

# Changing environmental context does not reliably affect memory

ANGEL FERNANDEZ and ARTHUR M. GLENBERG

*University of Wisconsin, Madison, Wisconsin*

Most current theories of human memory propose that context, defined here as the time and place at which an event was experienced, forms an integral feature of the mnemonic representation of events. One way of investigating context is by manipulating the environmental context (which typically means the room in which the experiment takes place). The predominant result of this manipulation reported in the literature has been consistent with theory: Memory performance is better when the learning and testing environments are the same than when they differ. This article reports eight experiments that in aggregate challenge the reliability of this same-context advantage. Experiment 1 reported a failure to obtain a same-context advantage. Experiments 2-7 investigated various features of the design that might have reduced the effect. None of these experiments produced a reliable same-context advantage. Experiment 8 repeated the methodology of a published report of a same-context advantage with more than double the number of subjects, but failed to replicate the effect. An analysis of features of the experiments led to two suggestions for future investigations of the effects of changes in environmental context on memory.

One of the most generally accepted propositions about mnemonic representations is that they encode multiple attributes (Bower, 1967; Underwood, 1969) rather than reflect a value on a single dimension such as strength. Indeed, much of what is stored in memory seems to consist of episodes (Tulving, 1983) or contextual markers (Anderson & Bower, 1972; Raaijmakers & Shiffrin, 1981) indicating time, place, and other conditions of the original event. Nonetheless, few analytic techniques have been devised to study which aspects of the presentation context are encoded, how they are represented, or their consequences for remembering. One apparent exception is work on the environmental context, the physical location in which an event takes place. It is the purpose of this paper to demonstrate, however, that the most popular technique for studying environmental context and memory does not produce robust or reliable results.

The proposition that presentation context is represented in memory is integral to most current theories of memory. For example, Anderson and Bower (1972) proposed that contextual markers are used to discriminate between previously presented and distractor items on recognition tests, as well as to guide retrieval on recall tests (through

marked associations as well as the "entry set" items). Glenberg (1979) proposed that context is one of three general types of information that plays an important role in producing spacing effects and recency effects (Glenberg, Bradley, Kraus, & Renzaglia, 1983). That memory for events encodes context, that is, that memory is for episodes, is basic to much of Tulving's work on the encoding specificity principle (e.g., Tulving, 1983; Flexner & Tulving, 1978). Raaijmakers and Shiffrin (1981) suggested that to-be-remembered (TBR) stimuli become associated with context, and that this context is an important component of virtually any retrieval cue. Metcalfe and Murdock (1981) proposed that initial TBR items are rehearsed with a representation of the situational context. This list does not exhaust the set of theorists utilizing context (for other examples see Kintsch, 1974, and Baddeley, 1982).

Given this unanimity of theory, one would suspect that a large and consistent set of results points directly to the conclusion that situational or environmental context is represented in memory and utilized in remembering. This appears to be the case in studies of animal memory (see Spear, 1978, for a review); in fact, situational context appears to play a critical role in conditioning-like processes such as the development of morphine tolerance (Siegal, Hinson, & Krank, 1981). There are far fewer studies on environmental-context effects in human memory, but they are widely cited. Most prominent in the recent literature on human memory are papers by Godden and Baddeley (1975) and by Smith (Smith, 1979b, 1982; Smith, Glenberg, & Bjork, 1978). The results of these studies are consistent and compelling: When the learning and testing contexts are the same, performance on recall tests is higher than when the test context is different from the learning

Data collection and preparation of this report were supported in part by NIMH Grant No. MH26643 and University of Wisconsin Graduate School Grants 120150, 131153, and 130614 to Arthur M. Glenberg, and by a Fulbright/Banco de Bilbao Scholarship to Angel Fernandez. Experiments 1-5 were extracted from Angel Fernandez's master's thesis (Fernandez, 1983) and were presented at the 1983 meeting of the Midwestern Psychological Association (Fernandez & Glenberg, 1983). Requests for reprints may be sent to either author at the Department of Psychology, University of Wisconsin, 1202 W. Johnson Street, Madison, WI 53706.

context. Recognition, however, is largely unaffected by changes in the environmental context.

In a clever series of experiments, Smith (1979b) ruled out two explanations of the decrement in memory performance associated with changed environments—that it can be solely attributed to psychological disruption (Strand, 1970), or that it is due to testing in unfamiliar places. Furthermore, Smith found that instructions to students to mentally reinstate the input context eliminated the decrement associated with recalling in rooms different from the study room, but only under conditions in which subjects could be expected to remember the study room. These are just the sort of results that directly implicate the representation and use of environmental context in encoding and retrieval.

Although the context change used by Godden and Baddeley (1975) involved moving from land to underwater, Smith's (1979b) work, as well as much of the animal work, has involved a more prosaic change from one room to another. Although in absolute terms the decrement in recall associated with testing in a room different from the study room is small, the proportion lost can be substantial (e.g., 25%). What is more, the effect is consistent with intuitions and everyday anecdotes about memory (e.g., forgetting the purpose of a trip from one room to another until reentering the original room).

We began the series of experiments reported here with the goal of investigating why changes in the environmental context (changes in room) do not disrupt performance on a recognition memory test (Smith et al., 1978). However, we failed to find the expected effects on free-recall tests as well (Experiment 1). This failure prompted a systematic search of a wide variety of cognitive factors we thought might be compensating for changes in the environmental context (Experiments 2-7). This search failed; we were never able to produce a reliable effect. The last experiment (Experiment 8) was a replication of two conditions in a previously published experiment reporting environmental-context effects (Smith, 1979b, Experiment 1). We more than doubled the number of subjects used in Smith's experiment, but once again failed to produce the effect.

Our data do not imply that the environmental-context effect is nonexistent, only that standard methods do not reliably produce it. Reasons why standard methods may be ineffective are presented in the General Discussion.

## EXPERIMENT 1

The first experiment was designed to determine why the environmental context fails to affect performance in recognition tests (Smith et al., 1978). One possibility, derived from two-process accounts of recognition memory (e.g., Mandler, 1980), was that the effect of context on recognition could be made apparent once the stored item lost strength or familiarity. We attempted to manipulate familiarity through number of presentations of the TBR items (one, two, or four presentations) and the duration

of the retention interval (1 or 7 days). To prevent rehearsal during the long retention intervals, an incidental learning procedure was used; subjects were not informed that memory was to be tested until immediately prior to the test. Our hypothesis was that the effect of context on recognition would be more salient after a longer retention interval, and that the effect would be stronger for items presented only once than for items with multiple presentations. Subjects were also tested by cued-recall and free-recall tests using the same materials and study procedures as the recognition test. These tests were included, in part, as a manipulation check to demonstrate the environmental-context effect on recall.

## Method

**Subjects.** Forty-eight students from introductory psychology classes at the University of Wisconsin—Madison participated in this experiment. They received course credit.

**Materials.** A total of 238 common nouns were used in the experiment. Of these, 108 words were used as distractors in the recognition test. The other 130 words, randomly paired, constituted the learning lists. Each pair was printed at the top of a punched card.

A learning list consisted of different subgroups of pairs. Three practice pairs were placed at the beginning of the list. Then came a primacy buffer composed of 2 pairs that were presented once and 2 pairs that were presented twice. These pairs were followed by 54 critical pairs: 18 pairs were presented once, 18 pairs were presented twice, and 18 pairs were presented four times. These pairs were distributed throughout the list. The average distance between repetitions of a pair was 7 items. Four additional pairs (2 pairs presented once and 2 pairs presented twice) formed a recency buffer at the end of the list. The pairs in the practice, primacy, and recency subgroups were the same for all subjects and were excluded from scoring and analysis procedures. Three different experimental lists were constructed so that each of the critical pairs could appear once in one list, twice in another, and four times in another. Each list was presented to a third of the subjects.

