

The effect of context on discrimination and bias in recognition memory for pictures and words

KELLY FEENAN and JOAN GAY SNODGRASS
New York University, New York, New York

When the context accompanying a to-be-remembered word is changed between study and test, recognition memory is impaired. The deleterious effect of context change on recognition memory can be viewed as support for encoding specificity theory, semantic theory, or the existence of two bases for recognition. A fourth possible interpretation, examined here, is that the effect of context change on recognition memory is due to an accompanying change in response bias, rather than a "true" decrease in sensitivity to old and new items. In two experiments, the effect of context change on discrimination and bias in recognition of simple line drawings and their names was examined. Bias was measured using two measures shown by Snodgrass and Corwin (1988) to be theoretically independent of their associated discrimination measures. Context change produced marked conservatism in response bias in both experiments but demonstrated an effect on discrimination in the second experiment only. The shift from a neutral to a conservative response strategy as a result of context change may also be seen in other experiments, in which the same experimental paradigm was used with a variety of stimulus materials. We suggest that the major effect of context manipulation is to produce a change in bias. A stimulus in a familiar environment appears to be more familiar than a stimulus in a strange environment, regardless of its old/new status. In addition, there appears to be a true decrease in discrimination with context change, but this is more difficult to detect. The finding that pictures, which are less polysemous than words, are as affected by context change as words are supports encoding specificity theory over semantic theory.

When concepts are studied in one context and tested in another, as opposed to being studied and tested in the same context, there is a decrease in recognition memory performance (Light & Carter-Sobell, 1970; Thomson, 1972; Tulving & Thomson, 1971). The deleterious effects of context change on recognition memory have been demonstrated by several investigators, who have used a variety of materials such as strongly associated, weakly associated, and unassociated word pairs (Tulving & Thomson, 1971; Thomson, 1972), homographs (Light & Carter-Sobell, 1970), and photographs of faces (Watkins, Ho, & Tulving, 1976; Winograd, & Rivers-Bulkeley, 1977).

The results of these studies have been interpreted in one of three ways: as support for the principle of encoding specificity (Tulving & Thomson, 1973); as support for semantic theory (Martin, 1975; Reder, Anderson, & Bjork, 1974); and as support for two bases for recognition (Humphreys & Bain, 1983; Mandler, 1980). The principle of encoding specificity is linked to Tulving's proposal regarding the two systems of episodic and seman-

tic memory (Tulving, 1972, 1983). Because episodes of learning are stored in the episodic system along with their contexts, retrieval of such episodes is dependent upon reinstatement of the contexts. Theorists who reject the notion of two separate memory systems argue that episodes of learning are stored in a single memory system, the semantic system, by tagging occurrences of words or other units with their meanings. The semantic theorists view the adverse effects of context manipulation in recognition memory for words as a change in the sense of the word that does not match the originally stored semantic interpretation. The third view is that recognition of an item as old can be based on one or both of two routes. These routes consist of familiarity and retrieval in Mandler's (1980) model, and item and relational information in the Humphreys-Bain model (1983). In both cases, context provides the opportunity for the retrieval or relational route to operate, and this opportunity is absent when context is deleted or changed.

A question that remains unaddressed by these interpretations is whether the recognition memory effect is due to response bias (Santa & Lamwers, 1974) or to a "true" change in sensitivity. In the present study, we examine the degree to which context change affects response bias in addition to or instead of discrimination. We also compare recognition performance for pictures and words in order to contrast the first two explanations of context change effects. Because words are more polysemous than pictures (contrast the word *saw* with the pictured represen-

The authors would like to thank Elsie Mercado for assistance and June Corwin for comments on an earlier version of this paper. This research was supported in part by a Research Challenge Fund grant from New York University and a University Research Initiative Program grant from the Air Force Office of Scientific Research to J. G. Snodgrass. Reprint requests should be sent to Joan Gay Snodgrass, 6 Washington Place, Room 857, New York University, New York, NY 10003.

tation of the tool), the experiments also permit a comparison of encoding specificity theory with semantic theory. Semantic theory predicts that because words may have more than one "sense," recognition performance for words in this study should be adversely affected by context change, whereas recognition performance for pictures, which have fewer semantic interpretations, should not be impaired. Encoding specificity theory predicts that because both pictures and words are stored in the episodic system along with their contexts, any change between study and test context should result in equal decrements in recognition performance.

Most investigators have adopted the implicit assumption that as long as false positive rates do not vary significantly across context conditions, bias is not a factor. Yet this ignores the fact that a necessary condition for bias invariance with any "reasonable" bias measure (the false positive rate alone is not a reasonable measure) is that there be an inverse relation between hit and false alarm rates—that is, as hit rates increase, the false alarm rates must decrease. The necessary and sufficient conditions for bias invariance depend on the particular bias measure adopted. In fact, a widespread empirical relation across a number of manipulations that affect discriminability of old and new items is the mirror effect: As hit rates for a category of old stimuli increase, false alarm rates for the same category of new stimuli decrease (Glanzer & Adams, 1985).

A notable characteristic of studies of context change in recognition is that the mirror effect is not obtained. As context is changed from study to test, hit rates decrease, but false alarm rates either remain constant or also decrease. Although measures of discrimination that are independent of bias usually show a decrease in discrimination as the test context is made increasingly different from the study context, computation of bias would also show a decreasing tendency toward "yea-saying"—that is, increasing conservatism. In one of these few studies to measure bias in addition to discrimination, Pellegrino and Salzberg (1975a) showed that both d' and $\log \beta$ (both of them computed on condition means rather than on individual subject's scores) showed a decrease across changes in context when the studied weakly and strongly associated pairs of Tulving and Thomson (1971) were presented with new associates.

Recently Snodgrass and Corwin (1988) have argued that all measures of discrimination, including the commonly used "corrected recognition score" (the difference between the hit and false alarm rates), are based on some model of the recognition memory process. They showed that this measure, which we here call P_r in conformity with their notation, is based on a two-high-threshold model of recognition memory. In this model, there is an old item threshold that can be exceeded only by old items and a new item threshold that can be exceeded only by new items, and the old and new item thresholds are equal. Snodgrass and Corwin also showed that several commonly

used bias measures are theoretically dependent on their associated discrimination measures, in the sense that the range of values that the bias measures can attain is decreased as discrimination decreases. Among these bias measures are the likelihood ratio measure β from signal detection theory and the nonparametric measures B'' and B'_H . In contrast, both the bias measure B_r from two-high-threshold theory and the intersection bias measure C from signal detection theory are independent of discrimination, and thus should be used with their corresponding measures of discrimination when bias measures are desired. Accordingly, in the experiments to be reported here, we present measures of discrimination and bias from both models. Although, as noted above, Pellegrino and Salzberg measured bias, they used the signal detection theory measure β (1975a) or the nonparametric measure B'' (1975b), neither of which is appropriate in the present situation because they are both theoretically correlated with discrimination measures and we expect context change to affect both discrimination and bias.

The experiments reported here were designed to examine the effects of context change on both discrimination and bias. Using a paradigm similar to that of Tulving and Thomson (1971), we studied the effects of context on recognition memory for words and pictures. In Experiment 1, we evaluated the effect of context change on studied single items, by pairing them with an associated item during test, and on studied associated pairs, by testing them either as singles or re-paired with a new associated item during test. Both pictures and words served as target stimuli. In Experiment 2, we first obtained ratings of association strength for the two types of associated pairs used in Experiment 1; then we evaluated the effect of context change on studied pairs but now with pair relation as a variable. In Experiments 1 and 2, discrimination and bias were measured under both of the two preferred recognition memory models—two-high-threshold theory, and signal detection theory with the intersection bias measure.

EXPERIMENT 1

Method

Materials and Design. All 260 pictures from the Snodgrass and Vanderwart (1980) set were used as stimuli. These pictures are simple black and white line drawings of common objects representative of a variety of categories (e.g., animals, clothing, vehicles, tools). The words used in this study were the names of the pictures. The pictures and their corresponding names were distributed equally in the study and test phases of the experiment; however, no item appeared as both a word and a picture in either the study or the test phase.

