

## Trace loss and the recognition failure of unrecalled words

IAN BEGG

McMaster University, Hamilton, Ontario L8S 4K1, Canada

This paper addresses the phenomenon of recognition failure from the perspective of a theory in which recognition and recall are assumed to involve independent retrieval processes. However, even given independent retrieval, measures of recognition and recall success will covary if any traces are lost from storage, simply because such traces are unavailable for any memory test. In support of the theory, rote learning produced higher covariation between recognition and recall (i.e., fewer recognition failures) than did meaningful elaboration during study. Further, recognition and recall were approximately independent of each other with meaningful elaboration and imagery encoding, regardless of whether the latter involved interactive or separate images. The results of three experiments are discussed in terms of the present "vandal" theory and other theories of recognition failure.

This article addresses two interrelated questions regarding the human memory system. The first question concerns memory storage at a general level: Does all forgetting result from failure to access traces that could be recovered with appropriate cues, or are some traces actually lost from storage? The second question more specifically concerns memory retrieval: Are recall and recognition accomplished by independent retrieval processes, or are the processes dependent in the sense that the recognition process is a component of the recall process? The central argument of the paper is that recall and recognition are independent processes for gaining access with the same memory information; however, there will be some degree of interdependence between measures of recall and recognition success simply because both processes must fail in the case of information that has been lost from storage. Before rationalizing the argument in detail, let us review prior research concerning the relation between recognition and recall.

The usual procedure for studying the relation between the two measures of retention is the "recognition failure" paradigm (e.g., Flexser & Tulving, 1978). The paradigm involves presenting a list of A-B pairs for study, testing for recognition of the B members, then testing for the recall of those same B members with their A partners as cues. Thus any B member could be retained or forgotten on both tests, it could be recog-

nized but fail to be recalled, or it could be recalled but be missed in recognition. If recall and recognition are stochastically independent events, then the overall proportion of B terms recognized on the first test,  $P(R_n)$ , should equal the likelihood of recognition for the subset of B terms recalled on the second test,  $P(R_n/R_c)$ . In contrast, if recall is dependent on recognition, recalled items should show a recognition advantage, with  $P(R_n/R_c)$  exceeding  $P(R_n)$ . As it turns out, recalled items do show a recognition advantage, but that advantage is quite modest. Specifically,  $P(R_n/R_c) - P(R_n) = .5 P(R_n) P(\bar{R}_n)$  over a large number of experiments (Tulving & Wiseman, 1975); for example, if .60 of all B terms are recognized on the first test, then .72 of the recalled items would be recognized, and .28 would be missed. In short, the two measures are not independent, but the obtained degree of dependency is not impressive. There are two basic ways of explaining the relation. First, recognition and recall could involve independent retrieval processes, in which case it is necessary to specify some other reason for the correlation between tasks. Second, the processes could be dependent, in which case it is necessary to specify factors that attenuate the correlation between tasks.

The prototype of dependency theories is the generation-recognition account (e.g., Bahrick, 1970; Martin, 1975), which characterizes recall as a staged process in which the subject generates response candidates, then recalls those that exceed a recognition criterion. Recalled items could fail to be recognized for several reasons, such as random variability in performance (Wallace, 1978), or the imposition of a more stringent criterion in recognition than in recall (Kintsch, 1978). Alternatively, since different retrieval cues are provided in the two tasks, different memory locations could be accessed (Reder, Anderson, & Bjork, 1974). Similarly, Rabinowitz, Mandler, and Barsalou (1977) argue that an item is remembered as a fairly literal

The research was supported by National Research Council (Canada) Grant A8122 to Ian Begg. Reprints may be obtained from the author, Department of Psychology, McMaster University, Hamilton, Ontario L8S 4K1, Canada. The manuscript was prepared in part while I was on sabbatical leave at the University of Toronto. While many colleagues at Toronto and McMaster, in addition to anonymous reviewers of a previous version, offered helpful criticism, the most sustained help was provided by Grant Harris, John Mitterer, Douglas Upfold, Ginny Begg, and Larry Jacoby.

perceptual trace that is adequate for recognition but not recall, and as a conceptual trace that is adequate for either task (see also Jones, 1978). The common feature of the accounts is the assumption that recall is dependent on recognition if the same memory trace or location is used in both tasks, but that the degree of dependency between measures of recognition and recall will be attenuated because of variability in the likelihood that the appropriate trace will be accessed in both tasks.

Other accounts assume that the retrieval processes are independent to begin with. For example, recognition of B requires that B as an encoded test cue contact the appropriate trace, while recall of B with A as a cue requires that A contact the trace. If each type of retrieval depends on the similarity between the trace and the encoding of the cue (e.g., Begg, 1978b; Flexser & Tulving, 1978), then the two types of access attempts will be independently likely to succeed. However, there could be dependency between the success of the two different tests if some items are better or stronger than others (Flexser & Tulving, 1978), or if the act of recognizing a test item has some chance of priming that item, making it more accessible in later recall (Humphreys & Bowyer, Note 1). The present account also assumes independence between recognition and cued recall processes, but offers a different explanation of why there is some correlation between measures of performance. Specifically, if any traces are lost from storage, both tasks will fail, thus producing a correlation between measures. The present account will now be detailed.

The organizational theory of concern has been presented in more precise form elsewhere (Begg, 1978a, 1978b; Begg & Young, 1977). For present purposes, a simple metaphor is adequate to illustrate the notion of retrieval independence. A telephone directory is a concrete memory system consisting of traces, each of which contains a name and an associated number. Normal use of the directory is a sort of cued recall, in which a name cue provides access with a trace containing that name and the desired number. Although it is rarely done, it is also possible to begin with a number and directly locate that number in the directory, thereby "recognizing" it as a list member. Obviously, the process of recognizing a number is unrelated to the process of recalling the same number with a name cue, and to the stage of recall in which the name cue is itself recognized. By the present account, the directory is analogous to the contents of memory in the case where all studied pairs have available traces. Any performance failures reflect difficulties in access, and under such conditions, recognition of a term is unrelated either to recall of the same term or recognition of other terms represented by the same trace.

The analogy between a directory and the contents of a human memory system is only reasonable if no trace

loss occurs in memory. However, directories themselves are frequently subjected to acts of vandalism in which some of their traces are physically removed. Let us consider two different acts of vandalism, and the consequences of those acts for the degree of dependency between measures of recognition and recall success; the processes remain independent, but their outcomes will covary. Assume a set of conditions under which numbers can be directly located with  $P(R_n) = .40$ , and names can be located, hence allow recall of their numerical associates, with  $P(R_c) = .80$ . Because the processes are independent,  $P(R_n/R_c) = P(R_n) = .40$ . Let one vandal tear out .20 of the pages at random, producing trace loss in all-or-none fashion. This procedure would reduce the level of  $P(R_c)$  to .64 and  $P(R_n)$  to .32. However, the likelihood of recognizing any remaining number would still be .40, and since recalled numbers necessarily remain,  $P(R_n/R_c) = .40$ , which exceeds  $P(R_n)$ . Similarly, if a different vandal defaced .20 of the names from another directory, and defaced an independent random sample of .20 of the numbers, the levels of recall and recognition would be reduced to .512 and .32 respectively, but  $P(R_n/R_c)$  would again remain .40. Thus regardless of whether trace loss is all or none or fragmentary,  $P(R_n/R_c)$  will exceed  $P(R_n)$ , producing a dependency between measures of performance. Further, the more trace loss there is, the larger the dependency will be. It is clearly the case with the vandalized directory that the measured dependency does not reflect any dependency between the retrieval processes.