For the recognition test, we prepared a booklet containing a pair of words on each page. The pairs were of various types: Three pairs appeared as they had been presented in the learning list; three pairs were formed from words that had been presented, but not paired together, in the learning list; six pairs were formed by putting together one word from the learning list and one distractor, with the distractor preceding the old word in half of the pairs; six pairs were formed by randomly pairing distractors. For each word in a pair, subjects were to circle either the letter O (for an "old" word, i.e., a word in the learning list) or the letter N (for a "new" word, i.e., a distractor). For each pair in the booklet, space was provided for subjects to express their judgment as to whether the two words in the pair had appeared together in the learning list: subjects could respond by circling the letter Y (for a "yes" response) or the letter N (for a "no" response). A third of the old words in the recognition test were selected from words that had been presented once in the learning list, a third were selected from words that had been presented twice, and a third were selected from words that had been presented four times. There were two structurally equivalent forms of the recognition test. One form tested half of the critical pairs and the other form tested the other half. Within each form, the order of the pairs was randomly determined.

For the cued-recall test, we prepared a second booklet. Each of the 27 pages in this booklet contained a single word, always the first member of a pair presented in the learning list. Following each word, space was provided for the subjects to write the second word in the pair. A third of the cues had been presented once in the learning list, a third had been presented twice, and a third had been presented four times. There were two equivalent forms of the cued-

recall test, each testing half of the critical pairs, corresponding to the two forms of the recognition test.

**Environmental contexts.** Room A was on the second floor of the University of Wisconsin's Brogden Psychology Building. It was a large room that contained two big closets, a file cabinet, a large gray table, and four chairs. The walls were painted black and were completely uncovered. The subjects sat at the table, facing a small tape recorder, and the experimenter sat in a different place.

Room B was a much smaller room in the basement of the same building. There were pictures on the walls, plastic plants, carpeting, a large green table, and four chairs. The experimenter sat at the same table as the subjects.

**Design.** The use of the two rooms, A and B, as both study and test locations resulted in a combination of four groups of 12 subjects each. The subjects in Group AA studied the lists and were tested in Room A. The subjects in Group AB studied the lists in Room A and were tested in Room B. The subjects in Group BB studied and were tested in Room B. The subjects in Group BA studied in Room B and took the memory tests in Room A. For the analysis, Group AA and Group BB were combined into one group (same-context), and Group AB and Group BA were combined into another group (different-context).

Half of the students in each group had a retention interval of 1 day and the other half had a retention interval of 7 days. For each subject, a third of the words were presented once, a third were presented twice, and a third were presented four times. The resulting experimental situation was a  $2 \times 2 \times 3$  factorial design, with environmental context (same vs. different) and retention interval (1 day vs. 7 days) as between-subjects factors and number of presentations (one, two, or four presentations) as a within-subjects factor.

**Procedure.** Each subject participated, either individually or paired with another subject, in two experimental sessions. At the beginning of the first learning session, subjects signed a consent form and were informed of the day and the time of the second session. Then instructions were played on a tape recorder in Room A and read by the experimenter in Room B. Due to the incidental nature of the task, the goal of the study was not explained to the subjects at this point; instead, they were told that the experiment studied creativity with words. The subjects were presented with a deck of punched cards containing the stimulus pairs. The subjects were instructed to use the words in each pair to create a sentence that met four criteria: (1) the sentence was to relate the words through their meanings; (2) the order of the words in the sentence was to correspond to the order of the words on the cards; (3) both words were to be used as nouns; and (4) at least one other word was to be put between the two nouns in the pair. After thinking of a sentence they were to write it on a blank page.

The subjects were instructed that when they encountered a previously presented pair they could either write the same sentence again or make a new one. They were encouraged to work at their own pace through the deck of cards, creating a sentence for each pair and writing it on a blank page before going to the next pair of words. They saw a total of 141 pairs of words.

The test session, either 1 or 7 days later, consisted of a recognition test and a cued-recall test given in alternating order between subjects. There was no time limit on these two tests. After testing half of the subjects, we noted the absence of context effects in the cued-recall test. As a check on the presence of the basic effect, we administered a free-recall test to the second half of the subjects. The free-recall test lasted 5 min and was administered before the other tests.

## Results and Discussion

The probability of a statistical type I error was set at .05 for all the analyses.

**Recognition data.** False-alarm data were submitted to an analysis of variance. The analysis revealed a significant effect of retention interval (RI) [ $F(1,44) = 9.24$ ,  $MSe$

$= 3.35$ ]; the mean false-alarm rate after a 7-day RI was .15, whereas the mean after a 1-day RI was .07. The mean false alarm rates were .11 and .10 for the same-context and different-context conditions, respectively; the difference was not significant ( $F < 1$ ). The absence of a significant context effect on false alarms permits a direct analysis of hits.

The analysis of variance on hits showed a significant effect of retention interval [ $F(1,44) = 18.58$ ,  $MSe = 8.97$ ]; the mean hit rate after a 1-day RI was .90 and the mean after a 7-day RI was .78. Number of presentations was also significant [ $F(2,88) = 78.85$ ,  $MSe = 2.01$ ]; the mean hit rate was .73 for words presented once, .86 for words presented twice, and .92 for words presented four times. However, environmental context did not have a significant effect on hits ( $F < 1$ ). The mean hit rates were .83 and .84 for the same-context and different-context conditions, respectively. None of the interactions was significant.

The absence of a context effect and the absence of significant interactions between environmental context and the other two factors are not consistent with our initial hypothesis. Although memory for words increased as a function of the number of presentations and decreased as a function of the duration of the RI (both facts consistent with a familiarity-decay hypothesis), context did not have a differential effect on the recognition of the words.

**Cued-recall data.** Memory scores after a 1-day RI were significantly different from memory scores after a 7-day RI [ $F(1,44) = 39.69$ ,  $MSe = 8.86$ ]. The mean proportion recalled for the group with 1-day RI was .61, whereas the mean for the group with the 7-day RI was .27. There was a significant effect of number of presentations [ $F(2,88) = 155.63$ ,  $MSe = 1.67$ ], with means of .17 for words presented once, .47 for words presented twice, and .68 for words presented four times.

The mean proportion recalled for subjects in the same-context condition was .39, whereas the mean for subjects in the different-context condition was .49. The difference between these two means was not quite significant [ $F(1,44) = 3.51$ ,  $MSe = 8.86$ ,  $p = .068$ ]. There were no significant interactions of context with the other two factors. The only significant interaction was between RI and number of presentations [ $F(2,88) = 6.95$ ,  $MSe = 1.67$ ]; the difference between words presented once and words presented twice was bigger in the 1-day RI condition than in the 7-day RI condition.

**Free-recall data.** Retention interval had a significant effect on memory scores [ $F(1,20) = 5.38$ ,  $MSe = 7.24$ ]. The mean proportion recalled for the 1-day RI group was .15, whereas the mean score for the 7-day RI group was .11. Number of presentations also had a significant effect [ $F(2,40) = 94.36$ ,  $MSe = 5.67$ ]; the mean proportions recalled were .02 for words presented once, .08 for words presented twice, and .27 for words presented four times.

