness by using opposition methods (Richardson-Klavehn et al., 1994). Clearly, the current self-report data only address the issue of global awareness. We take these data as supporting the

minimal conclusion that test-awareness, in the global sense, occurs frequently in indirect test subjects but does not always prompt intentional retrieval strategies. The exact conditions under which global test-awareness in indirect tests does and does not prompt an intentional retrieval strategy remain to be delineated. It is, however, apparent that attempts to "disguise" indirect tests—the intention being that subjects do not realize that their memory is being tested—may often not be as successful as experimenters have hoped. A more realistic approach may be to "come clean" and inform subjects that the test contains previously encountered items but to request that they nevertheless do not adopt a recollective strategy (as did Bowers & Schacter, 1990, in their test-informed condition). Such an approach follows if one recognizes that tests of repetition priming (and related tests) in normal subjects are useful indices of involuntary memory, but not necessarily of unconscious memory.

REFERENCES

- ALLEN, S. W., & JACOBY, L. L. (1990). Reinstating study context produces unconscious influences of memory. *Memory & Cognition*, **18**, 270-278.
- BLAXTON, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 657-668.
- BOWERS, J. S., & SCHACTER, D. L. (1990). Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 404-416.
- 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.
- COHEN, N. J., & SQUIRE, L. R. (1980). Preserved learning and retention of pattern analyzing skills in amnesics: Dissociation of knowing how and knowing that. *Science*, **210**, 207-210.
- DUNN, J. C., & KIRSNER, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, **95**, 91-101.
- DUNN, J. C., & KIRSNER, K. (1989). Implicit memory: Task or process? In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 17-31). Hillsdale, NJ: Erlbaum.
- EBBINGHAUS, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. (Original work published 1885)
- EICH, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition*, **12**, 105-111.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.
- GRAF, P., & MANDLER, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning & Verbal Behavior*, **23**, 553-568.
- GRAF, P., & SCHACTER, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 501-518.
- GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics*. New York: Krieger.
- HAMANN, S. B. (1990). Level-of-processing effects in conceptually driven implicit tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 970-977.
- HAYMAN, C. A. G., & TULVING, E. (1989a). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 228-240.
- HAYMAN, C. A. G., & TULVING, E. (1989b). Is priming in fragment completion based on a "traceless" memory system? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 941-956.
- HAYS, W. L. (1973). *Statistics for the social sciences*. New York: Holt, Rinehart, and Winston.
- HINTZMAN, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology*, **41**, 109-139.
- JACOBY, L. L. (1983a). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 21-38.
- JACOBY, L. L. (1983b). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior*, **22**, 485-508.
- JACOBY, L. L. (1991). A process dissociation framework: Separating automatic and intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.
- JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.
- JACOBY, L. L., TOTH, J. P., & YONELINAS, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, **122**, 139-154.
- JACOBY, L. L., & WITHERSPOON, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, **36**, 300-324.
- JACOBY, L. L., WOLOSZYN, V., & KELLEY, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, **118**, 115-125.
- JAVA, R. I. (1994). States of awareness following word stem completion. *European Journal of Cognitive Psychology*, **6**, 77-92.
- JOHNSON, M. K., & HASHER, L. (1987). Human learning and memory. *Annual Review of Psychology*, **38**, 631-668.
- KUČERA, H., & FRANCIS, W. (1967). *Computational handbook of present-day American English*. Providence, RI: Brown University Press.
- KUNST-WILSON, W. R., & ZAJONC, R. B. (1980). Affective discrimination of stimuli that are not recognized. *Science*, **207**, 557-558.
- MACLEOD, C. M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 13-21.
- MERICKLE, P. M., & REINGOLD, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 224-233.
- MOSCOVITCH, M., WINOCCUR, G., & MACLACHLAN, D. (1986). Memory as assessed by recognition and reading time in normal and memory-impaired people with Alzheimer's disease and other neurological disorders. *Journal of Experimental Psychology: General*, **115**, 331-347.
- RICHARDSON-KLAVEHN, A., & BJORK, R. A. (1988). Measures of memory. *Annual Review of Psychology*, **39**, 475-543.
- RICHARDSON-KLAVEHN, A., GARDINER, J. M., & JAVA, R. I. (1994). Involuntary conscious memory and the method of opposition. *Memory*, **2**, 1-29.
- ROEDIGER, H. L., III (1990). Implicit memory: Retention without remembering. *American Psychologist*, **45**, 1043-1056.