The black and white line drawings were reproduced and mounted on white 3×5 cards. The words were typed in uppercase orator type, reproduced and mounted on white 3×5 cards. The pictures and words were approximately equal in size.

From the set of 260 nouns, a set of 120 associated pairs was created to serve as the experimental stimuli. The remaining 20 items were used as filler items. The nouns were paired according to in-

tuitive criteria set by the experimenters. The association of the pairs could be categorical, such as violin-guitar, or the association could be an event, such as flower-watering can. Strength of association was not measured.

The order of each item in a pair determined the position of the items on the study and test cards. A left-hand member of a pair always appeared left of center on the cards, whereas a right-hand member always appeared right of center. Thus, single items always appeared to the right or left of center. The position of items within pairs for the study and test phases conformed to this original placement.

Sixty experimental words were presented for the study phase, 24 as singles and 36 as pairs. This resulted in a total of 42 study cards containing words (24 contained single words and 18 contained pairs). Sixty pictures were also presented for the study phase, 24 as singles and 36 as pairs, for a total of 42 study cards containing pictures. In addition to the 84 cards containing the experimental items, the study set contained 14 filler cards, 4 singles, and 3 pairs of words and pictures. Seven filler cards were presented at the beginning and 7 at the end of the study phase, to eliminate possible primacy and recency effects. The filler items were not presented for the recognition test.

The test set contained all 120 old words and pictures, plus 60 new words and 60 new pictures. The test set consisted of 168 cards, half of which contained words and half pictures. The 84 word cards contained 48 single words (24 old and 24 new) and 36 word pairs for a total of 72 words, half old and half new. The 36 test pairs were divided as follows: 6 old-old pairs, 6 new-new pairs, and 24 old-new pairs. The 84 picture cards had a distribution identical to that for words.

Table 1 illustrates the design of the experiment for words. The design for pictures is not shown, since it is identical to the design for words. There were two context categories for the study phase: (1) words studied in the absence of context as singles, and (2) words studied in the presence of context as pairs. When a pair was presented, each item served as a to-be-remembered target and as a context for the other member of the pair.

There were five context categories in the test phase: (1) word singles studied as singles and tested as singles (same context: singles); (2) word singles studied as singles and tested within a new context, with an associated distractor word (new context: singles); (3) word pairs studied as pairs and tested within the original context as pairs (same context: pairs); (4) word pairs studied as pairs and tested as single items (no context: pairs); and (5) word pairs studied as pairs and tested within a new context with addition of new associated distractor items (new context: pairs). Table 1 shows the old-new pairs twice, once as old items and once as new items.

Table 1 also illustrates the basis of the hit and false alarm rates for the five test categories. As in Tulving and Thomson (1971), false alarm rates were based on "yes" responses to new items that occurred in the same context as the corresponding old items. For both of the single test conditions, a "yes" response to the single old item was defined as a hit, and a "yes" response to a single new item was the corresponding false alarm. For the same context condition for pairs, a "yes" response to either item of an old pair constituted a hit (so "yes" responses to both items produced two hits), and a "yes" response to a new item paired with an old item was the corresponding false alarm. For the final two test conditions, in which test items were pairs presented within a new context, a "yes" response to the old member of a pair constituted a hit, and a "yes" response to either member of a new-new pair constituted the corresponding false alarm.

Eight different sets of study and test cards were created by counterbalancing pairs across stimulus form (pictures and words) and context condition (same context, no context, new context, and completely new), and by counterbalancing singles across stimulus type and context (same context, new context, and completely new). Be-

Table 1
Design of Experiment 1 for Words

Study Category	Test Category		
	Singles	Same Pairs	New Pairs
old singles	old ^a		old-new ^b new-old ^b
(24)	(12)		(12/12)
old pairs	old ^d	old-old ^c	old-new ^c new-old
(36)	(12)	(12)	(12/12)
new items	new	old-new new-old	new-new
(60)	(24)	(24/24)	(12)

Note—The design for pictures is not shown as it is identical to the design for words. The number of words presented in each study and test condition is shown in parentheses. The false positive rates for each condition are based on the new items in the corresponding column. Context conditions: ^asame context: singles. ^bnew context: singles. ^csame context: pairs. ^dno context: pairs. ^enew context: pairs.

cause context change is nested within study category (singles vs. pairs), words and pictures were not counterbalanced across pairs and singles. The filler items remained constant for all eight sets.

Subjects. Although 35 subjects participated in the experiment, the data from 3 subjects whose false alarm rates exceeded their hit rates were discarded, leaving a total of 32 subjects. The subjects were volunteers fulfilling a course requirement for an introductory psychology course. Each subject was randomly assigned to one of the eight study-test sequences resulting in 4 subjects per sequence. The design was within subjects; each subject was exposed to all five test conditions.

Procedure. Prior to the start of the study phase, the subjects were greeted and asked to sit facing the experimenter. The subjects were instructed that they would be shown words and pictures, and that some words and pictures would be shown as singles and some would be shown as pairs. The subjects were advised to pay close attention to the items presented, because their memory for the words and pictures would be tested later. The subjects were also instructed that when items were presented in pairs, the items were related in some way. Discovery of that relationship might help them recognize the items during the memory test.

Following the instructions, the study phase proceeded thus: The experimenter manually presented the complete study set one card at a time, with the 98 study cards randomized separately for each subject: the rate of presentation for cards that displayed a single item was about 1 sec, and for cards that displayed a pair of items, it was about 2 sec. Immediately following the study phase, the subjects were given a 10-min distractor task that consisted of solving problems from the space relations subtest of the Differential Aptitude Tests.

Immediately following the distractor task, the subjects were given the test phase of the experiment. They were instructed that they would be shown pictures and words as singles or pairs. Some of the singles would be old and some would be new. When pairs of items were presented, both items could be old, both items could be new, or one item of the pair could be old and one could be new. The subjects were instructed to classify single items as old or new and to classify each member of a pair as old or new, responding from left to right. The 168 test cards were randomized for each subject. The test cards were also in a random order with respect to the study cards.

The experimenter sat facing the subject and presented the test cards manually, one card at a time. The subject was allowed as much time as he or she needed to respond to each of the test cards. As the subject responded to each item presented as old or new, the

experimenter recorded the subject's responses on a data sheet. At the end of the experiment, the subjects were thanked and debriefed.

Results

As noted earlier, Snodgrass and Corwin (1988) found that only two models of yes/no recognition, two-high-threshold theory and signal detection theory with the intersection bias measure C , have measures of response bias that are independent of discrimination. The discrimination and bias measures for the two-high-threshold theory are P_r and B_r , respectively. P_r is the difference between the hit and false positive rates and is identical to the corrected recognition scores used by many investigators (e.g., Tulving & Thomson, 1971). B_r is defined as the probability of saying "yes" to an item while one is uncertain; it is expressed in terms of the hit and false alarm rate (FA) as $B_r = FA/(1 - P_r)$. For B_r , a completely neutral response bias has a value of .5, a value of less than .5 indicates a conservative response bias, and a value of greater than .5 indicates a liberal response bias.

The discrimination and bias measures for signal detection theory are d' and C (rather than β). The discrimination measure d' is defined as the distance between the means of the old and new distributions in units of their common standard deviation. This bias measure C is defined as the distance of the criterion from the intersection of the old and new distributions in units of their common standard deviation. A neutral bias has a zero value of C , a conservative bias has a positive value of C , and a liberal bias has a negative value of C .

Hit and false alarm rates were computed by adding .5 to each frequency and dividing by $N+1$ (N is the number of old items), to eliminate hit rates of 1.0 and false alarm rates of 0 (see Snodgrass & Corwin, 1988).