Environmental context did not have a significant effect ( $F < 1$ ). The mean proportion recalled in the same-context condition was .13, and the mean in the different-

context condition was .12. None of the interactions among the three factors was significant. Similar results were produced by an analysis of the probability that a pair was accessed during recall. For this analysis, recall of a single word from a pair and recall of both words from a pair was scored as a one. The sum of these scores was then divided by the number of pairs presented. The mean probability that a pair was accessed in the same-context condition was .15, and the mean for the different-context condition was .13. The difference was not significant.

The lack of a significant context effect in free recall is certainly an unexpected result since it is in opposition to the results of various previous studies (see Godden & Baddeley, 1975; Smith, 1979b; Smith et al., 1978). One implication of this failure to find the basic environmental-context effect is that the interpretation of the recognition results, and more specifically, the adequacy of the familiarity-decay hypothesis in relation to context, will have to be reevaluated. A more immediate concern is that the results demand a careful evaluation of experimental procedures and task requirements to determine which characteristics of the present experiment are critically different from previous studies of the context effect in recall. Experiments 2-7 were directed toward this concern.

## EXPERIMENT 2

The goal of this experiment was to test a possible cause for the failure to observe a context effect (in free recall) in the preceding experiment. In spite of room changes, the presence of the same experimenter in both contexts could have constituted an element of continuity between the two environments that obviated the effect. In that case, a change of experimenter, in addition to a change in location, should make the context effect manifest.

### Method

**Subjects.** Twenty-four students enrolled in introductory psychology classes at the University of Wisconsin—Madison participated in the experiment. The students were given course credit for their participation.

**Materials and environmental contexts.** A total of 33 pairs of words, randomly selected from the pairs used in Experiment 1, were used in the experiment. A learning list started with 3 practice pairs and 3 primacy pairs. Next were 24 critical pairs: 8 of these were presented once, 8 were presented twice, and 8 were presented four times. The list ended with 3 recency pairs. The practice, primacy, and recency pairs were the same for each subject and were excluded from the statistical analyses.

As in Experiment 1, three different learning lists were constructed, with each of the critical pairs appearing once in one list, twice in another, and four times in another. Each list was presented to a third of the subjects, and each pair was printed on a punched card as described in Experiment 1.

Two different rooms were used in the experiment. Room A was on the fifth floor of the Brogden Psychology Building. It was a large room that contained electronic equipment and several chairs, cabinets, and tables. The walls were painted black in two-thirds of the room and gray in the remaining part. The subjects sat at a gray table facing a cassette recorder, a round wall clock, and a work bench on top of which diverse tools and parts of electronic equip-

**Table 1**  
Experiment 2: Mean Proportions Recalled as a Function of Environmental Context and Number of Presentations

Context	Presentations			Probability of Pair Access
	1	2	4	
Same	.08	.30	.64	.35
Different	.07	.28	.61	.33

ment could be seen. The experimenter sat at a different table in front of the subjects. Room B was the same as in Experiment 1.

Two experimenters conducted the experiment. One was a female native speaker of English who was dressed in casual clothes. The other was a male nonnative (but fluent) speaker of English who wore a white laboratory coat. The female experimenter was assigned to Room B and the male experimenter was assigned to Room A.

**Design.** As in Experiment 1, an equal number of subjects were assigned to each of four groups, AA, BB, AB, or BA, indicating study and test rooms. For the analyses, subjects in Groups AA and BB were assigned to a single condition (same-context), and subjects in Groups AB and BA were assigned to a single condition (different-context).

A 2×3 factorial design was used with environmental context (same vs. different) as a between-subjects factor and number of presentations (one, two, or four presentations) as a within-subjects factor.

**Procedure.** Each subject participated, either individually or paired, in two experimental sessions. In the first session, the subjects signed a consent form and were then given the same instructions as in Experiment 1. After completing the sentence task, all subjects were told to come back 24 h later. They were unaware of the recall requirement until the beginning of the second session.

The second session consisted of a free-recall test. Subjects were given 5 min to complete recall of pairs and individual words they could remember from the previous day's session.

## Results and Discussion

An analysis of variance revealed an expected significant effect of number of presentations [ $F(2,44) = 114.99$ ,  $MSe = 4.1$ ]. The mean recall scores are listed in Table 1. Neither environmental context nor its interaction with number of presentations were significant ( $F < 1$  in both cases). The results replicate the absence of a context effect on free-recall memory. Our effort to make the context more distinct by changing experimenters along with rooms had no apparent effect on recall.

## EXPERIMENT 3

In the first two experiments, subjects were confronted with a task (creating a sentence with two previously unrelated words) that is likely to involve complex processes. First, the task requires the processing of more than one item at a time. Second, the task requires transformations in order to "compute" an adequate answer (i.e., a meaningful sentence). And third, the task requires the execution of an overt response. Multiple input, transformations, and overt responding have been shown to require processing demands detrimental to secondary task performance (Foss, 1969; Kerr, 1973; Posner & Keele, 1969; Posner & Klein, 1973). It may be that the demand on processing resources by the orienting task results in a decrement in resources that can be allocated to process-

ing environmental information, causing this kind of processing to be limited or completely absent.

Experiment 3 was designed to test the hypothesis that the level of difficulty of the incidental learning task determines whether or not the environmental context affects memory performance in a delayed recall test. The task was the same as in the first two experiments. Task difficulty was manipulated by providing subjects with associated word pairs or with randomly assigned word pairs. It was expected that when the task was difficult (creating a sentence using two randomly paired words), less attention would be paid to the surrounding context, resulting in the unavailability of contextual information at the time of the recall test. On the other hand, when the task was easier (creating a sentence using two associated words), a contextual-dependence effect should be observed, because more processing resources could have been allocated to the encoding of the environment.

### Method

**Subjects.** Sixty-four introductory psychology students at the University of Wisconsin—Madison served as subjects. They were given course credit for their participation.

**Materials and environmental contexts.** Two lists of easily relatable pairs (A and B) and two lists of more difficult-to-relate pairs (C and D) were used. Lists A and B contained pairs of words in which the second term was the noun that most frequently followed the first term (another noun) in a restricted association test (Riegel, 1965). Lists C and D contained the same words as Lists A and B, but the pairings were randomly determined. The increased difficulty of forming a meaningful sentence with pairs from Lists C and D, compared with the same task when the pairs came from Lists A and B, was independently confirmed in a preliminary stimulus-selection study (Fernandez, 1983). There were 34 pairs in each list. The first two were practice pairs and were not included in the analyses. Also excluded from the analyses were pairs 3 to 6 and 31 to 34, in order to eliminate serial-position effects.

Pairs were printed in uppercase letters at the top of punched cards, and the subjects wrote the sentences on the blank pages of a booklet. The experimental rooms, A and B, were the same as those in Experiment 2.

**Design.** Sixteen subjects were assigned to each of the study-test contexts AA, BB, BA, and AB. Half of the subjects in each group studied a list of easily relatable pairs (either List A or List B) and the other half studied a list of randomly paired words (either List C or List D). For the analyses, Group AA and Group BB were combined into one group (same-context), and Group AB and Group BA were combined into another group (different-context). A 2 × 2 factorial design was used, with environmental context (same vs. different) and difficulty of task (easy vs. difficult) as between-subjects factors.

**Procedure.** Each subject participated, either individually or paired, in two sessions. The first session lasted a maximum of 60 min. The second session was held the following day and lasted 15 min.

In the first session, the subjects signed a consent form and were informed of the time and waiting room for the session to be held the next day. Next, the subjects listened to the instructions that were read by the experimenter in Room B and played on a tape recorder in Room A. The instructions were similar to those in Experiment 1. Upon finishing the task, the subjects were given a reminder card for the next day's session and were told that they would do more work with pairs when they came back.

