Memory for Real-World Scenes: 
The Role of Consistency With Schema Expectation

Kathy Pezdek, Tony Whetstone, Kirk Reynolds, Nusha Askari, and Thomas Dougherty
The Claremont Graduate School

This study tested the generalizability of the consistency effect to real-world settings. The consistency effect refers to the finding that items inconsistent with expectations are better recalled and recognized than items consistent with expectations. In two experiments, subjects walked into a graduate student's office or a preschool classroom. Half of the items in each setting were consistent with expectations about that setting, and half were inconsistent. A recall and a same-changed recognition memory test followed immediately or 1 day later. In both experiments, the consistency effect was affirmed: items inconsistent with expectations were significantly better recalled and recognized than items consistent with expectations. This result is discussed in terms of differences in the encoding processes that operate on inconsistent and consistent items. The present study extends the generalizability of results from picture memory studies to real-world settings.

Two facts about memory for scenes and pictures are well established. First, people are impressively accurate in distinguishing scenes they have seen before from scenes they have not seen before (Nickerson, 1965; Shepard, 1967; Standing, Conezio, & Haber, 1970). Second, the perception and cognition of scenes is largely schema-guided (Biederman, 1972; Loftus & Mackworth, 1978, Pezdek et al., 1988). However, when we look more specifically at people's ability to remember descriptive information about items in a scene, the results are both less impressive and less straightforward (Mandler & Parker, 1976; Pezdek et al., 1988). For example, although people who have viewed a picture of a group of men talking on a city street are very likely to remember having seen a picture of a group of men talking on a city street, and although they are very likely to remember that the picture included men, a parked car, and a doorway (inventory information), they are not likely to be able to recognize changes in the physical appearance of either the men, the parked car, or the doorway (descriptive information).

One factor known to affect how accurately people remember the inventory and descriptive information about items in a scene is the extent to which each item is consistent or inconsistent with what one expects to occur in the schema activated by the scene. Consistency with expectation has been defined by Mandler (1984) as the probability of occurrence of an item in a scene—would a specific item typically be found in a particular scene? Mandler and also Loftus and Mackworth (1978) have differentiated the quality of consistency with expectation from qualities of (a) the schema relevance of items, (b) subschematic details of items, and (c) the importance of items in the scene. The general finding in this literature has been that items inconsistent with expectations are better recalled and recognized than items consistent with expectations. This effect, termed the consistency effect, has been reported for scenes (Friedman, 1979; Goodman, 1980; Hock, Romanski, Galie, & Williams, 1978), for faces (Light, Kayra-Stuart, & Hollander, 1979), as well as for scripts (Bower, Black, & Turner, 1979; Graesser, Gordon, & Sawyer, 1979).

One study that investigated memory for items that are consistent versus inconsistent with expectation reported results opposite to those summarized above, and this is the only study that has investigated the consistency effect in a real-world environment. In this experiment by Brewer and Treyens (1981), subjects were taken one at a time to a graduate student's office in a psychology building on campus. They were instructed to simply wait in the office until the experimenter checked to make sure that the previous subject had completed the experiment in an adjacent room. In the office were 61 items that had been previously rated on dimensions of schema expectancy and saliency. Each subject was left alone in the office for 35 s. When the experimenter returned, the subjects were led to another room and were given tests of (a) verbal recall, (b) drawing recall, and (c) verbal recognition.

Brewer and Treyens' major result was that schema expectancy (as well as saliency) was positively correlated with recall and recognition. That is, the items rated as most consistent with expectations about what should be found in a graduate student's office (e.g., chair, desk, table, shelves) were recalled and verbally recognized more frequently than items rated as inconsistent with expectations (e.g., wrench, umbrella, picnic basket, brick). Brewer and Treyens concluded that "it seems very likely that at least part of the general positive correlation of schema expectancy and recall was due to office-schema information serving as a retrieval mechanism in recall" (p. 228).
The purpose of the present study is to test the reliability of the effect reported by Brewer and Treyens (1981). The discrepancy between the results of Brewer and Treyens and the previous studies on the consistency effect is troubling. This discrepancy, if valid, would raise questions about the extent to which the results of picture memory studies generalize to memory for real-world scenes. Perhaps the notion of consistency operates differently in experimental drawings than in real-world scenes where, quite honestly, anything might appear anywhere.