Table 2 presents the mean proportion of hits and false alarms for paired and single items, along with the discrimination indices P_r and d' and the bias indices B_r and C . For both hits and false alarms, pictures are recognized better than words. For hits, pairs are recognized better than singles, and items tested in the same context are recognized better than items in a changed context. The results presented in Table 2 also show that across context change, false alarm rates either remain the same or decrease (i.e., changes in the false alarm rates do not "mirror" changes in the hit rates). Consequently the bias measures B_r and C show an increase in conservatism as the context at test becomes progressively less similar to the study context. This progression toward more conservative bias is seen for pictures and words and for singles and pairs. In addition, both discrimination indices show that pictures are recognized better than words, pairs are recognized better than singles, and items tested in the same context are recognized better than items tested in a changed context.

Analysis of single pictures and words. To assess the main and interactive effects on the four dependent variables (P_r , B_r , d' , and C) four 2 (form) \times 2 (context) within-subjects analyses of variance were performed on

Table 2
Performance Measures for Study-Test Conditions of Experiment 1

Stimulus-Context	Measure					
	Raw Measures		Discrimination		Bias	
	H	FA	P_r	d'	B_r	C
	Singles					
Words-same	.63	.17*	.45	1.38	.33	.33
Words-new	.57	.17†	.40	1.28	.29	.45
Pictures-same	.82	.13‡	.69	2.26	.46	.08
Pictures-new	.78	.11§	.67	2.21	.34	.24
	Pairs					
Words-same	.73	.25	.49	1.50	.51	.01
Words-no	.62	.17*	.44	1.33	.31	.36
Words-new	.59	.17†	.42	1.35	.30	.41
Pictures-same	.87	.13	.74	2.44	.54	.03
Pictures-no	.82	.13‡	.69	2.24	.46	.09
Pictures-new	.78	.11§	.67	2.22	.35	.23

Note— H = hit rate; FA = false alarm rate. P_r and B_r are based on a two-high-threshold model of recognition memory; C is the intersection bias measure from SDT. Hits and false alarms have been corrected to avoid 1s and 0s. Identical superscripts on the false alarm rates indicate identical conditions.

single items. All four measures showed a significant difference between pictures and words; only the bias indices showed a significant effect of context; and none showed a significant interaction.

Both discrimination indices showed that pictures were recognized better than words [for P_r , $F(1,31) = 65.57$, and for d' , $F(1,31) = 76.32$, both $ps < .001$]. Neither measure showed a significant effect of context [for P_r , $F(1,31) = 3.03$, and for d' , $F < 1$]. The interaction between form and context was not significant for either index.

Both bias indices showed that subjects responded with a more conservative response bias to words than to pictures [for B_r , $F(1,31) = 9.80$, and for C , $F(1,31) = 15.51$, both $ps < .001$]. Both bias indices also demonstrated a more conservative response strategy for new context items than for same context items [for B_r , $F(1,31) = 9.08$, and for C , $F(1,31) = 10.39$, both $ps < .01$]. The interaction between form and context was not significant for either index.

Analysis of paired pictures and words. Four 2 (form) \times 3 (context) within-subjects analyses of variance were performed on the four dependent variables for pairs. All four measures showed a significant difference between pictures and words, all but d' showed a significant effect of context, and none showed a significant interaction.

Both discrimination indices P_r and d' showed that pictures were recognized better than words [$F(1,31) = 66.88$ and $F(1,31) = 67.67$, respectively, both $ps < .001$]. P_r showed a marginally significant effect of context [$F(2,62) = 2.90$, $p = .06$], while d' did not [$F(2,62) = 2.20$]. The interaction between form and context was not significant for either measure.

The analysis of bias indices B_r and C demonstrated that, as for single items, subjects responded with a significantly

more conservative response strategy to words than to pictures [$F(1,31) = 8.31$, and $F(1,31) = 9.99$, respectively, both $ps < .01$]. Both bias indices also showed a significant change in response bias from same context, to no context, to new context. The subjects responded to items tested in the same context with a neutral bias, whereas they became increasingly more conservative as context was first removed and then changed [for B_r , $F(1,31) = 9.08$, and for C , $F(1,31) = 10.39$, both $ps < .001$]. For B_r , the interaction between form and context was not significant [$F(2,62) = 2.93$]. For C , however, the interaction was significant [$F(2,62) = 3.69$, $p < .05$], such that the response strategy for the no context and new context conditions was more conservative for words than for pictures.

In summary, for both singles and pairs, pictures showed their usual robust advantage over words in discrimination, but providing the same context at test as at study did not improve discrimination for single items and only marginally improved discrimination for pairs. Response bias was more conservative for words than for pictures and more conservative for changed than for same context.

Discussion

In this experiment, context manipulation produced impaired recognition performance only for picture or word pairs, and only when the effect was measured by P_r . Snodgrass and Corwin (1988) also report cases in which P_r appears to be more sensitive to expected effects on memory than d' . They suggest that because any discrimination measure is model-based, and because no generally accepted model of recognition memory exists, a model-free test of effects should be based on statistical support from both models. The context change effect thus fails to pass their model-free test. Thus, although the effects of context change were significant and equal for pictures and words, thereby supporting the encoding specificity position, the fact that context change reduced discrimination statistically for only one of the two measures of discrimination makes us hesitate to draw firm conclusions from the results of Experiment 1 alone.

In contrast, there is no question about the robustness of the effects of context change on each of the two bias measures, B_r and C . The subjects responded to same context items with a fairly neutral response bias but adopted an increasingly more conservative bias as context was first removed and then changed. Although Thomson (1972) commented on the changes in bias that subjects adopt when context is changed, he interpreted the bias change as yet another indication that more information is retrieved when the study context is reinstated at test. Our interpretation of the bias effects is more in line with traditional views: We believe that subjects are affected by the overall familiarity of the complex of information present at test. When both items are old, this produces an increase in yea-saying bias, but when one member of the complex is new, subjects are more likely to call the paired old item

new. We will return to a fuller discussion of the reason for these bias effects at the end of this paper.

EXPERIMENT 2

Only the related pairs in Experiment 1 showed reduced discrimination when context was changed. In order to explore this effect further, and to examine what effect type of relationship might have on context change effects, a second experiment was performed.

In creating the related pairs for Experiment 1, we were constrained by the stimulus set (260 pictures or picturable words) and thus were unable to use associative norms. In constructing pairs under these constraints, it became obvious that two types of relations could be formed—category or event. In the first type, members of a pair were related by belonging to the same category (arm-leg), whereas in the second type, pair members were related by belonging to an action-outcome sequence (hand-glove). In Experiment 1, changes across type of pair (i.e., from a category pair at study to an event pair at test) as a function of context change were not controlled. So that we could examine the effects of type of relations on discrimination and bias as a function of context change, Experiment 2 consisted of a rating study followed by a context change study. The purpose of the rating study was to obtain subjective ratings on the association strength for the two types of pairs used in Experiment 1. Once the ratings were obtained, the context change study was designed to systematically vary the relation between words or pictures presented as pairs.

Rating Study

Method

Materials. Sets of four concepts were selected from the Snodgrass and Vanderwart (1980) set, such that one pairing of the four concepts would result in two category pairs (e.g., hand-foot and shoe-glove) while the opposite pairing would result in two event pairs (e.g., hand-glove and shoe-foot). From the 260 concepts in the picture set, 40 sets of concept quadruplets were formed (a total of 160 concepts), which, when paired in each of the two possible ways, formed 160 pairs. The 160 pairs were presented as both picture pairs and word pairs, for a total of 320 pairs.

The 320 pairs were divided into four forms of 80 pairs each. The division was made so that an equal number of categorical and event pairs occurred in each form and no concept appeared as both a word and a picture or as a member of both a category and an event pair within a form.

Each form of the ratings task was presented in a 10-page booklet. Each page of the booklet presented eight pairs, two categorically related picture pairs, two event-related picture pairs, two categorically related word pairs, and two event-related word pairs. The order of each type of pair on a page was random, with the constraint that each row of two items contain one picture and one word pair, and the order of pairs from row to row alternate between picture pair left/word pair right and word pair left/picture pair right.