In the second session, held a day later, the subjects were administered a free-recall test. The subjects were instructed to write as many words as they could remember from those that they had

**Table 2**  
**Experiment 3: Mean Proportion Correct Free Recall as a Function of Environmental Context and Difficulty of the Task**

Context	Tasks	
	Easy	Difficult
Same	.24 (.24)	.17 (.22)
Different	.24 (.26)	.20 (.26)

*Note*—Probability of pair access is given in parentheses.

seen printed on the cards the day before. They were given 5 min to recall the words.

### Results and Discussion

Free-recall means for the different experimental conditions are shown in Table 2. An analysis of variance revealed that the only significant effect was that of difficulty of task [ $F(1,60) = 7.5$ ,  $MSe = 22.95$ ]; the overall means were .24 for the easy-task conditions and .18 for the difficult-task conditions. The effect of environmental context was not significant [ $F(1,60) = 1.5$ ,  $MSe = 22.95$ ]. The interaction between environmental context and difficulty of task was also not significant [ $F(1,60) = 1.5$ ,  $MSe = 22.95$ ]. The difference between the means for same and different context in the difficult task condition, although not significant, was not in the expected direction. These results fail to support the hypothesis that contextual dependence is a function of the difficulty of the orienting task.

## EXPERIMENT 4

This experiment tested the hypothesis that the absence of the environmental-context effect in the preceding experiments was due to the use of an orienting task that involved generation. According to one account of the generation effect (Greenwald, 1981), subjects associate generated responses with some easily accessible component of the self system, a component that can be used later as a retrieval cue. Perhaps subjects disregard contextual cues in favor of self-system cues at the time of the memory test. However, if self-system cues are not available, subjects might be more likely to use contextual cues.

This line of reasoning implies that, although contextual cues are always available, their use depends on the adoption of a particular retrieval strategy. This implication is congruent with the results of an experiment by Smith (1979b). In his experiment, the subjects could make use of contextual information to aid retrieval even when they were in a different room, provided they were prompted to use those environmental cues.

In the present experiment it was expected that when the subjects generated sentences to relate two randomly paired words they would use self-system retrieval cues and not be dependent on context. On the other hand, the subjects reading a sentence in which the TBR words were embedded would be dependent on contextual-retrieval aids, and they would perform more poorly than the subjects in the generation group when the test was in a different context.

### Method

**Subjects.** The subjects were 64 students in an introductory psychology course at the University of Wisconsin—Madison. They participated for course credit.

**Materials and environmental contexts.** Lists C and D from Experiment 3 were used. Each pair was printed on a punched card, and the subjects were given a booklet in which to write their sentences. In the condition in which the sentences were provided by the experimenter, a booklet containing sentences was used. The environmental contexts used in this experiment were rooms A and B, and they were used in exactly the same way as in Experiment 3.

**Design.** Half of the subjects were randomly assigned to the same context condition (learning and test in the same room, AA or BB). The other half were assigned to the different context condition (learning and test in different rooms, AB or BA). Within each of these two groups, half of the subjects generated sentences to relate pairs of TBR words. The other half read experimenter-supplied sentences in which the TBR words were embedded and underlined. A 2 × 2 factorial design was used, with environmental context (same vs. different) and orienting task (generation vs. reading) as between-subjects factors.

**Procedure.** The subjects participated either individually or in pairs in two experimental sessions. The first session lasted a maximum of 60 min. The second session was held 1 day later and it lasted approximately 30 min. Time of day and waiting room were the same for both sessions.

The instructions were read by the experimenter in Room B and were played on a tape recorder in Room A; the instructions were designed to induce incidental learning, and varied depending on the orienting-task condition to which the subjects were assigned. Participants in the generation condition were told that they were in a creativity experiment and received the same instructions as were given in Experiments 1-3, but with one additional requirement. They were asked to estimate how many seconds it took to generate a sentence and to write it in the booklet. The time-estimation task was designed to hide the real goal of the experiment from subjects in the reading condition, and was included in the generation condition to keep the tasks as similar as possible except in the generation versus reading manipulation.

The subjects in the reading condition were told that they were participating in a time-estimation experiment. They were presented with the deck of cards and a booklet in which to write sentences. In addition, they were presented with another booklet that contained sentences written by a participant in the generation group. They were told about the rules that the original writer of the sentences had had to follow. The subjects were instructed to look at a card and see if the words were included in the corresponding supplied sentence. Then they were to copy the sentence and estimate how long it had taken the original writer to think of the sentence and to write it. They were told to use their knowledge of the four criteria used in generating the sentence, and their actual writing of the sentence, to estimate the time. Having the subjects look at the cards, write the sentence, and be aware of the four criteria was expected to enhance the similarity between procedural aspects of the reading task and procedural aspects of the generation task. All subjects were given two practice trials, and they were encouraged to proceed at their own pace until the last pair was reached.

In the second session, all subjects received the same instructions. Immediately after they entered the assigned room they were administered a free-recall test that was identical to the one described in Experiment 3. After 5 min of free recall, the subjects were administered a cued-recall test. In this test, the subjects were given a booklet that contained on each page the first word of one of the 24 experimental pairs (practice, primacy, and recency pairs were excluded). Only one word was presented on each page of the booklet, and the subjects were told to write the second word of the pair. They were encouraged to write a word, even if they had to guess, and to proceed at their own pace until the end of the booklet, working at only one pair at a time and without going back to change a previ-

ously given response. The order of the cues was the same for all subjects, but because the order of presentation of the original pairs in the first session was random, the order of the test cues did not correspond to the learning order.

### Results and Discussion

Group means for the cued-recall test are shown in Table 3. The results of an ANOVA comparing the cued-recall scores showed that the only significant effect was the type of orienting task [ $F(1,60) = 18.83$ ,  $MSe = 24.55$ ]; the overall mean score of the generation group (.42) was greater than the mean score of the reading group (.20). The effect of environmental context was not significant ( $F < 1$ ), and neither was the effect of the interaction between environmental context and orienting task ( $F < 1$ ).

The significant difference in favor of the generation group replicates the finding that generation leads to better memory. Importantly, the difference verifies that subjects in the two groups processed the words differently. Thus the hypothesis regarding the use of self-system cues and contextual cues can be tested with the free-recall data. (The absence of significant effects of context in cued recall is consistent with the results of Smith et al., 1978, in which the effect of the context was small when the testing consisted of cued recall.)

The group means for the free-recall scores are shown in Table 4. An ANOVA comparing the scores for the four conditions produced no significant effects. The absence of a significant effect for type of orienting task [ $F(1,60) = 1.49$ ,  $MSe = 15.16$ ] contrasts with the significant effect observed in the cued-recall scores. It suggests that if the self system is responsible for the generation effect, it is only effective when cues associated with it are aided by other external cues, for example, the other member of the pair.

The results fail to support the central hypothesis of this experiment, namely that a context effect should be observed only in the reading condition. The interaction between environmental context and orienting task was not

**Table 3**  
Experiment 4: Mean Proportion Correct Cued Recall as a Function of Environmental Context and Orienting Task

Context	Tasks	
	Generation	Reading
Same	.43	.16
Different	.41	.23

**Table 4**  
Experiment 4: Mean Proportion Correct Free Recall as a Function of Environmental Context and Orienting Tasks

Context	Tasks	
	Generation	Reading
Same	.12 (.16)	.09 (.13)
Different	.13 (.19)	.11 (.16)

Note—Probability of pair access is given in parentheses.

significant ( $F < 1$ ). In fact, as in Experiment 3, free recall in the different context was slightly, but not significantly, greater than free recall in the same context.