By the present account, the contents of the memory system at the time of test are analogous to the contents of a vandalized directory. For mnemonic purposes, this explanation will be referred to as vandal theory. The major prediction from vandal theory is that the degree of dependency between measures of recall and recognition will be greater if conditions foster high rather than low degrees of trace loss. Equivalently, recognition failures of recalled words should be relatively infrequent with high degrees of trace loss, but quite common with better retention of traces. The first experiment assesses the prediction. It is assumed that meaningful processing of word pairs yields less trace loss than does rote repetition of the same pairs (cf. Dempster & Rohwer, 1974). Accordingly, rote repetition should show high dependency between measures of retention, and few recognition failures, while meaningful processing should show low dependency between measures but relatively many recognition failures.

## EXPERIMENT 1

The major purpose of Experiment 1 was to determine the degree of interdependence among measures of retention for A-B pairs studied by rote or meaningful processing. The basic test procedure involved a recognition test and a staged recall test, in which subjects first

classified potential cues as old or new, then recalled the B partners of the recognized cues (e.g., Bower, 1970). In addition to assessing predictions from vantage theory, the experiment also included manipulations designed to assess predictions from other accounts. Those manipulations will be considered briefly at this point.

One variable in the experiment was the test history of the pairs. Although all studied pairs were tested for recall of B with A as a cue, subsets of pairs were tested for recognition of the B members (B-ONLY), and A members (A-ONLY), both the A and B members (BOTH), or neither member was tested in recognition (NEITHER). The central condition is the B-ONLY condition, which corresponds to the usual procedure for assessing recognition and recall of the same items. For this condition, rote processing should produce higher levels of dependency between recall and recognition than should meaningful processing. The A-ONLY condition was included to assess the amount of variability in recognition performance from trial to trial. With these pairs, the same A terms were tested first for recognition as list items, then again for recognition when they appeared as cues. Clearly, there should be a high degree of dependency between two successive identical tests; for example, from Rabinowitz, Mandler, and Barsalou's (1977) Table 4, only .114 of items recognized on a second test failed to be recognized on a first test on which the overall miss rate was .393. The question is not whether there is any variability in recognition, but whether there is enough recognition failure of recognized words to account for the incidence of recognition failure of recalled words (cf. Wallace, 1978).

The BOTH items were included to allow analysis of the degree of dependency between recognition of the A and B terms. If all variance in performance reflects access failures, the two measures should be independent; however, any all-or-none trace loss would produce a correlation between measures. Thus vantage theory expects higher dependency between the measures after rote than after meaningful processing. It is also possible to determine the level of interdependence in recognition of partners in the B-ONLY condition by comparing initial recognition of the B terms to recognition of the A partners in the first stage of recall; the same prediction holds.

The NEITHER items were included to assess the extent to which the presence of items on the first test is sufficient to produce a dependency between measures of performance. It is clearly the case that previously tested items are recalled better than previously untested items (Broadbent & Broadbent, 1977; Postman, 1975; Humphreys & Bowyer, Note 1). Further, the assumption that items missed in recognition gain no recall advantage by being present on the recognition test (Tulving & Watkins, 1975; Wiseman & Tulving, 1976) is reasonable, because items missed in recognition are much less likely to be recalled than are untested items (Broadbent &

Broadbent, 1977; Rabinowitz, Mandler, & Patterson, 1977). Therefore, the advantage in recall for tested items comes from items recognized on the first test. This priming of some recognized items, perhaps because they receive additional study on the first test, is the reason given by Humphreys and Bowyer (Note 1) for the typical degree of dependency between measures of recall and recognition. It follows that if rote processing produces high dependency, as predicted by vantage theory, rote processing should also show a substantial influence of prior testing, with recall of B-ONLY items much better than recall of NEITHER items.

The final manipulation of the experiment involved a factorial manipulation of criterion stringency both in the recognition test and in the cue recognition stage of recall. Kintsch (1978) argues that the imposition of a more stringent criterion in recognition than in recall is sufficient to increase the incidence of recognition failure. However, if more stringent criteria reduce recognition level as much as they reduce recognition failure of recalled words, the degree of dependency between tasks would not be affected. Wiseman and Tulving (1976) found that more and less stringent criteria in recall differed in the level of recall, but not in the incidence of recognition failure. However, Rabinowitz, Mandler, and Barsalou (1977) found that the imposition of a more stringent recognition criterion in the first test increased recognition failure relative to the overall miss rate; unfortunately, their manipulation of criterion stringency is suspect, because it had little effect on initial recognition performance. By vantage theory, manipulation of criterion stringency only affects access with traces, so it should not produce differing degrees of interdependence over measures.

A final note pertains to the manner of presenting items. The usual procedure for demonstrating recognition failure is quite elaborate, with set-establishing lists and so on, as if investigators presuppose that recognition failures are in some way unusual or surprising events, requiring subterfuge in their demonstration. By the present account, such failures are to be expected as a matter of course if retention of traces is reasonably good, simply because the A and B members of a trace are independently likely to access that trace. Therefore, the present experiment is a quite straightforward study-test procedure.

### Method

**Subjects.** Eighty student volunteers were paid \$2/h; each was tested individually in a session lasting 1 h, with 10 subjects in each of eight conditions.

**Material and Procedure.** The basic materials consisted of a study list, two forms of a recognition test, and two forms of a recall test. A total of 280 nouns above 5.0 in rated imagery were selected from Paivio, Yuille, and Madigan (1968) to serve as study items and distractors, with all assignments of words to conditions being random. The study list consisted of 80 pairs of nouns, one member of which (B) was to be the response item in the later recall test, with the other (A) to be the recall cue. The

experiment was conducted in three stages, namely, study, recognition testing, and recall testing. During study, the list was presented at the rate of 10 sec/pair, during which time subjects either listed meaningful mediators linking the members, or repeated the pair aloud as often as possible, with half the subjects performing each processing task.

