Relational information and the context effect in recognition memory

MICHAEL S. HUMPHREYS
Northwestern University, Evanston, Illinois 60201

A role for relational information was examined for the paradigm in which recognition-memory performance on items tested in the same context in which they were studied is compared with performance on items tested in different contexts. Over a series of five experiments, randomly formed pairs were used to manipulate the context of high-frequency English words. Comparisons were made between instructional manipulations designed to influence the use of relational information, and between yes/no, confidence rating (both between- and within-subject), and forced-choice tasks. There was a context effect not due to the use of inappropriate response strategies. However, high-criterion subjects resembled those subjects who were specifically instructed to use relational information, while low-criterion subjects showed little or no context effect. A model specifying the relationship between item and relational information and how relational information influences decisions in recognition-memory paradigms was proposed.

In studies of word recognition, there is always some change from the context in which the to-be-remembered (TBR) word is studied to the context in which it is tested. In general, the TBR word is studied as part of some list and tested as part of another list. In addition to the differences between the lists, there will be some changes in the experimental setting and in the internal states of the subject. One way of varying the degree of similarity between the study and test trial contexts is by presenting other words at the same time and location as the TBR word is presented. The words which are presented in spatio-temporal proximity to a TBR word (either at the time of study or at the time of test) will be referred to as the context for that TBR word.

Light and Carter-Sobell (1970) manipulated context at the time of study by simultaneously presenting the TBR word along with some context material. They then tested the TBR word along with some of the material which comprised the study trial context, with new material not previously presented to the subject, or without any material. This basic design will be referred to as the Light and Carter-Sobell paradigm. Specifically, they took homographic nouns, paired them with an adjective and embedded the adjective-noun combination in a sentence for presentation to their subjects on a study trial. This was designed to induce a particular encoding of the homographic noun. On the test trial, the TBR word was presented along with the adjective with which it had been studied, with a new adjective not previously presented, or by itself. The condition where the TBR

word was tested along with the adjective with which it had been studied resulted in better performance then the conditions where the context had been deliberately changed. Similar results have been reported when context was manipulated by presenting pairs of words on the study trial and then testing the TBR words in their study-trial pairing, in new pairings, or by themselves. The conclusion from several experiments using associatively related words (Thomson, 1972; Tulving & Thomson, 1971) and using randomly formed word pairs (Thomson, 1972; Underwood, 1974; Winograd, Karchmer, & Steele Russel, 1971) has been that the condition where the TBR word was tested in its study-trial pairing was superior to the other conditions.

These results have typically been interpreted as being due to differential encoding. The following quote from Light and Carter-Sobell (1970) illustrates this position: "Only one memory representation per noun is tagged for recency during the study phase of an experiment. Correct recognition of study nouns depends primiarily on locating recency information associated with ... properties of the memory representations of test items. Recognition context (the semantic interpretation biased by the recognition adjective) specifies which memory representation will be examined for recency information during the recognition test. If the tagged memory representation is not the one examined for recency information S will be unable to located recency tags associated with ... features of the test noun" (p. 8).

Thus, the context material is seen as influencing recognition-memory decisions by influencing the information that is stored about a TBR word (the memory representation with which the information is stored) and by influencing the information that can be retrieved about a TBR word (which memory representation is examined). The context material is

not seen as a source of information in its own right. This position can be contrasted with a position which was suggested by Mandler (1972) in his discussion of the effects of organization on recognition-memory decisions (see also Mandler, Pearlstone, & Koopmans, 1969). While Mandler (1972) clearly intended that his discussion apply to the Light and Carter-Sobell paradigm, he did not explicitly discuss how organization would affect this paradigm, so the account which follows should be taken as an interpretation or extension of Mandler's (1972) position. In the Light and Carter-Sobell paradigm, organization refers to the relationship imposed on the study trial by presenting the TBR word along with some context material. The presence of some of this context material, when the TBR word is tested, provides the subject with some information in addition to that which is available when the TBR word is presented alone. This additional information can be used to help identify the TBR word. As Mandler (1972) suggests, the subjects might use the context material in an attempt to retrieve the TBR word and TBR words which would otherwise not be identified as old may be so identified when this retrieval check works. This suggestion is speculative, and for the purposes of this paper, the general position, that when making recognition-memory decisions subjects use information derived from the paired presentation of the context material and the TBR word on the study trial, will be explored. This general position will be referred to as the relational information hypothesis.

The overall goal of this study was to arrive at a plausible account of how the presence of some of the study-trial context material improves recognition-memory decisions. However, three questions of a more specific nature were first examined. (a) Is it possible to make a distinction between information about the paired presentation of two items (relational information) and information about the occurrence of an item (item information)? (b) If such a distinction can be made, can it be shown that subjects use relational information when they are asked to make a decision about whether an item is old or new? (c) Can response strategies be identified in the use of relational information?

Item vs. Relational Information

Consider the situation where subjects study pairs of words and are tested on a list which contains some of the original study pairs, some pairs formed by taking two old words which have been studied in different study pairs, and some distractor pairs (pairs which have one or two new words in them). When tested on such a list, it should be possible for subjects to make decisions about which pairs are the same as study pairs and about which pairs have two old words in them. If both of these types of decisions can be made with reasonable accuracy and if the decisions which

are made are sufficiently different, it would seem reasonable to attribute the difference to the use of more than one type of information. Item information will be defined as the information that is more heavily involved in the decisions about which pairs consist of two old words, and relational information will be defined as the information more heavily involved in the decisions about which pairs are the same as study pairs.

Evidence for the Use of Relational Information

Some of the studies already cited report results which seem to support the relational information hypothesis. Light and Carter-Sobell (1970, Experiment 3) changed the context of TBR words by presenting two kinds of new adjectives on the test trial. These new adjectives were designed to induce either the same encoding as the study adjective had induced or a different encoding. Thus, a subject might have studied the adjective-noun combination *sliced ham* and then been tested on *sliced ham, smoked ham,* or *radio ham.* The authors refer to these as the same adjective, same meaning, and different meaning conditions, respectively. The hit rates (HRs) for these conditions were .82, .73, and .63, respectively. The false alarm rates (FARs) were .09, .28, and .33, respectively. Considering both the HR and the FAR, there was a large change in performance going from same adjective to same meaning and a smaller change going from same meaning to different meaning. This comparison seems to show that it is the study trial adjective which carries most of the information, as predicted by the relational information hypothesis. The problem for the differential encoding position is that a change in the adjective which should have resulted in a small change in meaning produced nearly as much of a deficit as a change in the adjective which should have resulted in a dramatic change in the meaning.