- ROEDIGER, H. L., III, & MCDERMOTT, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63-131). Amsterdam: Elsevier.
- ROEDIGER, H. L., III, SRINIVAS, K., & WELDON, M. S. (1989). Dissociations between implicit measures of retention. In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 67-84). Hillsdale, NJ: Erlbaum.
- ROEDIGER, H. L., III, WELDON, M. S., & CHALLIS, B. H. (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 honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.
- ROEDIGER, H. L., III, WELDON, M. S., STADLER, M. L., & RIEGLER, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1251-1269.
- SCHACTER, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 501-518.
- 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.
- SCHACTER, D. L., & GRAF, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 432-444.
- SHIMAMURA, A. P., & SQUIRE, L. R. (1984). Paired-associate learning and priming effects in amnesia: A neuropsychological approach. *Journal of Experimental Psychology: General*, **113**, 556-570.
- SIEGEL, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- SLOMAN, S. A., HAYMAN, C. A. G., OHTA, N., LAW, J., & TULVING, E. (1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 223-239.
- SQUIRE, L. R., SHIMAMURA, A. P., & GRAF, P. (1987). Strength and duration of priming effects in normal subjects and amnesic patients. *Neuropsychologia*, **25**, 195-210.
- RINIVAS, K., & ROEDIGER, H. L., III (1990). Classifying implicit memory tests: Category association and anagram solution. *Journal of Memory & Language*, **29**, 389-412.
- STEIN, B. S. (1978). Depth of processing reexamined: The effects of precision of encoding and test appropriateness. *Journal of Verbal Learning & Verbal Behavior*, **17**, 165-174.
- WELDON, M. S., & ROEDIGER, H. L., III (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition*, **15**, 269-280.
- 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.
- WINNICK, W. A., & DANIEL, S. A. (1970). Two kinds of response priming in tachistoscopic recognition. *Journal of Experimental Psychology*, **84**, 74-81.
3. The program used for Experiments 1 and 2 was a general purpose experimental package written for the last author by a professional programmer. The programmer made the change in the mask without consultation as one of a number of updates. We reverted to the previous version of the program for Experiment 2.
4. Unlike analysis of variance, the signal detection model does not assume normality of raw hit and false alarm scores; it does, however, assume that the distributions of mnemonic information for old and new items that underlie subjects' responses are normal and of equal variance. In view of the demonstrable nonnormality of the current data, making the latter pair of assumptions seemed preferable to making the former.
5. One of our reviewers objected to our use of the term *priming* with reference to the intentional retrieval condition. We disagree with this objection because subjects in that condition were instructed to use their memory to *help them* identify test items and—unlike in direct data-driven tests—were not asked to use the flashed items *as cues for the retrieval only of studied items*, and because the term *priming* has lost its original connotation of the temporary sensitization of a preexisting abstract memory representation. We use the term *priming* to refer to facilitation effects, without a priori assumptions about the representations, processes, strategies, or states of awareness involved.
6. The comparisons between low-frequency critical items and medium- and high-frequency filler items were made using the data for old low-frequency items from subjects for whom 75% of test items were old and the data for new low-frequency items from subjects for whom 25% of test items were old.
7. Of the two experiments reported by Allen and Jacoby (1990), only the second showed a significant effect of percentage of old test items. That experiment showed a marginal interaction such that percentage of old test items exerted a greater effect on read items (data-driven encoding) than on anagram items (conceptually driven encoding). The fact that this interaction was opposite in form to the one reported here strengthens the argument that the interaction observed in Experiments 1 and 2 was attributable to our instructional manipulation.
8. The assumption of monotonicity of functions relating task performance to underlying processes is fundamental to the logic of reversed association. If this assumption is not accepted, the analysis presented here fails. However, as Dunn and Kirsner (1988) point out, this assumption is essential to psychological measurement in general.

NOTES

1. Schacter (1987) referred to this spontaneous awareness of the past as *involuntary explicit memory*. We prefer the term *involuntary conscious memory* because the term *explicit test* is often used to refer to memory tests involving intentional retrieval.

2. Because of the refresh rate of the monitor, there was unavoidable random variability in the presentation duration between trials. Exposure durations given here are therefore nominal figures.

(Manuscript received July 8, 1991;
revision accepted for publication August 3, 1993.)