Study was followed by a 5-min recognition test, during which subjects circled items they remembered as having appeared in the study list. For a stringent condition, the recognition test consisted of 80 list items and 80 new distractors; for a lenient condition, only the 80 list items appeared on the test. Finally, there was a 10-min staged recall test in which subjects were presented with a list of cues; subjects first performed a cue identification task in which they circled cues they remembered as having appeared in the study list, then a response production task in which they recalled whatever associated responses they could. As with the recognition test, the cue identification test was either lenient, with the 80 A terms from the studied pairs presented with no distractors, or stringent, with 80 distractors interspersed with the same 80 A terms; 40 of the distractors were new items, with 40 repeated from the recognition test, in order to determine whether there would be a higher false alarm rate for repeated than new distractors.

The basic design was a 2 by 2 by 2 orthogonal design with instructions, recognition criterion, and criterion for cue identification as independent factors. The experiment also included a within-subjects manipulation, with 20 of the studied pairs serving in each of four test conditions. All 80 pairs were tested in the same manner in recall, with the A member as a cue for the B associate. However, in the earlier recognition test, there was an A-ONLY condition, in which the A terms appeared as targets; a B-ONLY condition, in which the B terms appeared as targets; a BOTH condition, in which both pair members appeared as targets, interspersed throughout the test list; and a NEITHER condition, in which the pairs were untested in recognition. In all, subjects studied 80 pairs, were tested for recognition of 80 words, then were tested for recall of 80 words and identification of 80 cues.

## Results and Discussion

All means are presented as proportions accompanied either by standard deviations in parentheses or MSe. The  $\alpha$  level was set at .05 for all tests of significance. The experiment allowed calculation of several dependent measures of recall, recognition, and the relations between tasks, each of which was analyzed by a separate analysis of variance with criterion level and instructions as independent factors, and with test history and the basis of conditionalization, where appropriate, as repeated-measures factors. Several results allow for a simplified presentation. First, criterion level interacted with no other variables, and is therefore considered in a separate section, with averaged data reported in all other analyses. Second, meaningful processing exceeded rote processing in all retention measures by a considerable amount, as usual; this fact is simply mentioned for what it is worth, and will not be discussed further. Third, across the experiment, there was very little within-subjects variance in any analysis; in each case, in order to differ significantly on a post hoc *t* test, two individual means were required to differ numerically by .05 or greater. As a final point, the paper reports miss rates for recalled items,  $P(\overline{Rn}/Rc)$ , to make it comparable to most other published reports using similar paradigms to investigate recognition failure.

**Relations between tests.** The most important results of the experiment concern the extent to which various retention measures are independent or interdependent. First, consider the B-ONLY pairs. If recognition and recall are independent processes, then the initial miss rate,  $P(\overline{Rn})$ , should equal the miss rate for recalled items,  $P(\overline{Rn}/Rc)$ ; further, if the A and B partners are independently recognizable, then  $P(\overline{Rn})$  should also equal the miss rate for B terms whose A partners were recognized in the first stage of recall,  $P(\overline{Rn}/CRn)$ . The results for B-ONLY pairs appear in the top half of Table 1; numbers differing by .05 or greater differ reliably ( $MSe = .014$ ). It is quite clear that the overall and conditional miss rates are about equal after meaningful processing, but that the conditional rates are lower than the overall rate after rote processing. Although the results will be discussed in more detail below, the key finding is that the degree of dependency between measures of dependency is substantial after rote processing, but minimal after meaningful study, exactly as expected by *vandal theory*.

Following meaningful study, .80 (.11) of the A terms were recognized as cues in recall; the B partners of the recognized A terms were no more recognizable than average, .22 vs. .24. Likewise, with the BOTH pairs, the miss rate for B terms was .23 (.15), with .21 (.15) of the B terms missed given recognition of the A terms elsewhere in the test. Thus it appears that the members of pairs studied meaningfully are independently likely to be recognized. Further, for the .53 (.19) of the B terms that were recalled, the recognition miss rate, .21, was about the same as the miss rate for the larger set of B terms whose partners were recognized as cues, and the total set of all B items. In fact, the obtained value of .21 is so close to the estimate for independence that it actually exceeds the level predicted by Tulving and Wiseman's (1975) formula, .15. The general picture for pairs studied meaningfully is thus one of independence across all measures of retention.

In marked contrast to the results following meaningful study, rote study was followed by substantial interdependence between all measures of retention. Thus, .56 (.18) of the A terms were identified as cues in recall; their B partners had a lower miss rate, .31, than the average for all items, .39. Similarly, for BOTH pairs, .33 (.19) of the B terms whose A partners were recog-

Table 1  
Overall Miss Rate [ $P(\overline{Rn})$ ] Compared to the Miss Rate Given Cue Identification [ $P(\overline{Rn}/CRn)$ ], and the Miss Rate Given Response Recall [ $P(\overline{Rn}/Rc)$ ]

	$P(\overline{Rn})$	$P(\overline{Rn}/CRn)$	$P(\overline{Rn}/Rc)$
B-Only Pairs			
Meaningful	.24 (.15)	.22 (.15)	.21 (.15)
Rote	.39 (.21)	.31 (.22)	.16 (.24)
A-Only Pairs			
Meaningful	.21 (.13)	.11 (.12)	.11 (.15)
Rote	.40 (.19)	.17 (.17)	.05 (.12)

nized were missed, compared to the overall miss rate of .41 (.19). That is, there is more interdependence in the recognition of partners following rote than following meaningful study. Further, the .23 (.15) of B terms that were recalled were especially unlikely to be missed in recognition, with  $P(Rn/Rc) = .16$ , which is lower than the value of .27 predicted by Tulving and Wiseman's (1975) formula.

The results for B-ONLY items are quite simple. All measures are interdependent after rote processing, while the same measures are independent of each other after meaningful processing. Such a finding is precisely what vandal theory expects, because trace loss would yield indiscriminate interdependencies over all measures of retention. It is also worth noting that meaningful study had a lower overall miss rate than rote study, but a higher miss rate for recalled items. Therefore, the findings are inconsistent with any theory that predicts recognition failure solely on the basis of overall recognition levels.<sup>1</sup>

Next, consider the A-ONLY pairs, the results for which appear in the bottom half of Table 1. For these pairs, recognition miss rates on the first test refer to the same A terms that were presented for recognition in the first stage of recall. Therefore, some degree of dependency between measures should result. The point to note is that all the conditional probabilities are between .10 and .14 lower than the corresponding values for B-ONLY items; that is, there is certainly more dependency with A-ONLY pairs, but the dependency introduced by repeated identical tests of the same items is specific to those items. With meaningful study, .76 (.13) of the A terms were identified correctly in recall, and .52 (.19) of the A terms were effective cues for their response partners. However, the identified cues were no less recognizable overall, with a miss rate of .11, than the subset of cues that were effective in response production (miss rate = .11), consistent with the assumption that cue recognition and response production are independent of each other. With rote processing, .57 (.21) of the A terms were recognized as cues, and .17 (.17) led to correct recall of their B partners. In contrast to the results from meaningful processing, however, successful cues were much more unlikely to be missed in recognition (miss rate = .05) than were cues that were simply identified (miss rate = .17).