A second problem for the differential encoding position occurs with the variation on the basic paradigm introduced by Tulving and Thomson (1971). They observed that words that were studied alone and tested as part of a pair of words were not recognized as accurately as were words that were studied alone and tested alone (also see Thomson, 1972, Experiments 1-4). The problem for the differential encoding position is that this effect of adding context seems to depend on the presence of study-trial pairs, as Thomson (1972, Experiment 5) was unable to replicate the effect when he used a study list which consisted only of individual words. In the experiments where the effect was obtained, it appears as if the presence of study-trial pairs (some of which were present in the test list) may have biased the subjects towards using relational information when they made decisions about words that were tested as part of a pair.

In general, to support the relational information hypothesis, it will be necessary to demonstrate that the recognition-memory decisions subjects make in these paradigms are incompatible with an assumption that only one type of information is involved. In the following experiments, changes in performance which occurred as the result of experimental manipulations or which accompanied changes in criterion were examined. The basic idea was that a change in the ability to discriminate between old and new words when there was no change in the available information could be taken as evidence for more than one type of information being involved in the decisions.

Response Strategies

The context effect in the Light and Carter-Sobell paradigm is defined as a difference in performance between a TBR word that is tested in its study-trial context and a TBR word that is tested in a changed context. This difference could arise because the use of relational information improves performance on the TBR words which are tested in their study-trial context or because the use of relational information reduces performance on TBR words tested in a changed context. The latter will be said to result from the inappropriate use of relational information. The previously discussed effect of adding context appears to be inappropriate in this sense. Judging from the performance of subjects who only studied individual words, it would appear that there was information available to the subjects, if they had bothered to use it. Instead, it appears that their reliance on relational information depressed performance on those words which had been studied alone and tested as part of a pair. For an alternative way to use relational information, consider the situation where a subject is asked to discriminate between TBR words which are tested in their study-trial pairing and new words which are also tested as part of a pair. To the subject, some of the new words will probably appear to be old. It would probably help in discriminating between the TBR words and the distractor words if some of these old-looking new words could be rejected, because there was no relational information about the pair in which the new word was tested. Of course, if some of the TBR words were also tested in a new pairing, this use of relational information to reject distractors would be inappropriate.

In the following experiments, attempts were made to influence the use of relational information through the choice of paradigm and the instructions given to the subjects. The object was to see if there were changes in performance which could be attributed to the use of response strategies.

Summary of Experiments

A yes/no task was used in Experiment I along with instructional manipulations which were designed to demonstrate that both item and relational information could be used to make recognition-memory decisions. Experiments II, III, and IV were designed to see if there was evidence for the use of response strategies. A confidence-rating task was used in Experiments II and III, and a forced-choice task in Experiment IV. A mixed-list design was used in Experiment II, and an unmixed-list design in Experiment III. The type of task and the instructions given should have provided the subjects with information on which to base response strategies. In Experiment V, an alternative to the Light and Carter-Sobell paradigm was examined to see if a context effect could be obtained in the absence of differential relational information.

EXPERIMENT I

Experiment I consisted of three between-subject conditions which differed only in terms of the instructions used on the test trial. Each of the three groups studied a list of pairs and received a recognition-memory test on a list of pairs. The instructions for the test suggested the use of item information (identify the old words or identify the pairs with two old words) or the use of relational information (identify the pairs which are the same as study-trial pairs).

Method

Design and Procedure. A group testing session was employed, where all subjects were treated identically up to the time of the test. After the subjects were seated in the experimental room, test booklets were passed out with instructions not to open the booklets. The subjects were told that they would be shown a series of word pairs, one at a time, and were instructed to learn the pairings so that they could produce the word on the left if shown the word on the right and the word on the right if shown the word on the left. The pairs of the study list were then projected on a screen at a 3.2-sec rate. The test trial instructions were given after the study trial. The test was unpaced and consisted of four types of word pairs: (a) the test pair was the same as a study pair (intact); (b) the test pair consisted of one word from each of two study pairs (rearranged); (c) the test pair consisted of one word from a study pair and one word not included in the study list (old-new); (d) the test pair consisted of two words not included in the study list (new-new). As some of the subjects in the pair judgment condition had appeared to ignore the instructions, this condition was rerun, using instructions and examples designed to emphasize that both intact and rearranged pairs were to be circled.

Subjects. The subjects were 128 introductory students at Northwestern University who served as part of a course requirement. In the original experiment, 32 subjects had been randomly assigned to each of three instructional conditions. An additional 32 subjects were then used in a rerun of the pair judgment condition.

Lists and materials. A set of 160 four-letter words, all with noun functions, were chosen from the most frequent 2,500 words in the Thorndike-Lorge word count. These words were used in all subsequent experiments. In this experiment, 128 of the words were used. Each subject studied a list of 32 pairs and was tested on a list of 32 pairs. The test order was positively correlated with the study order as words presented in a given quarter of the study list were tested in the corresponding quarter of the test list.

Of the 32 pairs in each subject's test list, 8 pairs were intact, 8 were rearranged, 8 were old-new, and the remaining 8 were

Table 1
Study Pairs Used With Test Pairs
BOOK DART and LAWN NEST

Type of Test Pair*	Study Pairs	
Intact	BOOK	DART
	LAWN	NEST
Rearranged	BOOK	POLE
	BELL	DART
	RENT	NEST
	LAWN	WIRE
Old-New (Form A)**	BOOK	POLE
	RENT	NEST
Old-New (Form B)**	BELL	DART
	LAWN	WIRE

*These pairs were also used as New New pairs.
**Two forms were required so that the old word was equally often on the left and right of the test pair.

new-new. Each study word was included in at most one test pair, and when one member of a study pair was used to help form a rearranged or an old-new pair, the other member of that study pair was not included in the test list. Each subject was tested on the same set of 32 test pairs. However, the composition of the study list was systematically manipulated (both the words used and the pairs they were presented in changed), so over subjects each pair in the set of 32 test pairs served equally often as an intact, rearranged, old-new, and new-new pair, and the two words in each test pair served equally often as the old and new members of the old-new pairs. Eight different study lists were used to achieve this counterbalancing. To illustrate the relationship between study and test pairs, the study pairs that were used to assign the test pairs BOOK DART and LAWN NEST to intact, rearranged, and old-new functions are given in Table 1.