These results have three implications. First, memory for the TBR words is influenced by the generation task, perhaps because the task provides distinctive codes that may be related to the self system. Second, the self-system retrieval cues fail to be activated independently of external cues. (That is, in our experiment, recall in the generation condition exceeded recall in the reading condition only on the cued-recall test, during which external cues—the other member of the pair—were provided. The difference was not found in free recall. This finding is counter to Greenwald's 1981 hypothesis that cues associated to generation processes are more salient than other retrieval cues, including contextual cues.) Third, the results of this experiment, together with the results of Experiment 3, suggest that it is unlikely that characteristics of the orienting task, such as level of difficulty or generation of sentences, are factors responsible for the elimination of environmental-context effects.

## EXPERIMENT 5

In the preceding experiments pairs from the beginning and from the end of the list were excluded from the analyses, and memory scores only for words in the middle of the list were included in the different comparisons among conditions. Might environmental-context effects be restricted to primacy- or recency-serial positions? In one of the experiments reported by Smith et al. (1978), the highest difference in recall between same and different context was found when the comparison was made for words presented at the end of the learning session. Also, studies by Nixon and Kanak (1981) and by Dolinsky and Zabrocky (1983) suggest a relationship between context- and serial-position effects. We included primacy and recency pairs in the analysis of Experiment 5, in order to directly test the hypothesis that the environmental context selectively affects memory for TBR items presented in different serial positions in the input sequence.

Another change from the preceding experiments was the reduction of the retention interval to 5 min. Pilot work convinced us that length of the retention interval (over the range of 5 min to several days) does not interact with environmental context (same or different). This is borne out by a comparison of the pattern of results in Experiments 5-8, which used short retention intervals, with the results of Experiments 1-4, which used long retention intervals.

### Method

**Subjects.** The 32 subjects were introductory psychology students at the University of Wisconsin—Madison. They received class credit for their participation.

**Materials, environmental contexts, and design.** Lists C and D from Experiment 3 were used. Each list consisted of 68 randomly paired words. The order of presentation of all the pairs was randomly determined for each subject. Booklets were provided in which

the subjects were to write sentences. The experimental locations were Room A and Room B, as in the preceding experiment. Subjects were assigned to subgroups AA, BB, AB, and BA, as before. A  $2 \times 3$  factorial design was used, with environmental context (same or different) as a between-subjects factor and serial position (primacy, middle, and recency) as a within-subjects factor.

**Procedure.** Subjects participated, either singly or in pairs, for about 75 min. The subjects signed a consent form, then heard the instructions (read by the experimenter in Room B or played on a tape recorder in Room A). The directions were the same as for the generation condition in Experiment 4, but for the fact that time estimation was not required. The subjects were encouraged to work at generating sentences at their own pace. They were not told about the forthcoming memory test.

Once the subjects finished writing sentences, they were told to go to the waiting room while the next part of the experiment was prepared. After 5 min had passed, the subjects were given the memory test—half in the same room in which they had generated sentences, and half in a different room. Instructions for free recall were the same as in preceding experiments. Five minutes were allowed for recall.

## Results and Discussion

Each subject contributed three scores: the primacy score (proportion of words recalled from the first 6 pairs), the middle score (proportion of words recalled from the next 22 pairs), and the recency score (proportion of words recalled from the last 6 pairs). The two practice pairs were included in the analyses as part of the primacy score.

Group means are shown in Table 5. An analysis of variance comparing the free-recall scores of the six experimental conditions showed a significant effect of the environmental context [ $F(1,30) = 6.32$ ,  $MSe = .033$ ]. The mean unweighted proportions recalled were .42 in the same-context condition and .32 in the different-context condition. The effect of serial position was also significant [ $F(2,60) = 7.98$ ,  $MSe = .021$ ]. The overall mean proportions recalled were .37 for the primacy part of the list, .30 for the middle part, and .45 for the recency part.

The interaction between environmental context and serial position was not significant [ $F(2,60) = 1.77$ ,  $MSe = .021$ ]. However, a set of planned comparisons among means revealed a significant difference between recall in the primacy part of the list for same context and different context [ $t(90) = 2.67$ ]; there was not a significant difference between the two levels of context in the middle of the list [ $t(90) = .303$ ]; the difference was again significant at the end of the list [ $t(90) = 2.02$ ]. These results support the hypothesis of the selective sensitivity of different parts of the list to the environmental-context effect.

Several interpretations of the results are possible. The primacy effect observed in the same-context condition

Table 5  
Experiment 5: Proportion Free Recall as a Function of  
Environmental Context and Serial Position

Context	Position			Weighted Mean
	Primacy	Middle	Recency	
Same	.45 (.48)	.31 (.34)	.51 (.54)	.37 (.40)
Different	.30 (.33)	.29 (.32)	.39 (.45)	.31 (.35)

Note—Probability of pair access is given in parentheses.

could be explained by assuming that subjects pay more attention to the context at the beginning of the experiment. This possibility is consistent with Metcalfe and Murdock (1981) and is explored in the next experiment. There is also a potential explanation for the presence of a stronger recency effect in the same-context condition. If we assume that the representation of the environmental context changes over the course of input, the recency effect may be due to the similarity between the context as it is perceived toward the end of the list presentation and the context as it is perceived 5 min later at the time of the test. This explanation would be consistent with the ideas advanced by Glenberg et al. (1980) in accounting for temporal changes in the long-term recency effect.

### EXPERIMENT 6

The primary purpose of this experiment was to attempt to replicate the significant effects observed in Experiment 5 of environmental context in the primacy and recency portions of the list. Given replication, the second purpose was to determine if the context effect is tied to the beginning and the end of the input session or to the beginning and the end of a list of pairs. Toward this end, subjects were presented with two lists separated by a 2-min arithmetic task. Recall followed presentation of the second list. If the context effect is tied to the beginning and the end of the input session, then we should observe a context effect in the primacy part of the first list and the recency part of the second list. If the context effect is tied to the beginning and end of a list, then we should observe context effects in the primacy and recency parts of both lists.

#### Method

**Subjects.** A total of 32 subjects participated in the experiment. All were introductory psychology students at the University of Wisconsin—Madison; they received course credit for their participation.

**Materials, environmental contexts, and design.** Randomly paired nouns used in the previous experiment were used in this experiment. Two lists (List 1 and List 2) of 17 pairs each were prepared and printed on punched cards, one pair per card. The order of presentation of each pair and its membership in either list was randomly determined for each individual subject. A problem sheet with 50 simple arithmetic problems was constructed to be used between the presentation of List 1 and List 2. Each problem consisted of two or more digits separated by arithmetic signs (addition, subtraction, multiplication, and division), and a blank space where the subject could write the solution to the problem.

The two experimental environments were Room A and Room B, as in the preceding experiment. Eight subjects were assigned to each of the four conditions (AA, BB, AB, BA). A  $2 \times 2 \times 3$  factorial design was used, with context (same vs. different) as a between-subjects factor, and list (List 1 vs. List 2) and serial position (primacy, middle, and recency) as within-subjects factors.