The A-ONLY pairs provide results that indicated that there is indeed variability in recognition from test to test, but that there is simply not enough variability to generate the levels of recognition failure obtained with B-ONLY pairs. Consequently, dependency accounts that rely on variability to explain the attenuation of the correlation between measures are insufficient. Further, the A terms were less likely to be identified on both tests after rote than after meaningful study, making it implausible to explain the higher incidence of recognition failure of recalled items in meaningful conditions on

the basis of increased variability in which location (Reeder et al., 1974) or trace (Rabinowitz, Mandler, & Barsalou, 1977) is accessed. In general, dependency theories fare quite poorly with the present results.

**Other analyses.** This section considers some analyses that are less central to the vandal theory predictions but evaluate expectations from other accounts. First, priming theory (Humphreys & Bowyer, Note 1) is assessed; if dependency between measures results from learning of some items on the first test, then there should be more such learning after rote than after meaningful study. Second, the effects of criterion stringency are assessed to determine whether changes in stringency over tests are sufficient to affect the degree of dependency between tests (Kintsch, 1978).

*Prior testing.* In agreement with most prior studies, recall was higher for B-ONLY than for NEITHER items. After meaningful study the respective means were .58 (.19) vs. .48 (.18), with means of .23 (.15) vs. .15 (.14) after rote study. Since the latter difference is, if anything, smaller than the former, priming on the first test cannot account for the results reported in the preceding sections. The advantage of prior testing was, as usual, restricted to the items recognized on the first test, with missed items less than half as likely to be recalled as untested items. Additionally, in conditions that include distractors on both tests, repeated and new distractors were equally likely (.06) to be accepted as false alarms. The assumption that the prior test does not affect the traces of missed items is not compromised by any findings here.

*Criterion effects.* The general finding was that the presence of distractors on a test reduced measures of performance on that test but had no other consequences. For A-ONLY pairs, recognition miss rates were higher with a more stringent criterion (.35 vs. .24) [ $F(1,76) = 13.3$ ,  $MSe = .038$ ], and so were conditional miss rates [ $F(1,72) = 6.84$ ,  $MSe = .050$ ], with means of .18 vs. .10 given later cue identification, and .12 vs. .04 given later recall. For B-ONLY pairs, a stringent criterion also led to higher miss rates overall [ $F(1,76) = 9.47$ ,  $MSe = .022$ ] and conditional [ $F(1,72) = 5.98$ ,  $MSe = .052$ ]; overall miss rates were .36 vs. .27, miss rates conditional on cue identification were .31 vs. .23, and miss rates given recall were .23 vs. .14. In short, the effect of a more stringent criterion was to increase all miss rates between .08 and .11. Likewise, a more stringent criterion on the first stage of recall reduced the proportion of cues identified (.62 vs. .67) [ $F(1,72) = 4.89$ ,  $MSe = .020$ ] but did not affect any conditional measures. As a final point, false alarm rates were between .05 and .08 for all conditions. It may also be of interest to note that miss rates were substantial even if no distractors were present.

## Conclusions

The most important finding of the experiment is that rote study, which reduces retention relative to

meaningful study, increases the degree to which measures of retention are interdependent. Further, when there is interdependence, it applies indiscriminately over all retention tests. According to *vandal theory*, more trace loss should result in higher correlations between tests, regardless of whether the processes required by those tests are related or not. The fact that measures were virtually independent after meaningful processing presents difficulties for all theories that characterize the retrieval processes as being dependent. The results also are inconsistent with theories that explain the interest relation on the basis of criterion changes, performance variability, or the influence of the former test on the latter. Although the results are directly in accord with *vandal theory*, they do not contradict either Flexser and Tulving's (1978) goodness-of-encoding model or Jones' (1978) dual mechanism account; the latter will be addressed in Experiment 2, the former in Experiment 3.

## EXPERIMENT 2

Although Experiment 1 was successful in varying the degree of dependence between recognition and recall, rote processing was associated with generally lower levels of performance than was meaningful processing. Jones (1978) argues that performance can be mediated either by direct access with traces or, more indirectly, if traces allow generation-recognition processes to occur. His account leads to the prediction that lower levels of recall should show higher levels of dependency between recall and recognition, exactly as found in Experiment 1. However, the accounts are testably different. According to *vandal theory*, retrieval conditions that increase or decrease accessibility of traces will increase or decrease levels of performance but will not affect the degree of dependency between tasks, since that dependency varies with trace loss. The basic rationale is that if a trace is accessed under any circumstances, the trace must exist and be available in all other circumstances regardless of whether those circumstances foster access. Thus it should be possible to make access more or less likely within instructional conditions. Rote processing should show high dependency between tests, but meaningful processing should show low dependency, and further, the level of dependency within either instructional condition should be invariant. By Jones' (1978) account, dependency should vary inversely with the level of recall performance.

The strategy adopted for varying accessibility was to use pairs of members that are differentially recognizable. For example, rare concrete nouns such as "leopard" are more recognizable than common abstract nouns such as "truth" (Begg & Rowe, 1972). Thus in a pair such as "leopard-truth," "leopard" should be better recognized than "truth," but "truth" should be better recalled than "leopard" in cued recall (Townsend & Saltz, 1975). These differences, however, reflect differential accessi-

bility and should not relate to changes in the interdependency between measures of retention. As in Experiment 1, subjects either studied the pairs by rote or meaningful processing; this difference should influence interdependence as before.

## Method

**Subjects.** Twenty student volunteers were paid \$2/h; each was tested individually in a session lasting 45 min.

**Materials and Procedure.** Overall, subjects studied 78 pairs of nouns, then were tested by recognition and recall. In recall, one member of each pair was a recall cue; the response term from each pair was also a target in the recognition test, which included 78 distractors. Intervening between study and the recognition test was a 5-min filled interval, with a 6-min interval between recognition and recall. The pairs were presented at the rate of 10 sec/pair, with 7 min allowed for recognition and 5 min for recall. Each of the 78 pairs was printed on a separate index card, with presentation order randomized over pairs. The 78 response terms and the 78 distractors were randomly ordered on a recognition list; each item was rated on a scale from No-4 (sure new) to Yes-4 (sure old). The 78 cue terms were randomly ordered on a recall sheet, with a blank space next to each.