Results

In the word judgment condition, the subjects had been instructed to circle individual words. They should have circled both of the words in intact and rearranged pairs, one word (the old word) in old-new pairs, and neither word in new-new pairs. The data from this instructional condition were scored in two ways. The item score is the relative frequency with which individual words were circled, and the pair score is the relative frequency with which both words in a pair were circled. Note that for word judgments the pair scores should produce the same results as in the pair judgment condition, where the subjects were instructed to circle a pair when both members of the pair were old (rearranged and intact pairs). The relational judgment condition differed from the pair judgment condition, as the subjects had been instructed to circle the pairs which were the same as the study pairs (intact pairs).

The results are presented in Table 2 for each of the three instructional conditions and for both scoring methods in the word judgment condition. The entries in Table 2 are the relative frequency with which each type of pair or word within a pair was circled. While scoring the data, it was noted that eight subjects in the pair judgment condition circled no rearranged pairs

while circling an average of 5.3 out of the 8 intact pairs. It was felt that these subjects might have read the instructions carelessly and just assumed that they were only supposed to circle the intact pairs. For this reason, the pair judgment condition was rerun, and it is these results which are given in Table 2.

Old vs. new words. In the word judgment condition, the old words which were tested in rearranged and old-new pairs were circled 54% of the time. The new words which were tested in old-new and in new-new pairs were circled 14% and 10% of the time, respectively. The results from the pair scores and from the pair judgment condition indicated that the frequency with which both words in a pair were judged to be old was a function of the number of old words in the pair. In Table 2, for the word judgment (pair scores) and the pair judgment conditions compare the frequencies for rearranged pairs where there are two old words with the frequencies for old-new pairs where there is one old word and with the frequencies for new-new pairs where there are no old words. These results were, of course, expected and merely indicate that the subjects could perform in accordance with the instructions. However, the same pattern can be seen in the relational judgment condition, where 25% of the rearranged, 10% of the old-new, and 6% of the new-new pairs were circled. Most subjects showed this pattern, as 21 subjects circled more rearranged than old-new and 3 circled more old-new than rearranged; there were 8 ties. There were 12 subjects who circled more old-new than new-new, 4 who circled more new-new than old-new, and 16 who tied. Ignoring ties, these differences were significant, using a two-tailed sign test, at the .01 and .10 levels, respectively.

Intact vs. rearranged. In the word judgment condition, the difference between the frequency with which words from intact and rearranged pairs were circled was .15. This difference was significant,

Table 2
Relative Frequency With Which a Word or Pair Was Circled: Experiment I

	Type of Judgment			
	Word*			
Test Pair	Item Score	Pair Score	Pair	Relational
Intact	.69	.59	.75	.66
Rearranged	.54	.29	.61	.25
Old with New**	.54			
Old-New		.08	.30	.10
New with Old†	.14			
New-New	.10	.03	.10	.06

*For word judgments, the first column contains the frequency with which a word from that type of pair was circled, the second column the frequency with which both words in that type of pair were circled.
**The old word in an old-new pair.
†The new word in an old-new pair.

$F(1,31) = 22.06$, $Ms_e = .02$, $p < .01$. This estimate of the magnitude of the context effect was in close agreement with the estimate from the pair judgment condition, where the difference in the frequency with which intact and rearranged pairs were circled was .14. This difference was also significant, $F(1,31) = 9.83$, $Ms_e = .03$, $p < .01$. In the relational judgment condition and in the pair scores for the word judgment condition, the differences between intact and rearranged pairs were substantially larger, being .41 and .30, respectively. The difference between intact and rearranged pairs was significantly smaller in the pair judgment condition than it was in both of these conditions, $F(1,62) = 18.26$, $Ms_e = .07$, $p < .01$, and $F(1,62) = 7.31$, $Ms_e = .06$, $p < .01$, respectively. The difference between pair and relational judgments indicates that the subjects making relational judgments were probably using relational information. The difference between pair judgments and the pair scores for word judgments was surprising, as this scoring method should have simulated the pair judgment condition.

Intact vs. new-new pairs. The two conditions where subjects circled pairs and the pair scores for the word judgment condition showed little variability in the ability of subjects to discriminate between intact and new-new pairs. When the difference between the frequency with which intact and new-new pairs were circled was calculated, the largest difference was in the pair judgment condition (.65) and the smallest difference was in the word judgment condition (.56). When these same proportions were used to calculate d', signal detection theory's measure of sensitivity, the largest value was in the word judgment condition (2.03) and the smallest value was in the pair judgment condition (1.96). The measures of discriminability for the relational judgment condition were always intermediate between the values for the other two conditions.

Discussion

In the relational judgment condition, the subjects were apparently able to follow instructions and base their judgments on relational information. In support of this conclusion, there was a substantial difference between the frequency with which intact and rearranged pairs were circled, and this difference was significantly larger than the corresponding difference in the pair judgment condition. However, the frequency with which a pair was circled also varied with the number of old words in the pair. So the recognition-memory decisions made by the subjects in this condition would appear to be based on both relational and item information.

In the word and pair judgment conditions, the subjects were clearly able to discriminate between old and new words and between pairs with different numbers of old words in them. However, they also circled words in intact pairs more frequently than in rearranged pairs (word judgments) and they circled intact pairs more frequently than rearranged pairs (pair judgments). This is, of course, the Light and Carter-Sobell context effect, which has generally been interpreted as being due to differential encoding. Considering only this experiment, it is as plausible to assume that the context effect was due to the difference in relational information about intact and rearranged pairs as it is to assume that the differences between rearranged, old-new, and new-new pairs in the relational judgment condition were due to the use of item information.