Subjects and Design. A total of 60 subjects participated in the rating task, 15 assigned to each form. The rating task was used as a 10-min distractor task in an unrelated experiment. The subjects were volunteers fulfilling a course requirement for an introduc-

tory psychology course. Each subject was randomly assigned to one of the four test forms.

Procedure. Prior to the presentation of the rating task, the subjects were instructed that they would be asked to make judgments of relatedness for pairs of pictures and words on a 5-point scale, where 1 = slightly related, and 5 = very related. The subjects were urged to try to use the full range of ratings values. They were further instructed that all pairs were related in some way or another, but that some would appear to be more related than others, and that the words should be understood in their concrete senses—that is, as concepts that could be represented by simple pictures.

Immediately following the instructions, the subjects were given the test booklets. The subjects were free to examine all of the pairs before making their rating judgments. The subjects recorded their rating judgments directly beneath each pair in the booklets.

Results and Discussion

Table 3 presents the mean ratings of category and event-related word and picture pairs. Regardless of whether the concepts were presented as pictures or words, category pairs were judged to be more highly related than event pairs. Table 3 also shows that within each type of relationship, picture pairs were rated as slightly more highly associated than word pairs.

The analysis of the rating data was based on items rather than subjects. An item quadruplet served as the unit of analysis, such that the average rating across the two category pairs within a quadruplet was compared with the average rating across the two event pairs within the same quadruplet for pairs presented as pictures and as words. This produced four ratings for each of the 40 quadruplets (category pictures, category words, event pictures, and event words) matched by the four concepts within each quadruplet.

The results of a 2 (relation) \times 2 (form) within-items analysis of variance yielded significant effects for both relation [$F(1,39) = 7.08, p = .01$] and form [$F(1,39) = 5.76, p = .02$]; the interaction was not significant [$F(1,39) = 2.04$]. Thus, category pairs were rated as more highly associated than event pairs, and picture pairs were rated as more highly associated than word pairs. The pairs and their ratings are shown in the Appendix.

Context Change Study

Method

Materials and Design. The stimuli consisted of the 160 pairs of rated concepts plus 12 pairs of concepts used as filler items. As in Experiment 1, pictures and their corresponding names were distributed equally in the study and test phases of the experiment, and no item appeared as both a word and a picture in either the study or the test phase.

Both the study and test sequences were presented by means of an Apple Macintosh Plus microcomputer. The line drawings were

Table 3
Relatedness Judgments for Categorical and Event Word and Picture Pairs in Experiment 2

Condition	Picture Pairs	Word Pairs	<i>M</i>
Category pair	3.11	2.97	3.04
Event pair	2.69	2.66	2.67
<i>M</i>	2.90	2.81	

Table 4
Design of Experiment 2 for Words

Study Condition	Sample Pairs	Test Condition	
		Same Context	New Context
Category pair	arm-leg	arm-leg	leopard- <i>coat</i>
	leopard- <i>alligator</i> <i>skirt-blouse</i>	(apple- <i>grapes</i>)	(<i>scissors-dress</i>)
Event pair	hand-glove	hand-glove	apple- <i>grapes</i>
	apple- <i>pig</i> <i>scissors-dress</i>	(leopard- <i>coat</i>)	(<i>skirt-blouse</i>)

Note—Two pairs are shown in each test condition. The first pair provided the hit rates for the condition and the second pair (in parentheses) provided the false alarm rates. Old words are in roman type; new words are italicized.

prepared for presentation in the following manner: Each drawing was reduced to the desired size, digitally read into the computer with the Thunderscan digitizer, and stored as a MacPaint file. Each picture filled a square approximately 38 mm on a side (246 \times 246 pixel units). The words were typed in uppercase 24-point New York type and stored. The words filled a rectangle that varied from 30 to 38 mm. Thus, the pictures and words were approximately equal in size on the Macintosh monitor.

The experimental stimuli for the study sequence consisted of 40 pairs of items, 20 picture pairs and 20 word pairs of which 10 pairs were categorically related and 10 pairs were event-related. During the study presentation, the subjects saw a total of 52 pairs of stimuli, the 40 experimental pairs and an additional 12 filler pairs, half of which were presented at the beginning and half at the end of the study phase to absorb primacy and recency effects. The filler items were representative of the composition of the experimental set but were not presented for the recognition test. The recognition test stimuli consisted of the 80 old words and pictures, plus 40 new words and 40 new pictures.

As in Experiment 1, the order of each item in a pair determined the position of the item during the study and test presentations. A left-hand member of a pair always appeared left of center on the screen, whereas a right-hand member of a pair always appeared right of center on the screen.

Table 4 illustrates the design of the experiment for words. The design for pictures is not shown, since it was identical to the design for words. There were two types of pairs and one context category for the study presentation: (1) Words were studied within the presence of context as category pairs, and (2) words were studied within the presence of context as event pairs. Each member of a pair served both as a to-be-remembered target and as a context for the other member of the pair.

There were four test conditions: (1) words studied and tested as category pairs (same context); (2) words studied and tested as event pairs (same context); (3) words studied as category pairs and tested as event pairs (new context); and (4) words studied as event pairs and tested as category pairs (new context). Table 4 also illustrates the basis of the hit and false alarm rates for the four test conditions. As in Experiment 1, the hit and false alarm rates were defined with respect to the test contexts. For same context conditions (old-old pairs), a "yes" response to either item constituted a hit. So "yes" responses to both members of the arm-leg pair were counted as two hits for same context category pairs, and "yes" responses to both members of the hand-glove pair were counted as two hits for same context event pairs.

For new context conditions (old-new or new-old pairs), a "yes" response to the old member of the pair constituted a hit. So a "yes" response to leopard in the pair leopard-*coat* was counted as a hit for new context category pairs, and a "yes" response to apple in the pair apple-*grapes* was counted as a hit for new context event pairs. A "yes" response to the new member of new context pairs

was counted as a false alarm for the old context pairs having the same pair relation. So, as is shown in Table 4, a "yes" response to *coat* in the leopard-*coat* pair was counted as a false alarm for same context event pairs, and a "yes" response to *grapes* in the apple-*grapes* pair was counted as a false alarm for same context category pairs.

Finally, false alarms for new context (old-new or new-old) pairs were determined by "yes" responses to new-new pairs having the same pair relation as the new context pairs. So "yes" responses to either member of the pair *scissors-dress* were counted as false alarms for new context category pairs, and "yes" responses to either member of the pair *skirt-blouse* were counted as false alarms for new context event pairs. Note that because context was changed by re-pairing a category pair item so as to make it an event pair and vice versa, the pair relation for new context pairs used for hit and false alarm rates is opposite to the studied relation. This assignment is based on the principle that in order to determine appropriate false alarm rates, it is necessary to match the surface characteristics of the test pairs.

So that the reader can be assured that this scheme works, imagine a completely insensitive subject who responds positively to all category pairs, and negatively to all event pairs. The hit rate for same context (old-old) category pairs will be 1.0, but so will the false alarm rate. The hit rate for new context category pairs will be 0, but so too will the false alarm rate. Conversely, the hit rate for same context event pairs will be 0, but so will the false alarm rate. And the hit rate for new context event pairs will be 1.0, but so will the false alarm rate. Accordingly, discrimination will be constant (at 0) across both types of pairs and context conditions, in conformity with the assumption of this example that the subject's discrimination is at zero.

To control for possible item effects, eight different sequences of study and test sets were created, so that pairs were counterbalanced across the two stimulus forms, context conditions, and relationships. The filler items remained constant for all eight sets.

Subjects. A total of 36 subjects participated in the experiment. Of these, the data of 4 were discarded because their data were not recorded properly, leaving a total of 32 subjects whose data were analyzed. The subjects were volunteers fulfilling a course requirement for an introductory psychology course. Each subject was randomly assigned to one of the eight study-test sequences, resulting in 4 subjects per sequence. The design was within subjects; each subject was exposed to all eight test conditions.