**Procedure.** The subjects participated either singly or in pairs. They were presented with the two lists of TBR words and the arithmetic problems. Subjects were not informed of the real nature of the experiment; instead, they were instructed to generate and write down meaningful sentences using the two words in each pair. They

Table 6  
Experiment 6: Proportion Free Recall as a Function of Environmental Context, List, and Serial Position

Context	List 1 Position			List 2 Position			Weighted Mean
	P	M	R	P	M	R	
Same	.39	.19	.25	.21	.33	.25	.26 (.31)
Different	.33	.25	.29	.33	.36	.39	.32 (.37)

Note—P = Primacy; M = Middle; R = Recency. Probability of pair access is given in parentheses.

were told to proceed at their own pace, and to work with only one pair at a time, without altering the order of presentation. Once the subjects understood the instructions, they were given the cards containing the 17 pairs in List 1. After the subjects finished writing these sentences they were given instructions on how to solve the arithmetic problems, and they were given 2 min to solve as many problems as they could. Then they were given a short summary of the sentence-writing task and they were presented with the cards containing List 2. Upon completion of this task, the subjects were instructed to go back to the waiting room. Five minutes later the subjects were escorted to either the same- or different-context room and were given 5 min to write all the words they could remember from all the pairs presented to them in the first session of the experiment. They were not instructed to recall words from List 1 and List 2 separately.

#### Results and Discussion

Six scores were obtained from each subject in the experiment, three from List 1 and three from List 2. The scores from each list were the primacy score (proportion of words recalled from the first 3 pairs), the middle score (proportion of words recalled from the next 11 pairs), and the recency score (proportion of words recalled from the last 3 pairs). Group means for the different conditions are presented in Table 6.

An analysis of variance revealed no significant effects of list ( $F < 1$ ) or serial position ( $F < 1$ ). Regarding context, the unweighted mean proportion recalled was .27 in the same-context condition and .33 in the different-context condition; however, this difference was not significant [ $F(1,30) = 1.82$ ,  $MSe = .092$ ]. The interaction between list and serial position was significant [ $F(2,60) = 3.88$ ,  $MSe = .047$ ]. The mean proportions for primacy, middle, and recency positions were, respectively, .353, .218, and .270 in List 1 and .270, .347, and .317 in List 2.

Clearly, we did not replicate the results of Experiment 5. The interaction of environmental context and serial position is as elusive as the main effect of environmental context.

### EXPERIMENT 7

One element is present in all the experiments reported so far: The orienting task has involved the processing of sentences. Experiment 7 was designed to determine if the effect of the environmental context would be observed if the orienting task did not require processing of sentences. Except for the change of the orienting task, this experiment is identical to Experiment 5.

Table 7  
Experiment 7: Proportion Free Recall as a Function of  
Environmental Context and Serial Position

Context	Position			Weighted Mean
	Primacy	Middle	Recency	
Same	.24	.19	.26	.21 (.32)
Different	.26	.21	.26	.22 (.34)

Note—Probability of pair access is given in parentheses.

## Method

**Subjects.** The 48 subjects were introductory psychology students at the University of Wisconsin—Madison. They received course credit for their participation in the experiment.

**Materials, environmental contexts, and design.** Lists C and D (from Experiment 3) were used. Each list consisted of 34 randomly ordered pairs. The environmental contexts were Room A and Room B, as in the preceding experiment, and the subjects were randomly assigned to conditions AA, BB, AB, and BA. A  $2 \times 3$  factorial design was used, with environmental context (same vs. different) as a between-subjects factor, and serial position (primacy, middle, and recency) as a within-subjects factor.

**Procedure.** The experimental procedure parallels that of Experiment 5, with the exception of the instructions given to the subjects. In the present experiment, the subjects were told that they were participating in a size-judgment task, and they would have to compare the size of objects referred to by each of the two members of a pair. For each pair, the subjects were to write the noun that represented the smaller object, and then the other noun in the pair. The subjects were provided with the deck of cards containing the 34 pairs and were told to proceed, one pair at a time, at their own pace. They wrote each pair on a separate blank page.

After finishing the comparison and writing task, the subjects waited for 5 min in the waiting room. Then they were given a free-recall test—subjects in the same-context condition were tested in the room in which they had seen the pairs, and subjects in the different-context condition were tested in a different room.

## Results and Discussion

Each subject contributed three scores. The primacy score was the proportion of words recalled from the first 6 pairs in the list (including 2 practice pairs); the middle score was the proportion of words recalled from the next 22 pairs; and the recency score was the proportion of words recalled from the last 6 pairs in the list.

Group means are shown in Table 7. An analysis of variance revealed that environmental context did not have a significant effect ( $F < 1$ ), nor did serial position [ $F(2,92) = 2.7$ ,  $MSe = .017$ ]. Finally, the interaction between context and position was not significant ( $F < 1$ ).

These results indicate that a sentence-orienting task has little influence on the occurrence of the environmental-context effect. Furthermore, this experiment did not replicate the pattern of results obtained in Experiment 5 that suggested a relationship between environmental-context and serial-position effects.

## EXPERIMENT 8

Across seven experiments, we were unable to isolate the condition(s) consistently associated with a same-context advantage in recall. In fact, a different-context advantage was observed about as frequently as a same-

context advantage. Rather than continue the search for variables that might control the use of environmental context in remembering, we decided to investigate the replicability of the effect itself.

The procedures selected for replication were those used in Smith's (1979b) Experiment 1. This experiment was chosen because it was somewhat similar in methodology to our Experiments 1-7; thus, if a significant same-context advantage were found, it would be possible to titrate between the successful and unsuccessful procedures.

Subjects in Smith's (1979b) Experiment 1 were assigned to one of three main groups: ACA, ACB, and ACC. The letters refer to the environmental contexts in which subjects experienced the three phases of the experiment (list learning, context familiarization, and recall). Half of the subjects in each of Smith's groups received visual presentation of the TBR words and half received aural presentation.

The procedures used in Experiment 8 differed from Smith's (1979b) Experiment 1 in three respects. First, the ACC group was not included. Second, the TBR material was presented aurally to all subjects. (In the report of the experiment in Smith, 1979a, auditory presentation produced a larger same-context advantage than did visual presentation.) Third, Smith included 10 subjects in each of his main groups, whereas we included 24. In all other respects, our experimental procedures were very similar to Smith's.

## Method

**Subjects.** Forty-eight introductory psychology students at the University of Wisconsin—Madison served in the experiment. They received course credit for their participation.

**Materials.** The 90 words used in the original experiment (Smith, 1979b, Experiment 1) were used in the present one. The words were all four- and five-letter, unrelated, high-frequency English nouns. Eighty words were presented in the learning list, and 10 other words were presented as distractors on a recognition test. The words were presented aurally, using a tape recorder.

**Environmental context.** Three rooms were used in the experiment. Room A was a soundproof booth placed inside a room in which there was various computer equipment; the room was on the fourth floor of the psychology building. Room B was the same as in Experiments 1-7. Room C was on the second floor of the psychology building and it contained a few desk chairs, a slide projector, a cabinet, and a projection screen. Room B was the actual room used by Smith, and Rooms A and C were very similar to those he used.

**Design.** Subjects were randomly assigned to one of four groups. Group ACA studied the list of words, and was later tested for free recall, in Room A. Group BCB studied the list in Room B and was later tested in the same room. These two groups were combined to form the same-context condition. Group ACB studied in Room A and was tested in Room B. Group BCA studied in Room B and was tested in Room A. These last two groups were combined to form the different-context condition. Room C was always used as a neutral context between the learning and text contexts.

As in Experiments 5-7, serial-position effects were examined. A  $2 \times 3$  factorial design was used, with environmental context (same vs. different) as a between-subjects factor and serial position (primacy, middle, or recency) as a within-subjects factor.