The subjects in the word judgment condition were instructed to circle the old words, and the subjects in the pair judgment condition were instructed to circle the pairs in which both words were old. The pair scores for the word judgment condition should make the results from these two conditions comparable. However, this difference in the mechanics of responding apparently improved performance. That is, the subjects were better able to discriminate between rearranged and new-new pairs in the pair judgment condition than they were in the word judgment condition (pair scores), while the ability to discriminate between intact and new-new pairs did not change. There is some evidence that this improvement in performance was accompanied by a change in criterion. That is, the subjects making word judgments circled 42% of the words in the test list, but they only circled both words in 25% of the test pairs. The subjects making pair judgments circled 43% of the test pairs. There was no evidence to suggest that the subjects in the word judgment condition had "misunderstood" the instructions as some of the subjects in the original pair judgment condition seemed to have done. However, it is possible that occasionally they attended only to relational information, ignoring the item information that could have been used.

EXPERIMENT II

Experiment II differed from the word judgment condition of Experiment I in four major ways: (a) A confidence rating procedure was used where the subjects first made a decision about whether the word was old or new and then rated their confidence in their decision. (b) Three types of test pairs were used—intact, rearranged, and old-new—and the subjects made a decision only about the second member of each test pair (in old-new pairs, the second member was always new). (c) The subjects were tested individually and were required to pronounce each word out loud. (d) The relationship between the order of the pairs in the study list and the order of the pairs in the test list was systematically manipulated.

While the confidence rating procedure may be more

difficult for the subjects than the yes/no task used in Experiment I, it also may provide more information about the extent to which the words appear to be old or new. The requirement that a decision be made only about the second member of a test pair and the individual testing session were introduced to simplify the subject's task and to reduce the chance that an inappropriate response strategy would be used. Test order was introduced as a variable to determine if any systematic changes were occurring during the test session which might be attributable to the acquisition of a rule or response strategy.

Method

Design and Procedure. An individual testing session was used and the study pairs were projected on a screen one at a time. The subjects were required to read out loud both members of the study pairs, and they were instructed to learn the pairing. After the study trial, the subjects were fully informed about the types of test pairs to expect. Every subject was then tested on three types of pairs (intact, rearranged, and old-new). The test pairs were projected one at a time, and the subjects were required to read out loud both members of the test pair, to decide whether the word on the right was old or new, and to indicate their confidence in their decisions on a 3-point scale. The study pairs were presented at a 3.7-sec rate, and the test trial was unpaced. Two different testing orders were employed; one was positively correlated with the study order and the other was negatively correlated. To produce the positive correlation (first-in-first-out), items appeared in corresponding quarters of the study and test lists. To produce the negative correlation (first-in-last-out), the order of the test list was reversed.

Subjects. The subjects were 48 students from the same source used in the previous experiment. A counterbalanced order was used to assign an equal number of subjects to the two test orders.

List and materials. The test list for all subjects consisted of the same set of 32 pairs. Each study list consisted of 8 pairs which were tested as intact pairs; 16 pairs, each of which contributed 1 word to form 8 rearranged pairs; and 16 pairs, each of which contributed 1 word to form 16 old-new pairs. The left-right order of the study pairs was preserved in the test pairs. Thus, an intact test pair was studied in the same order in which it was tested, the left and right words in a rearranged test pair was presented on the left and right, respectively, of different study pairs, and the left word in an old-new test pair was presented on the left of a study pair. The word on the right of an old-new pair was always new. Four study lists were required to counterbalance the assignment of test pairs to intact, rearranged, and old-new functions, so that over subjects each test pair served each function equally often. Two test orders were required to produce the two order conditions. After testing half of the subjects, the presentation order of the items in each study list was reversed (the same test lists were used), so, over subjects, the same pairs appeared in the two halves of the study and test lists.

Results

Order conditions. The relative frequency with which the second word in intact, rearranged, and old-new pairs was identified as old for each order condition is presented in Table 3 as a function of test half. The HRs for both intact and rearranged pairs were roughly constant over the two halves of the test list for both order conditions, while the FARs increased sharply from the first half to the second half of the test list. In the first-in-first-out order, 15 subjects showed an increase in the FAR, 5 a decrease,

Table 3
Relative Frequency With Which the Second Word in a Test Pair was Identified as Old: Experiment II

Test Pair	First In First Out		First in Last Out	
	Test Half		Test Half	
	First	Second	First	Second
Intact	.83	.78	.84	.90
Rearranged	.67	.70	.75	.77
Old-New	.20	.34	.22	.35

and there were 4 ties. In the first-in-last-out order, 14 showed an increase, 5 a decrease, and there were 5 ties. Ignoring ties, this degree of consistency was significant, using a two-tailed sign test, at the .10 level for both order conditions. There were no major differences between the two order conditions. The HR was slightly higher in the first-in-last-out order, but this was partially offset by a higher FAR. The combined HR for the intact and rearranged pairs minus the FAR was not significantly different for the two order conditions [$F(1,46) = 1.02$, $Ms_e = .04$, n.s.]. The context effect (the difference in HR on intact and rearranged pairs) was also about the same in the two order conditions. Averaging over test halves, there was a difference of .13 for the first-in-first-out order and .11 for the first-in-last-out order. Both of these differences were significant, $F(1,23) = 9.23$, $Ms_e = .02$, $p < .01$, and $F(1,23) = 7.18$, $Ms_e = .02$, $p < .05$, respectively.

Confidence ratings. Since the correlation imposed between study and test orders had little effect on the size of the context effect, the confidence ratings from both conditions were combined. The ratings were treated as if they represented six levels of confidence in the oldness or newness of an item. They were numbered from -3 to +3, with the negative numbers indicating that the response was new. Thus, a rating of -3 indicates that the subject was highly confident that the item was new. The cumulative frequency of being at least as confident as a given confidence level i is presented in Table 4, as a function of type of pair. Plotting the cumulative frequencies for either the intact or rearranged pairs against the frequencies for the old-new pairs produces a memory operating characteristic (MOC) curve. Two measures of discriminability were calculated and are presented at the bottom of Table 4. These were the HR minus the FAR (HR-FAR) and the area (A) under the MOC curve. The area under the MOC curve for a variety of assumptions equals the probability correct in a two-alternative forced-choice task (see Green & Swets, 1966).