The present study includes two experiments. The experiments used real-world spaces—a graduate student's office in the psychology building on campus (Experiments 1 and 2) and a preschool classroom in the child care center on campus (Experiment 1). In Experiment 1 the same 16 items were arranged in each of the two rooms. Half of the items had been previously rated as consistent with the office and inconsistent with the preschool, and the other half of the items had been previously rated as consistent with the preschool and inconsistent with the office. Consequently, across subjects, the same 16 items were viewed in both settings, and, across subjects, each item served as both a consistent and an inconsistent item.

Immediately (Experiments 1 and 2) or 1 day after viewing the room (Experiment 1), subjects were administered a recall test followed by a visual recognition test. This study extends Brewer and Treyens' paradigm by including a visual recognition test rather than a verbal recognition test as used by Brewer and Treyens. The distractor items in the visual recognition test were token changed items, which allowed an assessment of, for example, whether subjects could detect when the telephone originally presented has been replaced by a physically different telephone. Thus, the present study assessed memory for descriptive or appearance information, as well as inventory information.

**Experiment 1**

**Method**

**Subjects and Design**

The subjects were 61 undergraduates who volunteered from classes at the Claremont Colleges. Subjects viewed either the graduate student office or the preschool classroom and were tested either immediately or after a 1 day's delay. The room viewed contained 16 items, 8 consistent and 8 inconsistent with expectations. The design was a 2 (setting) × 2 (delay) × 2 (consistency) mixed factorial design, with the first two factors manipulated between subjects and the third factor manipulated within subjects.

**Stimulus Items**

The presentation stimulus items were the same 16 items in each room, with 8 items consistent and 8 items inconsistent with expectations in each setting. These 16 items were selected by having a different group of subjects rate a larger set of potential items. The experimenters generated an initial list of the names of 25 items, each judged to be either (a) consistent with expectations about a graduate student's office and inconsistent with expectation about a preschool classroom or (b) consistent with expectations about a preschool classroom and inconsistent with expectations about a graduate student's office. Two actual items for each of these 25 named items were obtained (i.e., for "ashtray," two actual ashtrays were selected, an octagonal glass ashtray and a round tin ashtray). The duplicate of each of the items was selected for use as a token changed test item.

Twelve undergraduates then rated each of the 50 items in the graduate student's office, and 13 subjects rated each of the same 50 items in the preschool classroom. The consistency of each item was rated on a scale from 1 (high probability) to 7 (low probability). Consistency for each was defined as how likely it was that they would find that item in that specific setting, either the office or the preschool. The mean consistency ratings were calculated for each item. On the basis of these ratings, 16 items were selected for Experiment 1. The items selected were those that were rated as most distinctly consistent in one setting and inconsistent in the other setting. A list of the 16 presentation items and the mean consistency rating of each item in each setting is presented in the Appendix. For the 16 presentation items, when rated in the graduate student office the mean consistency scores for the 8 consistent items (2.33, range = 1.50-3.30) and the 8 inconsistent items (6.32, range = 5.75-6.83) were significantly different, t(12) = 27.40. Similarly, for the same 16 items rated in the preschool classroom the mean consistency scores for the 8 consistent items (1.81, range = 1.17-3.07) and the 8 inconsistent items (5.30, range = 3.69-6.71) were significantly different, t(11) = 13.58. Further, additional analyses were conducted to ensure that for each item, the consistency ratings did not differ for the original versus the corresponding token changed test item. For each of the 16 items in each of the two rooms separately, a t test for related measures was computed. The consistency ratings did not significantly differ for any of these pairs.