Procedure. The instructions for the experiment were presented on the monitor of the Macintosh. The subjects were instructed that they would be shown pairs of words and pairs of pictures via the Macintosh computer. They were instructed to pay close attention to the pairs, because their memory for them would be tested on a subsequent recognition test. The subjects were also instructed that the members of each pair were related in some manner and that discovery of the relationship might help them to recognize the items during the memory test.

Following the instructions, the study presentation proceeded as follows. First, the subjects were instructed to press the return key when they were ready to start viewing the study pairs. Once the return key was pressed, the 52 pairs of pictures and words (6 distractor pairs followed by 40 experimental pairs followed by 6 additional distractor pairs) were presented one at a time for 2 sec each. The set of 40 experimental pairs was presented in a quasirandom order, different for each subject, with the constraint that picture and word pairs alternate. Immediately following the study phase, the subjects were given a 2-min paper-and-pencil cancellation-of-nines distractor task.

Immediately following the distractor task, the subject was given the test sequence. The subjects were instructed that they would be presented with a series of pairs of pictures or pairs of words, one pair at a time via the Macintosh. Some of the pairs would be old,

some of the pairs would be new, and some pairs would contain a combination of old and new items. The subjects were instructed to respond "old-old" if both members of a pair were old, "new-new" if both members of a pair were new, "old-new" if the left-hand member was old and the right-hand member was new, and "new-old" if the left-hand member was new and the right-hand member was old.

The sequence of events for the test phase was as follows. First the message "Press the return key when you are ready to begin" appeared on the screen. Once the return key was pressed, the test pairs were presented one pair at a time. The sequence of test pairs was randomized separately for each subject and was random with respect to the study sequence. When each test pair was presented, the four possible responses (old-old, old-new, new-old, and new-new) appeared at the bottom of the screen. The subjects made their responses by clicking on their response choices with the mouse and then clicking on a button on the bottom of the screen labeled "OK" to go on to the next pair of items. They could change each response choice by clicking on another button before clicking "OK." The subjects were given as much time as they needed to respond to each of the 80 test pairs.

Results

As in Experiment 1, in order to examine the possibility of changes in response bias as a result of context manipulation, the discrimination and bias indices of the two-high-threshold and signal detection theories were examined.

Table 5 shows hit and false alarm rates, the two discrimination measures, and the two bias measures for each context condition. For both hits and false alarms, picture pairs were recognized better than word pairs. Changing the context between study and test produced a reduction in hits and an increase in false alarms (the mirror effect) for all conditions except category-new words. Both discrimination indices demonstrated that picture pairs were recognized better than word pairs and that changes in context reduced recognition performance.

Both bias indices demonstrated that subjects utilized a more conservative response strategy for words than for

Table 5
Performance Measures for Study-Test Conditions of Experiment 2

Relation-Context	Measure					
	Raw Measures		Discrimination		Bias	
	<i>H</i>	<i>FA</i>	<i>P_r</i>	<i>d'</i>	<i>B_r</i>	<i>C</i>
	Words					
Category-same	.68	.21	.46	1.40	.40	.19
Category-new	.59	.19	.40	1.26	.31	.37
Event-same	.73	.19	.54	1.76	.43	.14
Event-new	.60	.22	.38	1.13	.35	.29
	Pictures					
Category-same	.84	.12	.73	2.47	.47	.07
Category-new	.76	.14	.62	2.03	.39	.18
Event-same	.89	.16	.73	2.49	.61	-.15
Event-new	.85	.20	.65	2.12	.57	-.11

Note—*H* = hit rate; *FA* = false alarm rate. *P_r* and *B_r* are based on a two-high-threshold model of recognition memory; *C* is the intersection bias measure from signal detection theory. Hits and false alarms have been corrected to avoid 1s and 0s.

pictures. Although in Experiment 2 the pattern of hits and false alarms across context change exhibited the mirror effect for three of the four pair types, the magnitude of the increase in false alarm rates when hit rates decreased was not large enough to produce constant bias. Both bias measures showed increases in conservatism when context was changed for all four pair types.

Four 2 (form) \times 2 (context) \times 2 (relation) within-subjects analyses of variance were performed on the four measures of discrimination and bias (P_r , B_r , d' , and C). All four measures showed a significant picture advantage and a significant same context advantage. Relation was significant only for the two bias measures, B_r and C .

For P_r , the picture advantage yielded an $F(1,31) = 74.35$, $p < .001$, and the same context advantage yielded an $F(1,31) = 31.50$, $p < .001$. The main effect of relation was not significant [$F(1,31) = 1.14$]. None of the interactions was significant.

For d' , the picture advantage yielded an $F(1,31) = 79.33$, $p < .001$, and the same context advantage yielded an $F(1,31) = 27.74$, $p < .001$. The main effect of relation was not significant [$F(1,31) = 1.58$]. None of the interactions was significant.

For the bias index B_r , words produced more conservative response bias than did pictures [$F(1,31) = 28.58$, $p < .001$], and changed context produced more conservative response bias than did same context [$F(1,31) = 7.36$, $p < .05$]. In addition, category pairs produced more conservative response bias than did event pairs [$F(1,31) = 22.23$, $p < .001$]. The interaction between form and relation was also significant [$F(1,31) = 8.26$, $p < .01$], such that event pictures were responded to with a liberal response strategy, whereas category pictures and event and category words demonstrated a fairly conservative response strategy. The remaining interactions were not significant.

For the bias measure C , the main effects of form and context were also significant [$F(1,31) = 29.98$, $p < .001$, and $F(1,31) = 8.36$, $p < .001$, respectively]. As with B_r , the main effect of relation was significant [$F(1,31) = 21.02$, $p < .001$], as was the interaction between form and relation [$F(1,31) = 5.89$, $p < .05$]. The remaining interactions were not significant.

Discussion

Experiment 2, like Experiment 1, showed the expected picture superiority effect. However, unlike Experiment 1, Experiment 2 showed a robust decrement in recognition performance with context change for both measures of discrimination. Experiment 2, then, passes the Snodgrass-Corwin test for model-independent effects. Type of relationship had no effect on recognition performance. Thus, although categorical pairs were judged to be more highly related than event pairs, subjects recognized category and event pairs equally well when intact, and suffered the same decrement in recognition when context was changed. This is consistent with the results of Thomson (1972) and Tulving and Thomson (1971), who showed that strength

of association did not affect the magnitude of context change effects for the discrimination measure P_r .

In Experiment 2, as in Experiment 1, pictures were as affected by context change as words were, so the decrement in recognition of words in a changed context cannot be produced by a change in meanings as proposed by semantic theory.

Response bias in this experiment was significantly affected by context change. Interestingly, and somewhat inexplicably, it was only in the response bias measures that type of relationship had an effect. Here subjects showed a somewhat liberal response bias toward pictures presented as event pairs, in contrast with the consistently conservative bias they showed toward category pictures and to all word pairs. Why this should be remains a puzzle to us.

GENERAL DISCUSSION

Many investigators treat bias effects as nuisance variables and either attempt to eliminate bias by using forced-choice rather than yes/no designs, or simply base their conclusions on bias-free measures of discrimination. We believe, on the other hand, that bias effects can be as important to understanding the recognition memory process as discrimination effects. As long as it is possible to measure bias independently of discrimination, we believe there is merit in measuring bias in yes/no designs. However, the measurement of bias is based on two considerations. The first is the selection of the correct false alarm rate for comparison with each hit rate in this type of experiment, and the second is the selection of an appropriate bias measure.