The difference in the frequencies with which words in intact and rearranged pairs were given a +3 rating was .18. The same difference for a rating of +1 or higher (all old responses) was .12. This change in the size of the context effect occurred because the subjects

Table 4
Cumulative Relative Frequency of Being at Least as Confident in the Oldness of the
Second Word in the Pair as Level i: Experiment II

Confidence Level i	All Subjects			High Criterion Subjects			Low Criterion Subjects		
	Intact	Rearranged	New	Intact	Rearranged	New	Intact	Rearranged	New
+3	.66	.48	.06	.62	.37	.04	.71	.59	.08
+2	.78	.66	.20	.71	.55	.15	.85	.77	.26
+1	.84	.72	.28	.75	.60	.20	.92	.83	.35
−1	.89	.76	.36	.82	.65	.27	.96	.87	.45
−2	.96	.88	.62	.93	.82	.51	.99	.94	.72
−3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HR-FAR*	.56	.44		.55	.40		.57	.48	
A**	.87	.78		.86	.75		.88	.82	

*The difference between the frequency with which old words in intact or rearranged words were given a rating of +1 or higher and the same frequency for new words.
**The area under the MOC curve.

were less likely to use a +3 response on rearranged pairs than on intact pairs but were somewhat more likely to use a response of +1 or +2. However, both measures of discriminability indicated a clear difference between intact and rearranged pairs.

To further examine changes in the magnitude of the context effect with changes in criterion, the combined HR for intact and rearranged pairs along with the FAR were used to calculate β, signal detection theory's measure of criterion (see Green & Swets, 1966). This measure of criterion was calculated for individual subjects, who were then divided into two groups based on a median split of their scores. The cumulative frequencies for the three types of pairs for both criterion groups are presented in Table 4. The division into high and low criterion subjects had very little effect on intact pairs. There was a difference of .02 in both A and HR-FAR. However, on rearranged pairs, there was a difference of .07 in A and .08 in HR-FAR. The low-criterion subjects were better at discriminating between rearranged and old-new pairs than were the high-criterion subjects.

Discussion

In the word judgment condition of Experiment I, the difference between the frequency with which words were circled in intact and rearranged pairs was .15. The corresponding difference in this experiment was .12. Although the context effect was slightly smaller, the use of individual testing sessions and the requirement of only one response per pair did not eliminate it. In addition, there were little or no order effects except for the increase in the FAR from the first half to the second half of the test list. The overall pattern of a constant HR and an increasing FAR was similar to that found by McCormack and Swenson (1972). There was no indication that the size of the context effect was changed by the order manipulation. There was thus no evidence from the comparison with Experiment I and from the order manipulation to

suggest that a response strategy was involved in the context effect.

There was still a relationship between the size of the context effect and the criterion adopted by the subjects. The context effect was larger for the high-confidence old responses than it was for all old responses. When subjects were divided into two groups on the basis of their criterion, the size of the context effect increased for high-criterion subjects and decreased for low-criterion subjects. This change in the size of the context effect occurred because there was little change in indices of discriminability for intact pairs and a change in both indices for rearranged pairs. This was the same pattern that was found in Experiment I, both with respect to the change in the size of the context effect with a change in criterion and the stability with which intact pairs were discriminated from the distractor pairs.

EXPERIMENT III

Experiment III differed from Experiment II in that it used unmixed lists on the test trial instead of mixed lists. One group of subjects received a test list which consisted of intact and old-new pairs. A second group received a test list which consisted of rearranged and old-new pairs. For the subjects who had to discriminate between intact and old-new pairs, a strategy of rejecting pairs when the relational information indicated that the pair was probably not an intact pair would be appropriate. If the subjects in Experiment III were better able to discriminate between intact and old-new pairs than were the subjects in Experiment II, it would be evidence for the use of such a strategy. For the subjects who had to discriminate between rearranged and old-new pairs, a strategy which ignored item information in favor of relational information would be inappropriate and presumably unlikely. If discrimination between rearranged and old-new pairs was better in

Table 5
Cumulative Relative Frequency of Being at Least as Confident in the Oldness of the
Second Word in the Pair as Level i: Experiment III

Confidence Level i	All Subjects				High Criterion Subjects				Low Criterion Subjects			
	Condition and Type of Word				Condition and Type of Word				Condition and Type of Word			
	Intact		Rearranged		Intact		Rearranged		Intact		Rearranged	
	Old*	New†	Old**	New†	Old*	New†	Old**	New†	Old*	New†	Old**	New†
+3	.52	.05	.45	.07	.47	.01	.33	.03	.57	.10	.57	.10
+2	.69	.13	.65	.17	.63	.06	.53	.12	.75	.20	.77	.22
+1	.75	.17	.72	.22	.69	.08	.58	.13	.81	.26	.85	.31
−1	.82	.26	.78	.35	.78	.18	.66	.27	.86	.34	.90	.42
−2	.93	.57	.88	.64	.90	.49	.79	.57	.97	.65	.97	.70
−3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HR-FAR	.58		.49		.61		.45		.55		.54	
A	.85		.78		.86		.74		.84		.84	

*Old words tested in intact pairs.   **Old words tested in rearranged pairs.   †New words tested in old-new pairs.

Experiment III than it was in Experiment II, it would suggest that such a strategy had been used in Experiment II.

Method

Design and Procedure. A between-subject design was used where subjects studied the same list but were tested on different lists. For half of the subjects, the test list consisted of intact and old-new pairs, and for the other half, the test list consisted of rearranged and old-new pairs. The subjects on the study trial were instructed to learn the pairing and, after the study trial, were fully informed about the construction of the test list they were to receive. The test order was positively correlated with the study order. The rest of the procedure was the same as in Experiment II.

Subjects. The subjects were 48 students from the same source used in the previous experiments. A counterbalanced procedure was used to assign an equal number of subjects to the two conditions.

Lists and materials. Each subject studied a list of 48 pairs and was tested on a list of 24 pairs. Each test list consisted of either 12 intact and 12 old-new pairs or 12 rearranged and 12 old-new pairs. The test lists were constructed from a set of 36 pairs. To counterbalance the assignment of the pairs in the set of 36 to intact, rearranged, and old-new functions, so that each pair served each function equally often, required three forms, each consisting of two study and two test lists. All 36 pairs were required for the construction of the two test lists within each form. There were 12 pairs common to the two test lists (old-new pairs) and 12 unique pairs in each test list. Whether the unique pairs were intact or rearranged depended on the study list. When one list was studied, one set of unique pairs was intact and the other was rearranged. This assignment reversed when the other list was studied.