The major variables of the experiment were instructions, with half the subjects listing meaningful mediators for each pair and half repeating the pair aloud, and pair type: A total of 24 pairs contained members that differed in concreteness but not frequency, 18 had members differing in frequency but not concreteness, 24 had members differing in both, and 12 had members that differed in form class. In constructing the 24 pairs whose members differed only in concreteness, 72 nouns with frequency between 50 and 20/million were selected from Paivio et al. (1968); 36 were concrete ( $I > 6.0$ ) and 36 were abstract ( $I < 3.6$ ). Twelve nouns of each type were selected at random to be recall cues, 12 to be both responses and recognition targets, and 12 to be distractors. The 12 concrete cues were randomly paired with the 12 abstract responses, and the 12 abstract cues were paired with the 12 concrete responses. Similarly, 18 frequency-varied pairs were constructed: The words had imagery values between 3.6 and 6.0; half had frequencies greater than 100, half had frequencies between 0 and 20, and there were an additional nine words of each type for distractors. Another set of 24 pairs had a common (frequency  $\geq 50$ ) abstract ( $I < 3.6$ ) member, and a rarer (20  $>$  frequency  $>$  0) concrete ( $I > 6.0$ ) member, with another 12 words of each type to be distractors. The final set of 12 pairs each contained a rare concrete noun (as above) and an extremely high-frequency bigram, such as "he," "do," "if," and so on, all of which had frequencies over 1,000, with a further six words of each type for distractors. These pairs are not discussed further, since the false alarm rate for the bigrams was enormous (.43!).

Thus 78 pairs were constructed. Each pair had one member that should be more recognizable than the other; for half the pairs, the more recognizable member was both the recognition target and the response in recall, while in the other half, the less recognizable member was the target and response.

## Results and Discussion

**Preliminary analyses.** Preliminary analyses pertain to overall levels of recall and recognition, with the main purpose of determining whether the manipulation of accessibility worked. Recognition performance has  $P(\text{yes})$  as the dependent measure, since partitioning on the basis of confidence only affected the level of performance. Each analysis to follow is a 2 by 2 by 2 analysis of variance with instructions as an independent

factor, and test (recall or recognition) and response type (more or less recognizable member) as repeated factors ( $df = 1, 18$ ).

*Frequency-varied pairs.* Meaningful processing exceeded rote ( $F = 10.1$ ), recognition exceeded recall ( $F = 191$ ), and the two variables interacted ( $F = 16.4$ ) because the advantage for meaningful processing was greater in recall [.49 (.23) vs. .14 (.11)] than in recognition [.85 (.13) vs. .79 (.18)]. More importantly, the expected interaction between measure and response type was reliable ( $F = 68.1$ ); the rarer member was the better recognized [.90 (.11) vs. .74 (.22)] but worse recalled [.21 (.25) vs. .42 (.28)] member of the pair.

*Concreteness-varied pairs.* Meaningful processing exceeded rote ( $F = 17.9$ ), recognition exceeded recall ( $F = 198$ ), and the variables interacted ( $F = 5.10$ ), again because the instructional effect was greater in recall [.35 (.16) vs. .04 (.04)] than in recognition [.78 (.14) vs. .63 (.20)]. The interaction between measure and response type was again reliable ( $F = 6.40$ ), with the concrete response the better recognized [.77 (.20) vs. .63 (.24)] but worse recalled [.17 (.22) vs. .22 (.30)] of the two members.

*Both-varied pairs.* Meaningful processing exceeded rote [.62 (.10) vs. .38 (.09),  $F = 31.8$ ], and recognition exceeded recall [.76 (.15) vs. .24 (.20),  $F = 168$ ]. Rare concrete responses were better recognized [.84 (.16) vs. .67 (.19)] but worse recalled [.18 (.17) vs. .31 (.26)] than common abstract responses, yielding a reliable interaction ( $F = 33.3$ ).

*False alarms.* False alarms were generally higher for less recognizable items, with means of .13 (.17) vs. .08 (.10) for the frequency-varied pairs, .16 (.15) vs. .07 (.07) for the concreteness-varied pairs, and .21 (.15) vs. .08 (.11) for the both-varied pairs, simply indicating that the obtained differences in recognition hit rates are underestimates of the true differences in recognizability.

**Relations between tasks.** The preliminary analyses showed that the attempt to negatively correlate recognition and recall of the same items was successful. However, since those variations reflect differential accessibility, there should be little effect on the degree of dependency between recognition and recall. On the other hand, there should be more dependency between the measures after rote than after meaningful processing.

Table 2 presents the probability of recognition failure for recalled items, compared to the level predicted by Tulving and Wiseman's (1975) formula. The pattern is obvious. Following meaningful processing, the obtained values are either equal to or greater than predicted values; following rote processing, the obtained values are all smaller than the predicted values.

### Conclusions

The relative rareness of recognition failure after rote processing, with much higher levels for meaningful processing, replicates the results of Experiment 1. On the other hand, simple changes in overall recall and recognition levels reflecting variation in accessibility had no obvious relation to the degree of dependency between recognition and recall. On balance, trace loss is the single most important determinant of the relative incidence of recognition failure, as expected by vandal theory.

### EXPERIMENT 3

The purpose of the third experiment was to contrast the present vandal theory with Flexser and Tulving's (1978) goodness-of-encoding theory. To this point, all dependencies attributed to the differential likelihood of trace loss are consistent with the goodness account. Specifically, the model contains a parameter,  $p$ , the likelihood that any particular feature of the item B is encoded at study; as  $p$  increases, the measures approximate independence. In other words, the better the encoding of the study item, the more likely it is that recall and recognition will be independent. Meaningful processing could thus simply produce better traces than rote processing, with the present results to be expected from the account. The point of departure between the alternate conceptions is that the goodness account applies to the entire trace of each A-B episode, while the vandal account applies to the association between A and B independently of how well the separate items are encoded.

To illustrate, let us return again to the directory metaphor, with  $P(Rc) = .80$ , and  $P(Rn) = P(Rn/Rc) = .40$ . Suppose a vandal were to tear pages in half vertically, and in so doing dissociate .20 of the names from

Table 2  
Recognition Failures in Experiment 2

Pair Type		$P(\overline{Rn}/Rc)$			
		Meaningful Processing		Rote Repetition	
		Obtained	Predicted	Obtained	Predicted
Frequency Varied	Common Response	.13	.13	.05	.20
	Rare Response	.06	.05	.00	.06
Concreteness Varied	Abstract Response	.22	.20	.00	.30
	Concrete Response	.08	.08	.00	.21
Both Varied	Abstract Common Response	.19	.14	.19	.30
	Concrete Rare Response	.03	.03	.12	.17

their numerical associates, but leave all the items in the directory. This vandalistic act would reduce  $P(Rc)$  to .64, but would have no effect on either  $P(Rn)$  or  $P(Rn/Rc)$ . In other words, if the difference between conditions is only in the degree of association or dissociation between members, but not in the likelihood that the traces of those members will be lost, the conditions should differ only in recall, not either in recognition or in the degree of dependency between measures. However, since the more integrated traces are certainly better than less integrated traces, the former should be closer to independence between recall and recognition than should the latter according to Flexser and Tulving (1978).