In experiments in which study and test contexts are the same, hit and false alarm rates are determined by study category and thus are paired in a straightforward manner. An item is either presented for study or not presented and subsequently at test is either old or new. However, in experiments that manipulate the context in which a studied item is tested, determination of the appropriate hit and false alarm pairs becomes more complicated. When context is manipulated, there are several ways an old item may be viewed at test. For instance, a pair of words studied as a pair can be tested as an intact pair in the same context as study, or each word of the studied pair can be tested separately, as singles with context deleted, or, finally, the study pair can be tested with the addition of a new context word in an old-new, new-old pair. In each case, the hit rate is generated by the identification of the old item as old.

Although generation of the hit rate in experiments that manipulate test context is relatively unambiguous, selection of the appropriate false alarm rate may not be as intuitively obvious. The choice of corresponding false alarms in this type of experiment must be based on the surface characteristics of the items used to generate the hit rates. For instance, if a test context of a pair of items is an old item paired with a new item, the hit rate is gener-

Table 6
Summary of Performance Measures for Six Experiments and for Seven Types of Stimuli

Stimulus-Context	Measures			
	Raw Measures		Discrimination P_r	Bias B_r
	H	FA		
Unitized photographs of faces-Same ^a	.81	.17	.64	.47
Unitized photographs of faces-No	.67	.13	.54	.28
Unitized photographs of faces-New	.75	.12	.63	.32
Unitized photographs of faces-Same ^b	.81	.21	.60	.52
Unitized photographs of faces-No	.72	.14	.58	.33
Unitized photographs of faces-New	.70	.18	.52	.37
Nonunitized photographs of faces-Same ^c	.84	.17	.67	.51
Nonunitized photographs of faces-No	.73	.16	.57	.37
Nonunitized photographs of faces-new	.81	.16	.65	.46
Line drawings-Same ^d	.87	.13	.74	.53
Line drawings-No	.82	.13	.69	.46
Line drawings-New	.78	.11	.67	.35
Strongly associated words-Same ^e	.88	.29	.59	.70
Strongly associated words-no	.70	.24	.46	.44
Strongly associated words-New	.66	.16	.50	.32
Weakly associated words-Same ^f	.85	.23	.62	.60
Weakly associated words-No	.77	.24	.53	.51
Weakly associated words-New	.58	.21	.37	.59
Words-Same congruous sentence ^g	.93	.19	.74	.73
Words-Different congruous sentence	.75	.16	.59	.39
Words-Different incongruous sentence	.69	.18	.51	.37
Words-Same incongruous sentence ^h	.93	.12	.81	.63
Words-Different incongruous sentence	.82	.18	.64	.50
Words-Different congruous sentence	.78	.16	.62	.42
Words-Same congruous sentence ⁱ	.83	.22	.61	.56
Words-Different congruous sentence	.71	.15	.56	.34
Words-Different incongruous sentence	.62	.22	.40	.37
Words-Same incongruous sentence ^j	.85	.21	.64	.58
Words-Different incongruous sentence	.77	.22	.55	.49
Words-Different congruous sentence	.67	.15	.52	.31

Note— H = hit rate; FA = false alarm rate. P_r and B_r are based on a two-high-threshold model of recognition memory and have been computed from the average hit and false alarm rates.
^aWinograd and Rivers-Bulkeley (1977, Experiment 1). ^bWinograd and Rivers-Bulkeley (1977, Experiment 2). ^cWinograd and Rivers-Bulkeley (1977, Experiment 2). ^dThis paper, Experiment 1. ^eTulving and Thomson(1971). ^fTulving and Thomson (1971). ^gMorris (1978, Experiment 1). ^hMorris (1978, Experiment 1). ⁱMorris (1978, Experiment 2). ^jMorris (1978, Experiment 2).

ated by correct identification of the old member of the pair as old, and the false alarm rate is generated by responses of "old" to a new item paired with a new item. In this example, hit and false alarm rates were determined by their respective test categories and were based on identical test contexts.

As an example of the pitfalls inherent in experiments that manipulate context, we will describe an experiment by Morris (1978). Subjects studied target words in sentences that provided either congruous or incongruous contexts; tests were of target words in the same sentence context as that during study, target words in a new sentence context with the same congruous or incongruous relation as that during study, and target words in a new sentence context with the relation that was different from that during study. By the logic described above, each of these test conditions calls for a different false alarm rate. For a

studied word in a congruous context, the same context test takes the false alarm rate from new words in old congruous sentences, the new-same relation test takes the false alarm rate from new words in new congruous sentences, and the new-different relation test takes the false alarm rate from new words in new incongruous sentences. However, Morris used the same false alarm rate (the second one) to compute corrected recognition scores (P_r) for all conditions. Because false alarm rates varied rather widely (in his Experiment 1, they varied from .12 to .19), use of the inappropriate false alarm rates for two of the three test conditions served to enhance the apparent "transfer appropriate processing" produced when words studied in incongruous sentences were tested in congruous sentences.

In the experiments reported here, we have shown that a change in context between study and test usually, but

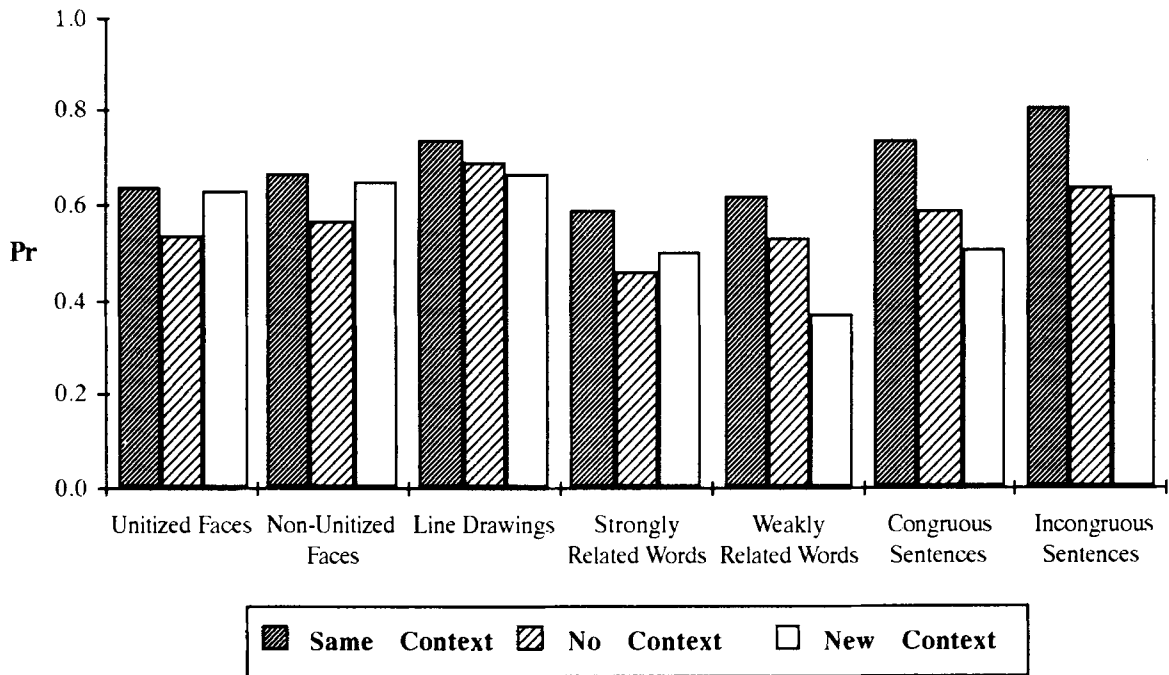


Figure 1. The effect of context change on the discrimination measure P_r for seven types of stimuli used in four experiments. Unitized and nonunitized photographs of faces: Winograd and Rivers-Bulkeley, 1977, Experiment 2. Line Drawings: this paper, Experiment 1. Strongly and weakly associated words: Tulving and Thomson, 1971. Congruous and incongruous sentences: Morris, 1978, Experiment 1.

not invariably, produces a decrease in discrimination, but always produces a change in response bias toward conservatism. Next, we will show that the same pattern of results has been obtained for a variety of stimuli in experiments using the same basic methodology. We chose as the target design the pair condition of Experiment 1, because this is a common design in the literature, and also because our results showed much smaller discrimination effects than bias effects.