Results

The cumulative frequency of being at least as confident as level i as a function of type of pair for all subjects and for high- and low-criterion subjects is presented in Table 5. The two measures of sensitivity, HR-FAR and A, are also presented. The division into criterion groups was again based on a median split of the β values. The median for the subjects receiving rearranged pairs was 1.00, and the median for the subjects receiving intact pairs was 1.25.

The subject's ability to discriminate intact or rearranged pairs from old-new pairs was essentially

the same as it was in Experiment II. Compare the indices of discriminability at the bottom of Tables 4 and 5. There is thus no evidence from this comparison for the use of either response strategy.

The division of subjects into criterion groups which reduced the size of the context effect in Experiment II all but eliminated it in this experiment. This was the result of a sharp increase in the indices of discriminability for the low-criterion subjects who received rearranged pairs and a slight decrease in these indices for the low-criterion subjects who received intact pairs. The high-criterion subjects, especially those who received the intact pairs, were very accurate when they used the high-confidence old response. The subjects receiving intact pairs used a response of +3 on old and new words in a ratio of 67.0 to 1. The same ratio for the subjects receiving rearranged pairs was 11.8 to 1.

Discussion

The results from the last two experiments with respect to the division of subjects into high- and low-criterion groups can be summarized as follows. With high-criterion subjects, the HR on words in rearranged pairs was substantially less than was the HR on words in intact pairs. With low-criterion subjects, the HR on words in intact pairs increased, but this was largely offset by the increase in the FAR. However, the increase in the HR for rearranged pairs was substantially greater than was the increase in the FAR. The same pattern emerged when the area A under the MOC curve was used as the measure of discriminability. This post hoc division into criterion groups cannot establish that it was the criterion adopted which was influencing the size of the context effect. However, in Experiment I, the instructional manipulation apparently produced a change in criterion, and the same pattern of changes in discriminability emerged. Experiment I thus offers experimental evidence supporting a change in the size

of the context effect with a change in criterion. Experiments II and III offer correlational evidence supporting the same conclusion.

In the mixed-list design of Experiments I and II, the depression in performance on rearranged pairs, when a high criterion was adopted, could have occurred because the high-criterion subjects were ignoring item information in favor of relational information. The subjects who were tested on rearranged pairs in Experiment III had been told that the list contained only rearranged and old-new pairs. Simply ignoring item information in favor of relational information was presumably unlikely, so the depression in performance on rearranged pairs was probably not the result of subjects using this response strategy.

EXPERIMENT IV

A forced-choice task was used to further reduce the chance that item information might be ignored. The subjects studied pairs and were tested on a triple of words, one of which was designated as the cue, and the other two of which were designated as the test set. One member of the test set was new and the other was old. The cue and the old word in a test set formed either an intact or a rearranged pair. Forcing a subject to choose between two alternatives should provide an incentive to use all relevant information. When the cue and the old word in the test set formed a rearranged pair, item information would be relevant. When the cue and the old word formed an intact pair, both item and relational information would be relevant.

Method

Design and Procedure. The subjects studied a list of 48 pairs and were tested on a 32-item forced-choice list. Each of the test items consisted of the first word from one of the study pairs (the cue), followed by a set of two words. One word in each set was old (the target), and the other was new (the distractor). The target had either been paired with the cue in the study list forming an intact pair or the target and cue had been studied in different pairs forming a rearranged pair. The subjects were asked to circle the old word in each set. They were told that exactly one word in each set was old, and that, while the cue might help them to identify the old word in some sets, it would be of no help in other sets. The test items were presented on a sheet of paper, and the test was unpaced. The rest of the procedure was the same as in Experiment III.

Subjects. The subjects were 30 undergraduates from the same source used in the previous experiments.

Lists and materials. The entire pool of 160 words was used. The 32 pairs on the test lists (cue plus target) were drawn from a set of 40 pairs. There were 10 study lists, 2 with each of the 5 test lists. When one study list was used, the cue and the target formed an intact pair for half of the test items and a rearranged pair for the other half. When the other study list was used, the relationship between cue and target reversed. A counterbalanced procedure was used, so over subjects each of the 40 pairs served as an intact and rearranged pair equally often and each of the 160 words served equally often as a distractor.

Results

Averaging over subjects, the probability of correctly identifying the target was .85 when the cue and the target formed an intact pair and .80 when they formed a rearranged pair. This difference was significant, $F(1,29) = 4.75$, $Ms_e = .01$, $p < .05$. A majority of subjects showed the effect, since 17 were more accurate with intact pairs, 6 were more accurate with rearranged pairs, and there were 7 who tied. Ignoring ties, this degree of consistency was significant using a two-tailed sign test at the .05 level.

Discussion

Words which were tested in their study trial pairing were more likely to be identified as old than were words which were tested in a different pairing. This was observed in the forced-choice task of this experiment as well as the unmixed-list confidence rating task of Experiment III, the mixed-list confidence rating task of Experiment II, and the mixed-list yes/no task of Experiment I. While there was not one bit of evidence which clearly supported the hypothesis that subjects might be ignoring item information in favor of relational information, the overall pattern of results suggests that this might have occurred. The four experiments can be considered as a series where this strategy becomes increasingly less likely. In the word judgment condition of Experiment I, the difference in the frequency with which words from intact and rearranged pairs was circled was .15; in Experiment II, the comparable difference was .12, and in Experiment III, it was .09. When the area under the MOC curve was calculated, the difference between intact and rearranged pairs was .09 in Experiment II and .07 in Experiment III. In Experiment IV, the difference in the probability correct in the two-alternative forced-choice task was .05.

The conclusion from all four experiments is that with the materials and conditions of these experiments there was a small but reliable context effect. This context effect was not attributable to the use of inappropriate response strategies, though these strategies may have played minor roles in some of the experiments.