A manipulation that affects the degree of association between items, but not the recognizability of those items singly, is interactive as opposed to separate imagery. Pairs learned by interactive imagery are much better than pairs learned by separate imagery in later cued recall, but the individual words are equally well recognized in the two conditions (Begg, 1978a; Bower, 1970; Dempster & Rohwer, 1974). Thus, according to vandal theory, the extent to which pieces of information are well integrated with each other in memory should not of itself increase or decrease the extent to which measures of recall and recognition are interdependent. The concept of integration also suggests that there may be some loss of identity of the terms that are integrated; perhaps the mother in an integrated image of "railroad-mother" is less identifiable as a mother on a later test than is a mother imaged alone. In fact, Baker and Santa (1977a, 1977b) found that well integrated information may suffer in later recognition of the separate parts in a new context. However, such fluctuation in recognition clearly affects access and should have no consequence on later intertest dependencies.

In summary, subjects processed pairs of items either jointly or separately, then were tested for recognition and recall of the individual items. In order to make the study more comparable to the work of Baker and Santa (1977a, 1977b), both members of each pair were presented simultaneously, rather than successively as in Begg's (1978a) experiments. Begg's manipulation allows both separate and integrated study of pairs in the joint conditions and might, therefore, reduce the likelihood of a negative effect of integration on recognition of single items. At any rate, the degree of dependency between recognition and recall should be invariant over the instructional manipulation.

## Method

**Subjects.** Forty student volunteers were paid \$2/h, with each tested in an individual session lasting 1 h.

**Materials.** The basic material consisted of two study lists, a recognition test, and a recall test. The study material and distractors for the retention tests were nouns above 5.0 in rated imagery (Paivio et al., 1968), with all pairings and assignments of items to conditions determined by a random schedule. The 78 pairs of nouns constituting the study material were arranged in two lists, with 52 pairs in each. A total of 26 pairs were on both lists, 26 were only on List 1, and 26 were only on List 2;

the three sets, respectively, are referred to as repeated pairs, List 1 pairs, and List 2 pairs. One member of each pair was designated at random to be the A term, the other the B term. The recognition test contained 78 distractors, and 78 list items as targets. For 39 pairs, 13 from each set, the A terms were the targets, with the B terms as targets for the remaining half of the pairs; the two conditions, respectively, are referred to as A-ONLY and B-ONLY test conditions. For the recall test, the 78 A terms appeared as recall cues, with an additional 78 distractors; each item had a blank space for recall of a B term.

**Procedure.** Subjects first studied List 1; half the subjects produced a verbal associate for each of the 104 words in the list and half produced a verbal associate for each of the 52 pairs, with 10 sec/pair allowed for both the separate and joint conditions. Then subjects studied List 2; half of each of the above two groups generated an image for each of the 104 words and half generated an image for each of the 52 pairs, with 10 sec/pair again allowed for both the separate and joint conditions. In each case both pair members were presented on a card side by side. Ten subjects were in each of the four combinations of initial study tasks.

Following study, subjects were given the recognition test list, with 5 min allowed for completion. Then subjects were given the recall test, with 5 min allowed for cue identification, followed by an unpaced period during which subjects attempted to recall the B terms associated with identified cues.

In all, the design included two between-subjects variables, and two within-subjects variables. The between-subjects variables were defined by factorial combination of whether each study list was processed by separate or joint processing of pair members. One within-subjects variable was defined by whether pairs were studied on both lists, or only on List 1 or List 2. The other within-subjects variable was defined by whether both the recognition targets and the to-be-recalled items were the same B terms, or whether the recognition targets and recall cues were the same A terms.

## Results and Discussion

The results from the experiment are presented in Table 3. Although the table contains a large quantity of data, the major results are easy to summarize. For B-ONLY items, the different measures of retention are all independent of each other, regardless of the type of initial processing. In recall, the likelihood that items recognized on the first test would be recalled later was nowhere reliably higher than the overall recall rate. Likewise, the two stages of recall, namely, cue identification and recall given cue identification, were not reliably different for items recognized on the initial test than for all items. As in Experiment 1, all measures were independent of each other, and further, there was no suggestion that better integrated traces are more independent than poorly integrated traces.

The A-ONLY items, whose A terms were tested both as recognition targets and as cues, do show evidence of dependency in recall, with recognized terms more likely than average to produce correct recall for List 1 [ $F(1,38) = 3.15$ ,  $p = .09$ ], for List 2 [ $F(1,38) = 10.3$ ], and for repeated pairs [ $F(1,36) = 8.31$ ]. It is clear, however, that the basis of the dependency in recall is that the previously recognized cues were more likely than average to be recognized again in the first stage of recall; previously recognized items were more likely than overall to be recognized as cues for List 1 [ $F(1,38) = 67.3$ ], for List 2 [ $F(1,38) = 158$ ], and for repeated pairs [ $F(1,36) = 51.6$ ]. When recall is conditionalized on



**Table 3**  
Overall Recall Performance Compared to Recall Given Prior Recognition

Item Set	Processing Task	Recognition Test	Recall		Cue Identification		Response Production	
			Overall	Conditional	Overall	Conditional	Overall	Conditional
B-Only Test Condition								
List 1	Separate	.56 (.24)	.02 (.07)	.01 (.04)	.54 (.17)	.48 (.22)	.04 (.13)	.06 (.22)
	Joint	.43 (.14)	.08 (.11)	.11 (.17)	.33 (.18)	.33 (.20)	.17 (.25)	.24 (.36)
List 2	Separate	.69 (.18)	.04 (.09)	.05 (.14)	.50 (.19)	.50 (.22)	.06 (.17)	.07 (.18)
	Joint	.49 (.20)	.17 (.18)	.20 (.22)	.39 (.21)	.46 (.21)	.37 (.25)	.35 (.32)
Repeated	Separate	.88 (.06)	.04 (.10)	.04 (.10)	.81 (.24)	.80 (.24)	.04 (.10)	.04 (.10)
	Mixed*	.78 (.14)	.27 (.18)	.28 (.19)	.83 (.14)	.81 (.19)	.37 (.29)	.34 (.22)
	Joint	.64 (.21)	.42 (.20)	.42 (.25)	.67 (.17)	.66 (.18)	.63 (.25)	.62 (.30)
A-Only Test Condition								
List 1	Separate	.54 (.23)	.004 (.02)	.01 (.03)	.59 (.23)	.78 (.30)	.01 (.04)	.01 (.04)
	Joint	.39 (.13)	.05 (.08)	.08 (.15)	.42 (.19)	.75 (.26)	.07 (.14)	.09 (.16)
List 2	Separate	.59 (.17)	.02 (.07)	.05 (.15)	.58 (.20)	.82 (.19)	.05 (.18)	.07 (.21)
	Joint	.44 (.17)	.09 (.09)	.19 (.23)	.49 (.16)	.83 (.20)	.22 (.21)	.23 (.26)
Repeated	Separate	.85 (.17)	.02 (.07)	.02 (.07)	.88 (.12)	.93 (.09)	.03 (.09)	.03 (.09)
	Mixed*	.77 (.14)	.19 (.17)	.21 (.24)	.77 (.14)	.88 (.13)	.24 (.23)	.24 (.24)
	Joint	.72 (.12)	.40 (.23)	.51 (.28)	.79 (.13)	.92 (.14)	.51 (.27)	.54 (.26)

\*Mixed is averaged over both orders of separate processing on one trial and joint on the other.

cue identification, the advantage for previously recognized items is not reliable in any analysis.