**Procedure.** The subjects participated, singly or in pairs, in a 30-min experimental session, and were randomly assigned to one of

the two experimenters conducting the study. The first phase of the session, the learning phase, lasted 10 min. During that period the subjects were presented with a list of 80 words and were told to memorize the words as well as they could. The words were presented at the rate of 1 every 3 sec. The subjects were then given a recognition test consisting of 10 words from the TBR list and 10 new words. The words in the recognition test were presented at the rate of 1 every 6 sec; the subjects had to indicate whether a word was "old" or "new" and to indicate their confidence in that judgment by circling a number on a rating scale. This recognition test was used to suggest to the subjects that the memory part of the experiment was over. The recognition test and the single order in which the words were presented were identical to those used in Smith (1979a).

The subjects were taken to Room C and instructed to draw two different views of the room. Five minutes were allowed for each drawing. Then the subjects were taken either to the room where they learned the list (same-context condition) or to a new room (different-context condition) for a free-recall test. They were given 10 min to recall all the words presented in the experiment, including distractors used in the recognition test.

### Results and Discussion

Words presented in the recognition test (10 distractors and 10 words from the TBR list) were eliminated from the main data analyses. Three scores were contributed by each subject: the primacy score (the proportion of words recalled from the first 10 words in the study list), the middle score (the proportion of words recalled from the next 50 words in the list), and the recency score (the proportion of words recalled from the last 10 words in the list).

The means of the six conditions are listed in Table 8. In general, there was better retention in the same-context condition than in the different-context condition (mean proportions .22 and .18, respectively), but an ANOVA revealed that this difference was not significant [ $F(1,46) = 3.31$ ,  $MSe = .019$ ]. (The weighted means, .19 and .17 for same and different context, respectively, are clearly not significantly different.) There was a significant effect of serial position [ $F(2,92) = 27.94$ ,  $MSe = .376$ ], with primacy scores higher than both middle and recency scores. The interaction between environmental context and serial position was not significant ( $F < 1$ ).

In order to gain statistical power, we conducted parallel analyses of covariance for three dependent measures: recall of TBR words not presented on the recognition test, recall of TBR words presented on the recognition test, and recall of distractors from the recognition test. Each analysis included one experimental factor (same or different context) and three covariates: score on the recognition test, experimenter (one of two), and place of the free-recall test (Room A or B). The only significant effect of environmental context was for recall of distractors

[ $F(1,43) = 4.63$ ,  $MSe = .019$ ]. However, this was a different-context advantage; mean proportion recalled of distractors in the different-context condition was .25 and mean proportion recalled of distractors in the same-context condition was .16.

These results fail to replicate the findings of Smith (1979b, Experiment 1). Because this experiment used an intentional learning procedure, whereas Experiments 1-7 used incidental learning, differences in intentionality of learning can be eliminated as the sole factor precluding observation of a reliable same-context advantage.

### GENERAL DISCUSSION

The combined results of these eight experiments demonstrate that manipulations of environmental context do not produce reliable results. A possible reason for our failure to produce reliable results is that our experiments lacked statistical power, but this seems unlikely. For example, it might have been the case that the same-context condition generally facilitated recall, but the facilitation was not large enough to reach significance. However, this was not the case. Considering the results from the various free-recall conditions in our eight experiments, there was a same-context advantage of more than 2% in 6 comparisons, a different-context advantage of more than 2% in 6 comparisons, and less than a 2% difference in 11 comparisons. A more formal power analysis indicates sufficient power in most of the experiments.<sup>1</sup> Experiment 8 duplicated Smith's (1979b) methodology. He reported a same-context advantage of about 1 standard deviation. Using this as an estimate of the effect size, the power of Experiment 8 to detect a similar effect (assuming a two-tailed, .05 test) was .94. Experiment 5 produced a significant same-context advantage in the primacy and recency portions of the serial position curve. The average effect size in these positions was .92 of a standard deviation. Experiments 6 and 7 were methodologically similar to Experiment 5; using .92 as an estimate of the effect size, the power of Experiment 6 was .74 and the power of Experiment 7 was .89. Assuming an effect size of 1 standard deviation, the power of Experiments 1-4 varied from .71 to .98. Although consistent with Smith's results and with Experiment 5, an estimated effect size of 1 standard deviation may be much too large. Assuming an effect size of .5 standard deviation yields power estimates of .22 to .52 for Experiments 1-4. These estimates are uncomfortably low. However, even with these low estimates, the probability of finding at least one same-context advantage across the four independent experiments is .86. Thus it seems unlikely that our failure to find a reliable same-context advantage was due to low power.

Another possibility is that the use of pairs as TBR items (in Experiments 1-7) may have reduced the same-context advantage. R. A. Bjork (personal communication, December 12, 1984) has suggested that recall of words may have required (1) accessing the representation of one member of a pair (which should have been sensitive to context manipulations), and then (2) retrieving the other mem-

Table 8  
Experiment 8: Mean Free Recall as a Function of  
Environmental Context and Serial Position

Context	Position			Weighted Mean
	Primacy	Middle	Recency	
Same	.337	.169	.166	.192
Different	.270	.161	.110	.169

ber of the pair using mechanisms which may not have been sensitive to context manipulations. Thus, presenting pairs as TBR items, but scoring recall of individual words, may have diluted the effect. Interestingly, a similar analysis led us to retain the use of pairs in the experiments. We reasoned that with single words as stimuli, the subjects would be likely to form just a few subjective organizations with each organization encompassing many TBR words. Although accessing each organization might be sensitive to context manipulations, the organizational structure, once accessed, can cue recall, thus obviating the use of contextual cues. On the other hand, forcing subjects to focus on individual pairs should have resulted in the formation of many memory representations (perhaps one for each TBR pair), each of which should be sensitive to context manipulations.

Bjork's suggestion can be tested by scoring pair access, that is, recall of at least one member of a pair, rather than recall of individual words. If context only affects pair access, then these scores should reveal a greater same-context advantage than recall scores of individual words. Data on accessing pairs (that is, scoring for recall of at least one) are presented in Tables 1, 2, 4, 5, 6, and 7. In none of the tables was a same-context advantage enhanced by analysis of pair access. In fact, the trend seems to be that analysis of pair access decreases the same-context advantage (or increases the different-context advantage).

As in many of Smith's (1979b; Smith et al., 1978) experiments, we collected some of our data from pairs of subjects. Perhaps when a pair of subjects in the different-context condition were together for both study and recall, even when recall was in a different room, enough of the context was preserved to dampen the same-context advantage (R. A. Bjork, personal communication, December 12, 1984). We were able to test this supposition with the data from Experiments 3 and 4. In those experiments, a sizable proportion of subjects in the different-context conditions were initially in pairs, but were tested separately (this was never the case with subjects in the same-context conditions). In Experiment 3, the difference in proportion recalled between subjects in the same-context condition and subjects in the tested-separately different-context condition was  $-.005$ , a slight different-context advantage. In Experiment 4 the difference was  $-.03$ .

Given the evident unreliability of the same-context advantage, why has it become a staple of our theories and a chapter in the learning and memory mythology? One possibility is that previous reports of a same-context advantage have included a mix of disruption effects (Strand, 1970), real instances of the same-context advantage under conditions that are as yet unknown, and statistical Type I errors. Indeed, in our series of experiments there were instances of a numerical same-context advantage (Experiments 2 and 8) and a statistically significant same-context advantage (Experiment 5). After obtaining the data from Experiment 5, our conviction in regard to the reliability of the data was great enough that we reported

them (Fernandez & Glenberg, 1983) as evidence of a successful replication of the same-context advantage. It now seems that our conviction was misplaced. Because the results of Experiment 5 were not replicated in Experiment 6 or in Experiment 7 (in fact, the order of the weighted means was reversed in both of these experiments), the significant result in Experiment 5 is most parsimoniously interpreted as a Type I error.