EXPERIMENT V

The Light and Carter-Sobell paradigm in which recognition-memory performance on items tested in their study-trial context is compared to performance on items tested in different contexts confounds the opportunity for differential encoding with the loss of relational information. When the context is changed (which may change the encoding), there is a loss of relational information. However, effects attributable to differential encoding can be observed without this confounding. For example, Horowitz and Manelis (1972) compared performance on words like *chair* after they had been studied as part of pairs like *high chair, deep chair,* or *lazy chair*. The authors refer to

these adjective-noun phrases as idioms, meaningful phrases, and anomalous phrases, respectively. When tested by themselves, the components of idioms were not as well recognized as were the components of the meaningful and anomalous phrases. Similar results have been reported by Winograd and Raines (1972), who presented homographs in settings designed to induce high- and low-frequency encodings and then tested the homographs by themselves. Performance was worse when the study trial setting had been designed to induce a low-frequency encoding. In these studies, then, there is evidence for the detrimental effect of a change in context which is not confounded with differential loss of relational information.

In the next experiment, performance on words which were studied as part of a pair and were tested alone was compared with performance on words which were studied alone and tested alone. To the extent that differential encoding occurs with this material, words that were studied alone should have been encoded with a higher frequency encoding than were words that were studied as part of a pair, and the first member of a study pair should have had a higher frequency encoding than the second member (for the latter prediction, see Reder, Anderson, and & Bjork, 1974). Both of these predictions depend on the assumption that when a word is studied as part of a pair the other member of the pair will influence the encoding of that word, and when the pairs are randomly formed the result is likely to be a low-frequency encoding. When the words are presented one at a time, the influence one word has on the encoding of the next is assumed to be attenuated in comparison to the simultaneous presentation of a pair of words.

Method

Design and Procedure. Half of the subjects studied 32 pairs (pair condition) and were tested on 64 individual words (32 old and 32 new). Half of the old words in the test list came from the first position in the study pair, and half came from the second position. The other subjects studied the words one at a time (single condition). The subjects in the two conditions studied the same words and were tested on the same words. The study and the test orders for the two conditions were the same, with the left-right ordering of the words in the study pairs being preserved in the single condition by presenting the two words in adjacent study-list positions. A pair was presented for 3.25 sec, and, as there was a .7-sec slide-change time, the time per word was 1.98 sec. Single words were presented for 1.95 sec, including the .7-sec slide-change time. Subjects were tested individually and were required to pronounce each word in the study list. Subjects in the pair condition were instructed to learn to produce the word on the right if shown the word on the left. Subjects in the single condition were instructed to learn the words. No details about the nature of the test were given to either group until after the study trial. On the unpaced test trial, the words were presented one at a time, and the subjects responded old or new and then rated their confidence in their response on a 3-point scale.

Subjects. The subjects were 32 students from the same source used in the previous experiments. A counterbalanced order was used to assign them to the two conditions.

Lists and materials. A total of 128 words were chosen from the

pool of 160 words. These were randomly formed into 64 pairs. The study list consisted of 32 of these pairs and the test list consisted of 32 old words (one from each study pair) and 32 new words (one from each pair not used in the study list). Only one member from each pair was used in a test list, but over study and test lists both members of a pair were used equally often as old and new words. This counterbalancing required two study and two test lists. The study order and the test order were positively correlated.

Results

The cumulative frequency of being at least as confident as level i for each condition and position within a pair is presented in Table 6, along with two measures of discriminability (HR-FAR and A). For old items in the pair condition, position refers to the position within the study pair. For new items and old items in the single condition, the breakdown by position serves as a control for the fact that different words were used in the left and right positions of the study pairs. Examination of Table 6 reveals no evidence for an effect of position within a study pair in the pair condition and no evidence of a difference between the sets of words used in the left and right study positions. In addition to an absence of an effect of position within the study pair, there was little effect of the study conditions. The subjects in the single condition were slightly more accurate in identifying the old items, but this was compensated for by an increase in the FAR. Neither measure of discriminability supported a conclusion of superior performance in the single condition. Combining the data over the left and right pair positions resulted in values for A of .85 and .84 and in values for HR-FAR of .58 and .56 for the single and pair conditions, respectively.

The first member of the pair (the item on the left) was not recognized better than the second member, as Reder et al. (1974) predicted. The instructions to learn to produce the second member when shown the first and the subjects' habitual left-right reading order

Table 6
Cumulative Relative Frequency of Being at Least as Confident in the Oldness of the Word as Level i: Experiment V

| Confidence Level i | Pair Condition | | | | Single Condition | | | |
| | Left | | Right | | Left | | Right | |
	Old	New	Old	New	Old	New	Old	New
+3	.50	.05	.54	.06	.57	.05	.52	.06
+2	.69	.15	.70	.14	.73	.15	.70	.17
+1	.73	.20	.76	.18	.82	.24	.80	.23
-3	.82	.27	.81	.27	.87	.38	.85	.37
-2	.94	.60	.93	.54	.95	.57	.94	.59
-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HR-FAR	.54		.57		.58		.57	
A	.84		.85		.86		.84	

Note. Left and right for the old words in the pair condition refer to whether the word was on the left or right of the study pair. For the new words and in the single condition, left and right designate the appropriate control words for the old words in the pair condition.

should have contributed to a process whereby the second member was encoded in relation to the first member. In addition, breaking up the study-trial pairing did not produce a decrement in recognition-memory performance, at least in comparison to a condition where items were studied by themselves and tested by themselves. Two conclusions appear plausible. Either the presentation of a random pair of words with instructions to learn the pairing did not increase the frequency with which low-probability or nondominant encodings were used or the subjects were not having any difficulty recovering or using stored information when presented with a set of isolated words on the test trial.