The key finding is that recall and recognition are independent measures regardless of the degree of integration between pair members, and despite large fluctuations in the level of recall and recognition performance. The only dependency found was that items recognized on one test are likely to be recognized again on another, which should surprise nobody. The results are the same as those found in Experiment 1 for meaningful processing. Thus, whether meaningful processing is produced by direct instructions, or imaginal or verbal processing of items, measures of recognition and recall are independent. Apparently, there is very little trace loss under these conditions.

**Other analyses.** Several other results obtained, which will be briefly detailed here.

**Processing tasks.** In agreement with the findings of Baker and Santa (1977b), joint processing leads to reduced recognizability of single items. In initial recognition of B-ONLY pairs, the effect was reliable for List 1 pairs [ $F(1,38) = 6.82$ ], List 2 pairs [ $F(1,38) = 12.4$ ], and repeated pairs [ $F(3,76) = 6.22$ ]. In cue identification, the effect was still present but less robust; for List 1 pairs,  $F(1,38) = 11.9$ , for List 2 pairs,  $F(1,38) = 4.08$ ,  $p = .05$ , for repeated B-ONLY pairs,  $F(3,36) = 3.13$ , and for repeated A-ONLY pairs,  $F(3,36) = 1.53$ ,  $p = .22$ . By and large, then, integrated processing does lead to some loss of recognizability of single items. However, integrated processing led to better recall than separate processing for List 1 pairs [ $F(1,38) = 5.23$ ], for List 2 pairs [ $F(1,38) = 9.43$ ], for repeated B-ONLY pairs [ $F(3,36) = 7.94$ ], and for repeated A-ONLY pairs [ $F(3,36) = 8.56$ ]. Thus joint processing, if it is accomplished in a truly integrated fashion, may lead to a tradeoff in which items are highly accessible once cues are recognized, but in which those cues are somewhat less likely to be recognized.

**Prior testing.** The general finding was that previously tested items are more likely to be correct in the appropriate stages of recall than are previously untested items. Thus A-ONLY items, whose cues were previously tested, led to higher likelihood of cue identification than B-ONLY items, whose cues were not previously tested; this difference was reliable for List 1 pairs [ $F(1,38) = 5.80$ ] and List 2 pairs [ $F(1,38) = 6.78$ ], but not for repeated pairs. Conversely, the B-ONLY pairs were better recalled than the A-ONLY pairs, with reliable differences for List 1 [ $F(1,38) = 4.65$ ] and List 2 [ $F(1,38) = 6.26$ ], but not for repeated pairs. As in Experiment 1, there was some effect of prior testing, but that effect influenced only the level of performance and not any relations between tests.

### Conclusions

The obvious conclusion is that recall and recognition are independent processes, and that performance in recognition and recall tests will also be independent if trace retention is good. The degree of integration between items has no effect on dependency between measures, if integration is manipulated within a level of meaningful processing. However, if the strength of association between items is confounded with the "goodness" of encoding of individual items, as in comparing meaningful to rote processing, the latter shows considerable levels of dependency, or low incidence of recognition failure.

### GENERAL DISCUSSION

The present account will be reviewed, along with the evidence that favors the account over competitors. Then some more general points will be addressed. Vandal theory is basically a description of the contents of memory at the time of testing retention. Some pairs will be memorially represented by single traces of both members, some pairs will be represented by sep-

arate traces of each member, other pairs will have only one member memorially represented, and others will be unrepresented. Such a state can occur in a variety of ways, including forgetting of entire joint or separate traces, forgetting of parts of joint traces, dissociation of joint traces into constituents, or any combination of those factors. Regardless of the contents of memory, the process by which a test item is recognized is independent of the process by which that same item is retrieved by a different recall cue. However, if an item is recalled, that item and the cue are memorially represented by a joint trace, and the item is therefore available for recognition whether or not that retrieval attempt succeeds. Moreover, single items represented by joint traces are not more recognizable than single items represented by single traces, although the latter are inaccessible in cued recall. Items could fail to be recalled if the cue does not contact an available joint trace, or if only single traces exist, or if no traces exist. Since all recalled items are available for recognition, but some unrecalled items may have been lost, recalled items will generally be more recognizable than average. Of course, if all traces are available, recalled items will be no more recognizable than average, and could even be less so, in the hypothetical case where integration of cue and response is so strong that the response is very poorly recognized in isolation.

The central finding of the present study is that measures of recall and recognition are independent after meaningful processing of items, regardless of whether that processing treats pair members separately or jointly, and regardless of whether access in retention is good or poor. The fact that the measures are independent is clearly in support of the assumption that the retrieval processes are independent. Theories that assume dependency between the processes are seriously compromised by these results. Although some reduction in measured dependency could result from such factors as variability in performance, criterion change, or variability in which memory trace or location is accessed, it is simply inelegant, if not implausible, to assume that some combination of these factors is exactly enough to remove all traces of an inherent dependency.

Although dependency theories can be dismissed as accounts of the present data, this should not be taken as an outright rejection of the theories for all purposes. For example, if study material consisted of prominent members of common categories, it would be extremely sensible for subjects to employ generation-recognition strategies; in fact, I have proposed exactly that explanation elsewhere to account for performance with categorical material (Begg, 1978b). However, with arbitrarily paired items, it is difficult to conceive of a strategy of generating from semantic memory a small enough number of response candidates to be useful. Thus with categorical material, one should expect quite low levels of recognition failure. Similarly, if the study list consisted of sentences, with a recognition test for one word, followed by a cued test for the same word with the

remainder of the sentence as a cue, relatively low levels of recognition failure would be expected, again because the task can be accomplished by a generation-recognition procedure in which the processes are dependent.

As a further point, the fact that different members of a pair are independently recognizable should not be taken as evidence against their joint storage in memory. Thus with highly integrated pairs, recall of one should guarantee recall of its partner; for example, Begg (1972) found that recall of nouns from pairs like "white horse" was only .40, but recall given that the adjective was recalled rose to .87. Of course, dependency in the sense of all-or-none recall is unrelated to dependency over test occasions. As a concrete example, the probability that a drawn heart is red, 1.00, is greater than the overall probability of red, .50, on a draw, thus showing a dependency between suit and color; however, the probability of drawing a red card on a second draw is independent of what happened on the first draw (unless a vandal removed a higher proportion of one type of card than another).