Can all of the reports of a same-context advantage (that are not due to disruption) be Type I errors? This question is probably impossible to answer, because we do not have data such as the number of failures to obtain a same-context advantage coupled with reasonable methodologies and reasonable statistical power. Nonetheless, the answer, based on other considerations, would seem to be no.

One other consideration is the theoretical necessity, given our current understanding of memory, for contextually marked memories. Memories are not only abstract, but episodic memories, at least, seem to encode information pertaining to the time and place in which the represented event was experienced. It is probably the encoded contextual information that embellishes remembering of episodes with feelings of immediacy and personal experience. At a more experimental level, it is probably the encoded contextual information that endows us with the ability to estimate frequency of events and to engage in various discriminations such as recency of events, list discrimination, etc. In short, our current scientific conceptions of memory demand some representation of context in memories.

Although context may be a required component of episodic memories, it need not have a direct effect on recall. In Underwood's (1969) terminology, context may be a discriminative attribute (providing a basis for discrimination between events) but not a retrieval attribute. However, effects of change in environmental context are minimal on a recognition test (Smith et al., 1978), which should be sensitive to discriminative attributes. Also, Smith (1982) reported that presenting lists of words in separate rooms did not enhance performance on a list-discrimination test relative to presenting all of the lists in a single room.

A second reason for not treating all instances of a same-context advantage as statistical aberrations is the frequency of anecdotal reports of the phenomenon. Many people report having experienced an improvement in memory for events associated with a specific environmental context upon reentering that context. Although some of these reports are no doubt apocryphal, the frequency of the reports and the subjective experience associated with the phenomenon are too powerful to be relegated to fiction. These reports tend to share three characteristics which distinguish them from experiments investigating environmental context effects. First, improvements in memory upon reentering a context are typically associated with long retention intervals. For example, many years may elapse between secondary-school graduation and the class reunion, during which memories come flooding back. Se-

cond, the improvement is associated with a unique context in terms of the physical qualities of the context (e.g., a school building), the time during which the event was experienced, and the effect of the event on the experiencer. Third, the improvement is associated with multiple retrieval attempts over an extended period of time (e.g., during the class reunion), during which retrieval cues provided by the physical context, interactions with others, and successfully retrieved memories are varied, richly interconnected, and in flux.

Given these considerations, the major conclusion that we draw from our experiments is this: Manipulation of the match between study and test rooms does not capture the critical features of natural changes in context that produce the same-context advantage. There seem to be two related problems with using standard-memory techniques to investigate environmental-context effects. First, memory in natural situations represents an interlocking series of coherent events. We interpret and remember specific events as being embedded within a structure of more global (and more local) events interrelated by physical causes and psychological goals of the actors taking part in the events. Also, the environmental context is usually closely related to the events that occur within the environment. That is, the environment may be perceived as causing the event or enabling the event to occur. Thus, it is likely that relationships between the environmental context and events are integral to the representation of naturally occurring events.

In most laboratory studies of memory, however, individual events correspond to words on a list. These words are probably not perceived as causing or enabling one another, and the specific words presented almost certainly are not perceived as being closely related to the environment in which they occur. In other words, laboratory tasks do not preserve the strong links between naturally occurring events and their environments that are presumed to underlie anecdotal reports of context effects.

Second, there has been very little analysis of critical features of the experimental context from the subject's point of view. From the experimenter's point of view, changing the room produces a massive change in stimulation. The critical feature for the subject, however, may be that the events take place within the context of an experiment, and this feature does not change with a change of room.

In summary, our data indicate that changing the environmental context by changing the room in which an experimental procedure takes place produces unreliable effects on recall. We propose that future studies of effects of the environment are likely to be more fruitful if (1) they deal with memory for events that are perceived to be related to the environment, and (2) manipulations of the environmental context involve the components of the context (environmental, experimental, temporal, subject's goals, etc.) that are likely to be perceived by the subject as relevant to the memory task (cf. Baddeley's,

1982, distinction between interactive and independent context).

## REFERENCES

- ANDERSON, J. R., & BOWER, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*, 97-123.
- BADDELEY, A. D. (1982). Domains of recollection. *Psychological Review*, *89*, 708-729.
- BOWER, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press.
- DOLINSKY, R., & ZABRUCKY, K. (1983). Effects of environmental context changes on memory. *Bulletin of the Psychonomic Society*, *21*, 423-426.
- FERNANDEZ, A. (1983). *Modulating factors in the environmental context effect*. Unpublished master's thesis, University of Wisconsin, Madison.
- FERNANDEZ, A., & GLENBERG, A. M. (1983, May). *Cognitive boundaries on the environmental context effect*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago.
- FLEXSER, A. J., & TULVING, E. (1978). Retrieval independence in recognition and recall. *Psychological Review*, *85*, 153-171.
- FOSS, D. J. (1969). Decision processes during sentence comprehension: Effect of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior*, *8*, 457-462.
- GLENBERG, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95-112.
- GLENBERG, A. M., BRADLEY, M. M., KRAUS, T. A., & RENZAGLIA, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 231-255.
- GLENBERG, A. M., BRADLEY, M. M., STEVENSON, J. A., KRAUS, T. A., TKACHUK, M. J., GRETZ, A. L., FISH, J. H., & TURPIN, B. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 355-369.
- GODDEN, D. R., & BADDELEY, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*, 325-331.
- GREENWALD, A. G. (1981). Self and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 15). New York: Academic Press.
- KERR, B. (1973). Processing demands during mental operations. *Memory & Cognition*, *1*, 401-412.
- KINTSCH, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271.
- METCALFE, J., & MURDOCK, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *20*, 161-189.
- NIXON, S., & KANAK, N. J. (1981). The interactive effects of instructional set and environmental context changes on the serial position effect. *Bulletin of the Psychonomic Society*, *18*, 237-240.
- POSNER, M. I., & KEELE, S. W. (1969). Attention demands of movements. *Proceedings of the XVIIth Congress of Applied Psychology*. Amsterdam: Zeitlinger.
- POSNER, M. I., & KLEIN, R. (1973). On the functions of consciousness. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.
- RAAIJMAKERS, J. G., & SHIFFRIN, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93-134.
- RIEGEL, K. F. (1965). *The Michigan restricted association norms* (Report No. 3, USPHS Grant MH07619). Ann Arbor, MI: University of Michigan, Department of Psychology.
- SIEGAL, S., HINSON, R. E., & KRANK, M. D. (1981). Morphine-induced attenuation of morphine tolerance. *Science*, *212*, 1533-1534.

- SMITH, S. M. (1979a). *Context dependence in episodic memory*. Unpublished doctoral dissertation, University of Wisconsin, Madison.
- SMITH, S. M. (1979b). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 460-471.
- SMITH, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, **10**, 405-412.
- SMITH, S. M., GLENBERG, A., & BJORK, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, **6**, 342-353.
- SPEAR, N. E. (1978). *The processing of memories: Forgetting and retention*. Hillsdale, NJ: Erlbaum.
- STRAND, B. Z. (1970). Change of context and retroactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, **9**, 202-206.
- TULVING, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- UNDERWOOD, B. J. (1969). Attributes of memory. *Psychological Review*, **76**, 559-573.

#### NOTE

1. Most of the following analyses were suggested by an anonymous reviewer.

(Manuscript received August 27, 1984;  
revision accepted for publication April 22, 1985.)