GENERAL DISCUSSION

Three questions about the relational information hypothesis were raised in the introduction. These questions were concerned with the distinction between item and relational information, with evidence for the use of relational information, and with the role of response strategies. In Experiment I, the difference between performance on intact and rearranged pairs was considerably larger in the relational judgment condition than in the pair judgment condition. In addition, the subjects in the word judgment condition were less willing to indicate that both members of a pair were old and showed a larger context effect (pair scores) than did the subjects in the pair judgment condition. In Experiments II and III, this same interaction was present when subjects were divided into groups on the basis of an estimate of the criterion they had adopted. The conclusion is that decisions about whether a test pair is the same as a study pair differ from decisions about whether the words in a test pair are old. Furthermore, it looks as if subjects who adopt high criteria make decisions about whether the words in a test pair are old in the same manner as do those subjects who are instructed to make decisions about whether the test pair is the same as a study pair. A reasonable interpretation of these observations is that there is a distinction between item and relational information, the extent to which these two types of information are used is partially under subject control, and relational information is sometimes used when subjects make decisions about words which are tested in pairs.

The evidence for inappropriate response strategies has already been discussed. While there was no firm evidence supporting their use, the possibility that such strategies may be used should be considered in designing context experiments. In particular, it should be noted that comparisons between forced-choice, yes/no, and confidence-rating tasks as well as between mixed and unmixed lists may be needed to disentangle the various alternatives. In Experiment V, the attempt to produce a context

effect in the absence of differential relational information failed. The importance of this experiment is that the relational information hypothesis could have been rejected if a sufficiently large context effect had been found.

The following model is a tentative proposal as to how relational an item information might be related and how the two types of information might be used in making recognition-memory decisions. The states of memory which are assumed to be present when an old word in an intact or rearranged pair is tested and when a new word in a distractor pair is tested, along with the probabilities of being in these states are given in Equation 1.

	M*	M	$\overline{M}$	
Intact	ab	a(1 − b)	1 − a	
Rearranged	ac	a(1 − c)	1 − a	(1)
New	0	e	1 − e	

The model is formally equivalent to the three-state low- and high-threshold model discussed by Krantz (1969). The states represent the information a subject has about a test item. When an old item is tested either in an intact or a rearranged pair, item information will be present with probability a. When a new word is tested, item information will be present with probability e. When there is information about the occurrence of an old word in the study list, there will sometimes be relational information present which unambiguously identifies that word as having been in the study list (the word is said to be in state M*). The relational information is present with probability b when the word is tested in an intact pair and with probability c when tested in a rearranged pair. Superior performance for the words tested in intact pairs is accounted for by assuming $b > c$. Relational information is never present, p(M* | New) = 0, when a new word is tested. State M is defined as the state where item information is present but relational information is not present. State $\overline{M}$ is defined as the state where neither item nor relational information is present.

It is an open question at this point whether the parameter c, the probability that relational information is present when a word is tested in a rearranged pair, should be set greater than zero. The subjects report that they can sometimes positively identify these words as old because they can remember their reaction when they studied the word or because they can recall the other member of the study pair. In any case, such a state is required if it can be shown that subjects can respond with zero or near zero FARs, even when the HR is appreciably below one. Setting the parameter e greater than zero

reflects the assumption that item information is sometimes present when a new word is tested.

In a two-alternative forced-choice task, the subject is assumed to choose an item in State M* over an item in either M or $\overline{M}$ and to choose an item in state M over an item in state $\overline{M}$. If both items are in state $\overline{M}$ or both in state M, the subject is assumed to choose the correct alternative with probability 1/2. The presence of the state M* is an advantage in the forced-choice task, since it permits the subject to choose the correct alternative even when the new item is in state M. In a yes/no task, a subject could eliminate FAs by only saying yes (old) when in state M*. This would be a high criterion and it would produce the largest difference in the HRs between intact and rearranged pairs. If some FAs are tolerated, the subject might say yes in state M* and in a proportion of the time when in state M. As this proportion increases (as the criterion becomes lower), the difference in the HRs between intact and rearranged pairs will decrease. When the subject always says yes in states M* and M, there will be no difference in these HRs. Further changes in criterion are accomplished by assuming that, in addition to always saying yes in states M* and M, the subject occasionally says yes in state $\overline{M}$. These further changes in criterion produce the same effect on the HRs for words in intact and rearranged pairs.

This model can account for the following results. (a) In a forced-choice task, the target was more likely to be identified as old if the cue and the target formed an intact pair than if they formed a rearranged pair. (b) There was a large difference in the HRs between words in intact and rearranged pairs for high-criterion subjects, but little or no difference for low-criterion subjects. (c) When the criterion changed, it was performance on rearranged pairs which changed, not performance on intact pairs. (d) The high-criterion subjects were very accurate when using the high-confidence old response. The model does not provide an adequate account of the changes in the area under the MOC curve, and in general it does not provide an adequate quantitative account. There are two problems, however, with using this data to test quantitative predictions. The most fundamental problem is that it is data obtained by averaging over subjects and even if individual subjects were behaving in accordance with the model, the group average

would not. The second problem is with the use of a confidence rating task. The subjects were not trained, and it appeared that at times they were uncomfortable with the task. For these reasons, quantitative tests of the model will have to wait for the collection of more adequate data.

REFERENCES

GREEN, D. M., & SWETS, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

HOROWITZ, L. M., & MANELIS, L. Toward a theory of redintegrative memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 6). New York: Academic Press, 1972.

KRANTZ, D. H. Threshold theories of signal detection. Psychological Review, 1969, 76, 308-329.

LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 1-11.

MANDLER, G. Organization and recognition. In E. Tulving & W. Donaldson (Ed.), Organization of memory. New York: Academic Press, 1972.

MANDLER, G., PEARLSTONE, Z., & KOOPMANS, H. S. Effects of organization and semantic similarity on recall and recognition. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 410-423.

McCORMACK, P. D., & SWENSON, A. L. Recognition memory for common and rare words. Journal of Experimental Psychology, 1972, 95, 72-77.

REDER, L. M., ANDERSON, J. R., & BJORK, R. A. A semantic interpretation of encoding specificity. Journal of Experimental Psychology, 1974, 102, 648-656.

THOMSON, D. M. Context effects in recognition memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 497-511.

TULVING, E., & THOMSON, D. M. Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 1971, 87, 116-124.

UNDERWOOD, B. J. The role of the association in recognition memory. Journal of Experimental Psychology, 1974, 102, 917-939.

WINOGRAD, E., KARCHMER, M. A., & STEELE RUSSELL, I. Role of encoding utilization in cued recognition memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 199-206.

WINOGRAD, E., & RAINES, S. R. Semantic and temporal variation in recognition memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 114-119.