Table 6 shows results from four experiments in which the same basic task as that in the experiments reported here was used with five types of stimuli: unitized photographs of faces (i.e., pairs of photographs that were studied as a "unit"; Winograd & Rivers-Bulkeley, 1977, Experiments 1 and 2), nonunitized photographs of faces (pairs of photographs that were not expressly studied as a unit; Winograd & Rivers-Bulkeley, 1977, Experiment 2), and strongly and weakly associated words (Tulving & Thomson, 1971). Results for the pairs condition for Experiment 1 in this paper are shown for comparison. Although the exact experimental conditions varied slightly from experiment to experiment, all items in each of these experiments were studied as pairs and were tested in the same context, or with context deleted as singles, or with the addition of a new context item. We computed P_r and B_r from the average hit and false alarm rates reported in each paper. We chose P_r and B_r rather than

d' and C for this analysis, because these measures are the same whether computed on individual hit and false alarm rates and then averaged or computed on the averaged hit and false alarm rates. The same is not true of d' and C . To demonstrate the importance of selecting the correct false alarm rate for comparison with each hit rate, Table 6 also includes the hit rates re-paired with the appropriate false alarm rates for Experiments 1 and 2 from Morris (1978).

Table 6 reveals several striking similarities across the four experiments. First, both the hit rates and the discrimination index P_r decreased when context was either deleted or changed. Second, the false alarm rates were generally higher for the same context condition than for the changed context conditions. Third, for each type of stimulus, there was a large change in response bias (B_r) between the same context and changed context conditions, with subjects becoming more conservative as context was removed or changed. Table 6 also shows that the re-paired hit and false alarm rates from Morris (1978) exhibit the same general pattern as do the results of other experiments in this genre. For both experiments, both discrimination (P_r) and bias (B_r) decreased as context was changed.

Figure 1 shows the effect of context change on P_r for the same seven types of stimuli, and Figure 2 presents the same results for B_r . Comparison of Figures 1 and 2 shows that for all the experiments presented, even slight

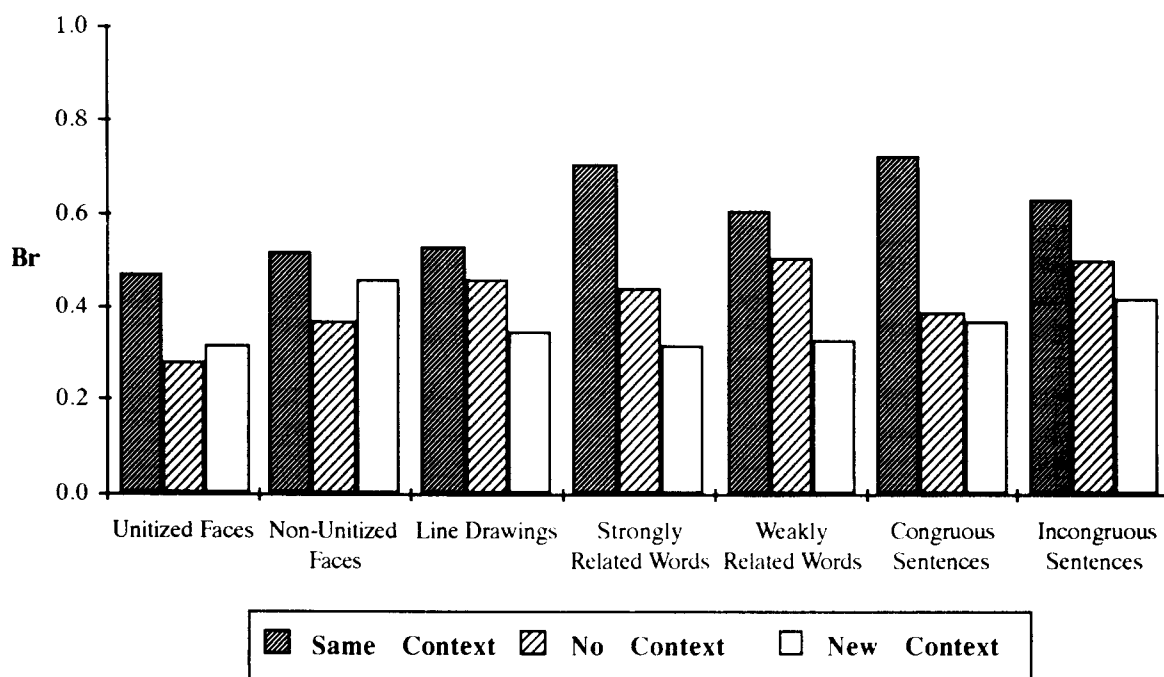


Figure 2. The effect of context change on the bias measure B_r for seven types of stimuli used in four experiments. Unitized and nonunitized photographs of faces: Winograd and Rivers-Bulkeley, 1977, Experiment 2. Line drawings: this paper, Experiment 1. Strongly and weakly associated words: Tulving and Thomson, 1971. Congruous and incongruous sentences: Morris, 1978, Experiment 1.

decrements in discrimination are accompanied by marked shifts from neutral to conservative response bias as test context is changed.

The results of Table 6 and Figures 1 and 2 demonstrate the importance of examining bias as well as discrimination in this type of experiment. In the present experiments, we found that response bias became more conservative when context was changed. As we have shown, these same bias effects can be observed across a number of experiments using a variety of materials.

Turning next to the selection of the appropriate bias measure, we have been at pains to show that as long as discrimination and bias effects can be measured independently, examination of both bias and discrimination is equally important for understanding recognition memory performance. As shown by Snodgrass and Corwin (1988), B_r and C are the bias measures of choice because they are theoretically independent of their respective discrimination measures P_r and d' . However, they may not be independent empirically. As we have shown, changes of context regularly decrease both discrimination and bias. This means that empirically they will show correlation (the sign will depend on which measure of bias is used). For example, the correlations (Pearson r) between discrimination and bias based on the pair conditions from Experiment 1 were +.314 for the P_r/B_r pair and -.323 for the d'/C pair. For Experiment 2, the correlation between P_r and B_r was +.226, and the correlation between d' and C was -.252. The r for each of the discrimination-bias pairs was significant at $p < .05$.

Tulving and Thomson (1971), who found a negative correlation between d' and β in their study, suggested that demonstrations of empirical dependence between measures of discrimination and bias invalidate their use. They were correct in their particular criticism of β , in that Snodgrass and Corwin (1988) showed that β is not theoretically independent of d' . However, they were incorrect in their general criticism: even theoretically independent measures may show empirical dependence if the experimental manipulation affects both discrimination and bias.

In conclusion, we have shown that changing context at test affects bias measures as much as or more than their associated discrimination measures and that these bias effects can be observed in experiments other than our own. We have also shown that context change effects are equally robust for stimuli with few meanings (pictures) and for stimuli with many meanings (words).

REFERENCES

- GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8-20.
- HUMPHREYS, M. S., & BAIN, J. D. (1983). Recognition memory: A cue and information analysis. *Memory & Cognition*, *11*, 583-600.
- LIGHT, L. L., & CARTER-SOBELL, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning & Verbal Behavior*, *9*, 1-11.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271.
- MARTIN, E. (1975). Theoretical notes: Generation-recognition theory

- and the encoding specificity principle. *Psychological Review*, **82**, 150-153.
- MORRIS, C. D. (1978). Acquisition-test interactions between different dimensions of encoding. *Memory & Cognition*, **6**, 354-363.
- PELLEGRINO, J. W., & SALZBERG, P. M. (1975a). Encoding specificity in associative processing tasks. *Journal of Experimental Psychology: Human Learning & Memory*, **1**, 538-548.
- PELLEGRINO, J. W., & SALZBERG, P. M. (1975b). Encoding specificity in cued recall and context recognition. *Journal of Experimental Psychology: Human Learning & Memory*, **104**, 261-270.
- REDER, L. M., ANDERSON, J. R., & BJORK, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, **102**, 648-656.
- SANTA, J. L., & LAMWERS, L. (1974). Encoding specificity: Fact or artifact? *Journal of Verbal Learning & Verbal Behavior*, **13**, 412-423.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.
- SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.
- THOMSON, D. M. (1972). Context effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior*, **11**, 497-511.
- TULVING, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.
- TULVING, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- TULVING, E., & THOMSON, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, **87**, 116-124.
- TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and the retrieval processes recognition memory. *Psychological Review*, **80**, 352-373.
- WATKINS, M. J., HO, E., & TULVING, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning & Verbal Behavior*, **15**, 505-517.
- WINOGRAD, E., & RIVERS-BULKELEY, N. T. (1977). Effects of changing context on remembering faces. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 397-405.