Although it is possible to conclude that recall and recognition are basically independent processes for retrieving the same memory information, there are several accounts of why the success of the two types of task would covary. Humphreys and Bowyer's (Note 1) account is that items recognized on the first test receive additional study and thereby become more recallable; by the present metaphor, that is analogous to circling recognized numbers in the directory, making recall somewhat easier later. Although there is some recall benefit for previously recognized items, that benefit affects only the level of recall, and not the degree of dependency between tasks; that is, there was no more benefit for rote-processed than for meaningfully processed items in Experiment 1, but much more dependency, and in Experiment 3, there were reliable testing effects but no dependency. Consequently, the account is inadequate.

The final competitor to be considered is Flexser and Tulving's (1978) goodness-of-encoding model. There are two major shortcomings of the model. First, although the model assumes retrieval independence, it has quite some difficulty in predicting independence between measures of performance, since with retrieval independence their parameter  $k$  equals 0 and predicts performance along the Tulving and Wiseman (1975) function. Unless  $k$  is allowed to become negative, or there is a fortuitous choice of their other parameters, measured independence is beyond the theory. The second shortcoming of the model is that the parameters that refer to the likelihood of feature selection, hence memorability, apply to the entire event in which a pair is encountered, simply summing over the two items, with no separate parameter for whether or not that summing occurs. In the case of separate processing, feature selection can be quite good without the formation of any joint episodic trace of the pair of items. Their theory is basically a strength theory, but without

any measure of associative strength as opposed to item strength. As a result, the model can be rejected as it now stands.

The present theory is certainly a simple one, since trace loss is the sufficient condition for interest dependency. Trace loss may also be a necessary condition with randomly paired items, since variability in access alone is not sufficient to produce such dependencies. At some level of analysis, trace loss is an intuitively plausible thing to have happen. However, it is no more possible to prove the absence of traces than to prove the nonexistence of unicorns. Nevertheless, the burden of proof rests with those who would argue that all information resides in storage, even if that information is not accessible under any attempted retrievals. As a disclaimer, I would certainly not wish to argue that all forgetting represents trace loss—only some. The more independent failed attempts to access some information there are, the more likely it becomes that the information simply does not exist.

Finally, although the incidence of trace loss was assumed here to be greater for rote than for meaningful processing, the experiments could have been conducted in other ways. Thus abstract items may be more easily forgotten than concrete, with the result that there should be relatively few failures with abstract items unless some meaningful form of processing is guaranteed. Likewise, nonsense material, numbers, and other easily lost items should show few recognition failures. The present approach offers a technique for deciding whether or not differences between conditions represent differential trace loss, although to avoid circularity, it will be necessary to devise some other index of trace loss.

#### REFERENCE NOTE

1. Humphreys, M. S., & Bowyer, P. A. *The relationship between recognition and recognition failure: An artifact of sequential testing*. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, November 1978.

#### REFERENCES

- BAHRICK, H. P. Two-stage model for prompted recall. *Psychological Review*, 1970, 77, 215-222.
- BAKER, L., & SANTA, J. L. Context, integration and retrieval. *Memory & Cognition*, 1977, 5, 308-314. (a)
- BAKER, L., & SANTA, J. L. Semantic integration and context. *Memory & Cognition*, 1977, 5, 151-154. (b)
- BEGG, I. Recall of meaningful phrases. *Journal of Verbal Learning and Verbal Behavior*, 1972, 1, 431-439.
- BEGG, I. Imagery and organization in memory: Instructional effects. *Memory & Cognition*, 1978, 6, 174-183. (a)
- BEGG, I. Similarity and contrast in memory for relations. *Memory & Cognition*, 1978, 6, 509-517. (b)
- BEGG, I., & ROWE, E. J. Continuous judgments of word frequency and familiarity. *Journal of Experimental Psychology*, 1972, 95, 48-54.

- BEGG, I., & YOUNG, B. J. An organizational analysis of the form-class effect. *Journal of Experimental Child Psychology*, 1977, 22, 503-519.
- BOWER, G. H. Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 529-533.
- BROADBENT, D. E., & BROADBENT, M. H. P. Effects of recognition on subsequent recall: Comments on "Determinants of recognition and recall: Accessibility and generation" by Rabinowitz, Mandler, and Patterson. *Journal of Experimental Psychology: General*, 1977, 106, 330-335.
- DEMPSTER, R. N., & ROHWER, D. W. Component analysis of the elaborative encoding effect in children's learning. *Journal of Experimental Psychology*, 1974, 103, 400-408.
- FLEXSER, A. J., & TULVING, E. Retrieval independence in recognition and recall. *Psychological Review*, 1978, 85, 153-171.
- JONES, G. V. Recognition failure and dual mechanisms in recall. *Psychological Review*, 1978, 85, 464-469.
- KINTSCH, W. More on recognition failure of recallable words: Implications for generation-recognition models. *Psychological Review*, 1978, 85, 470-473.
- MARTIN, E. Generation-recognition theory and the encoding specificity principle. *Psychological Review*, 1975, 82, 150-153.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*. Monograph Supplement, 1968, 76(1, Pt. 2).
- POSTMAN, L. Tests of the generality of the principle of encoding specificity. *Memory & Cognition*, 1975, 3, 663-672.
- RABINOWITZ, J. C., MANDLER, G., & BARSALOU, L. W. Recognition failure: Another case of retrieval failure. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 639-664.
- RABINOWITZ, J. C., MANDLER, G., & PATTERSON, K. E. Determinants of recognition and recall: Accessibility and generation. *Journal of Experimental Psychology: General*, 1977, 106, 302-329.
- REDER, L. M., ANDERSON, J. R., & BJORK, R. A. A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, 1974, 102, 648-656.
- TOWNSEND, D. J., & SALTZ, E. The effects of stimulus frequency on subsequent recall of backward associates. *Canadian Journal of Psychology*, 1975, 29, 3-10.
- TULVING, E., & WATKINS, M. J. Structure of memory traces. *Psychological Review*, 1975, 82, 261-275.
- TULVING, E., & WISEMAN, S. Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society*, 1975, 6, 79-82.
- WALLACE, W. P. Recognition failure of recallable words and recognizable words. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 441-452.
- WISEMAN, S., & TULVING, E. Encoding specificity: Relation between recall superiority and recognition failure. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 349-361.

#### NOTE

1. Although at present no theory explains recognition failure on the basis of overall recognition level, Tulving and Wiseman's (1975) formula does relate the two. The formula cannot, however, be a complete account because the parameter  $c$  cannot exceed 1.00 without generating predicted values of  $P(R_n/R_c)$  in excess of 1.00. Flexser and Tulving (1978), as well as the present experiments, present points above the maximum value predicted by the formula with  $c = 1.00$ .

(Received for publication June 9, 1978;  
revision accepted February 6, 1979.)