APPENDIX

**The 160 Pairs of Concepts Used in Experiment 2,
Along with Their Mean Association Ratings**

Category Pairs	Pictures	Words	Event Pairs	Pictures	Words
1 airplane-helicopter	4.27	4.47	airplane-sun	3.13	2.00
1 moon-sun	4.67	4.20	moon-helicopter	1.33	1.57
2 apple-grapes	3.53	4.07	apple-pig	1.53	1.47
2 fox-pig	3.07	2.40	fox-grapes	1.73	1.71
3 barn-windmill	2.87	3.80	barn-window	2.47	1.67
3 door-window	4.33	4.40	door-windmill	1.47	2.00
4 baseball bat-bicycle	2.33	2.87	baseball bat-arm	2.80	3.71
4 leg-arm	4.40	4.00	leg-bicycle	2.87	2.73
5 basket-bottle	3.20	2.13	basket-kangaroo	2.07	1.73
5 cow-kangaroo	2.73	2.07	cow-bottle	3.20	2.21
6 bed-lamp	3.00	3.73	bed-bear	1.33	1.20
6 butterfly-bear	1.60	1.47	butterfly-lamp	1.00	1.50
7 book-envelope	2.47	2.00	pen-envelope	2.73	3.40
7 pen-pencil	4.73	4.60	book-pencil	3.80	2.43
8 baby carriage-wagon	3.40	3.20	baby carriage-doll	4.27	3.87
8 clown-doll	3.73	2.73	clown-wagon	1.67	2.50
9 carrot-mushroom	3.07	2.80	carrot-rabbit	3.93	3.67
9 frog-rabbit	2.93	2.27	frog-mushroom	1.73	2.14
10 church-house	3.13	2.87	church-bell	3.80	3.80
10 key-bell	1.73	1.47	key-house	3.67	3.71
11 ear-eye	3.80	4.33	ear-record player	4.07	3.33
11 television-record player	3.93	3.87	television-eye	3.67	3.64
12 helmet-mitten	2.07	2.27	helmet-motorcycle	3.27	4.27
12 sled-motorcycle	2.53	2.40	sled-mitten	3.60	2.93
13 horse-gorilla	2.27	2.67	horse-boot	2.13	1.93
13 suitcase-boot	2.13	1.87	suitcase-gorilla	1.20	1.64
14 lion-tiger	4.27	4.20	lion-chain	2.27	1.93
14 gun-chain	2.80	1.53	gun-tiger	1.40	2.00
15 ostrich-skunk	2.13	2.13	ostrich-hair	1.20	1.13
15 nose-hair	3.27	2.87	nose-skunk	3.73	3.57
16 pocketbook-coat	3.07	2.87	pocketbook-alligator	2.60	1.87
16 leopard-alligator	2.53	2.47	leopard-coat	1.13	2.64
17 rooster-owl	3.27	2.60	rooster-clock	3.33	2.27
17 glasses-clock	1.40	1.53	glasses-owl	1.80	2.21
18 scissors-needle	3.00	3.93	scissors-dress	2.47	2.00
18 button-dress	3.47	3.47	button-needle	3.87	3.43

APPENDIX (Continued)

Category Pairs	Pictures	Words	Event Pairs	Pictures	Words
19 seal-dog	2.53	2.67	seal-ball	3.93	3.67
19 whistle-ball	3.40	1.87	whistle-dog	2.60	3.00
20 watering can-ladder	1.33	1.60	watering can-flower	4.13	3.53
20 tree-flower	3.53	4.00	tree-ladder	2.67	3.14
21 lips-finger	2.60	2.27	lips-glass	3.47	2.47
21 cup-glass	4.33	3.93	cup-finger	2.27	2.20
22 screw-nail	4.47	4.07	screw-screwdriver	4.60	4.53
22 hammer-screwdriver	4.33	3.86	hammer-nail	4.93	4.80
23 sock-blouse	2.47	2.27	sock-clothespin	2.87	2.53
23 hanger-clothespin	2.73	3.36	hanger-blouse	4.07	4.20
24 star-mountain	1.87	1.60	star-kite	1.93	2.27
24 balloon-kite	4.27	3.79	balloon-mountain	2.07	1.67
25 tennis racket-football	3.33	2.60	tennis racket-thumb	1.87	2.73
25 toe-thumb	3.67	4.29	toe-football	3.20	3.33
26 watermelon-pumpkin	3.13	3.73	watermelon-refrigerator	3.33	2.27
26 stove-refrigerator	4.13	3.57	stove-pumpkin	1.73	1.47
27 candle-flag	1.33	1.27	candle-piano	1.87	2.53
27 drum-piano	3.53	3.93	drum-flag	3.33	2.27
28 fence-leaf	2.00	1.53	fence-squirrel	1.67	2.07
28 bird-squirrel	2.60	2.79	bird-leaf	3.13	2.80
29 fork-knife	4.20	4.60	fork-potato	3.87	2.80
29 bread-potato	2.47	2.57	bread-knife	3.80	4.33
30 garbage can-barrel	3.27	2.53	garbage can-goat	1.40	1.60
30 monkey-goat	2.73	1.79	monkey-barrel	2.47	2.20
31 lobster-chicken	2.40	2.27	lobster-pot	3.07	3.00
31 frying pan-pot	4.67	3.50	frying pan-chicken	2.47	3.73
32 pitcher-broom	1.27	1.00	pitcher-camel	1.60	1.40
32 cat-camel	2.00	1.57	cat-broom	1.67	1.73
33 raccoon-donkey	2.20	2.13	raccoon-cap	1.40	2.60
33 hat-cap	4.53	4.57	hat-donkey	1.27	1.47
34 shirt-pants	4.00	4.00	shirt-dresser	3.47	3.13
34 chair-dresser	2.53	3.21	chair-pants	1.27	2.00
35 sweater-jacket	4.13	3.93	sweater-sheep	3.87	3.67
35 penguin-sheep	2.33	2.00	penguin-jacket	1.60	2.27
36 table-desk	3.80	3.73	table-paintbrush	2.20	1.80
36 ruler-paintbrush	1.73	2.29	ruler-desk	2.53	2.87
37 bee-ant	4.33	3.07	bee-cake	1.47	1.33
37 sandwich-cake	2.93	2.64	sandwich-ant	1.80	2.13
38 cigar-cigarette	4.53	4.33	cigar-vase	1.60	1.20
38 ashtray-vase	1.73	2.00	ashtray-cigarette	4.67	4.87
39 fish-seahorse	3.40	3.27	fish-sailboat	3.67	3.00
39 anchor-sailboat	4.20	4.21	anchor-seahorse	3.00	2.73
40 glove-shoe	3.27	2.53	glove-hand	4.67	4.67
40 food-hand	3.67	4.07	foot-shoe	4.33	4.73
<i>M</i>	3.11	2.97		2.69	2.66
<i>SD</i>	0.93	0.98		1.06	0.97

Note—The scale for judging the degree of association was 5 = highly related, and 1 = unrelated.

(Manuscript received April 1988;
revision accepted for publication January 5, 1990.)