**Settings**

Two experimental settings were used. A graduate student's office in the psychology building on campus was one setting. This room was 6.17 × 3.15 m, with overhead lighting and no windows. The room was entered from a corridor on the ground floor in the building. Four "room frame" items were included in the room—a desk, a desk chair, an easy chair, and an empty set of book shelves. These items served to define the setting as an office (cf. Minsky, 1975) and were not included as stimulus items to be recalled or recognized.

The second setting was a room in the preschool section of the campus child care center. The classroom was 4.57 × 3.05 m, with two windows and overhead lighting. The preschool was a one-block walk from the psychology building. Participants walked across a play yard and through another classroom before reaching the experimental classroom. This served to establish the authenticity of the experimental setting. Four room frame items were included in this setting: a set of wooden shelves, a toy kitchen set, a rocking chair, and an alphabet board. These room frame items were not included as stimulus items to be recalled or recognized.

For the presentation phase, in each of the two rooms the same 16 items specified above were placed throughout the room, primarily on room frame objects.

**Procedure**

**Presentation phase.** Subjects participated in small groups of 1 to 3 people. Each group assembled in the hallway of the psychology building with one of the experimenters. They were then directed to either the graduate student office or the preschool classroom. Before
entering the room, subjects were instructed as follows:

In a minute I am going to take you into a graduate student's office/preschool classroom. I want you to study this environment carefully as I am going to ask you some questions afterward about this room and your reaction to it. You will have one minute in the room. You may move about the room if you wish, but please do not talk to anyone during this time. After viewing the room for 1 min they were directed out through the doorway.

Test phase. Subjects were tested either immediately or 1 day later. The test phase consisted of a recall test followed by a recognition test. In the recall test, subjects were asked to write down a list of as many of the items as they could remember from that room.

While subjects completed the recall test, a second experimenter entered the experimental room through another door and substituted token changed items for 4 consistent and 4 inconsistent items. The experimenter also placed cards numbered from 1 to 16 in front of the test items. When subjects completed the recall test, they were led back into the room for the recognition test. In the test room were 16 items, half old and half token changed new items. Four of the old items and 4 of the new items were consistent, and the other half were inconsistent. Subjects were instructed that they would see 16 items with a number in front of each. They were to look at each item and decide if it was the same as an original item or changed. A changed item was defined as one that had the same name as an original item but looked physically different from the corresponding original item. Subjects circled same or changed for each of 16 numbered items on the sheet provided.

Results and Discussion

Recall Test

The recall data were scored in terms of the number of old items correctly recalled. In each condition the maximum recall score was 8. Few intrusion errors occurred, and these were not analyzed. Throughout this study the rejection region for all analysis was $p < .05$.

The mean number of items correctly recalled in each condition is presented in the top portion of Table 1. A 2 (setting) × 2 (delay) × 2 (consistency) analysis of variance was first performed on the correct recall data. As predicted by the consistency effect, $d'$ was greater for inconsistent items ($d' = 3.35$) than for consistent items ($d' = 2.66$), $F(1, 57) = 7.55, M_{SE} = 1.87$. The only other significant effects were the main effect of setting, with $d'$ greater for the office ($d' = 3.35$) than for the preschool ($d' = 2.65$), $F(1, 57) = 4.76, M_{SE} = 3.15$, and the main effect of delay, with $d'$ greater immediately ($d' = 3.56$) than after 1 day ($d' = 2.44$), $F(1, 57) = 12.03, M_{SE} = 3.15$.

Separate 2 (setting) × 2 (delay) × 2 (consistency) analyses of variance were also performed on the hit rate [$p(\text{"same"}/\text{same})$] and false alarm rate [$p(\text{"same"}/\text{changed})$] data. No effects were significant in the analysis of the hit rate data. Most notably, consistency did not significantly affect the hit rate data ($p = .18$), $F(1, 57) = 1.85, M_{SE} = 0.01$. On the other hand, the analysis of the false alarm rate data yielded a significant effect of consistency, $F(1, 57) = 4.37, M_{SE} = 0.50$, with a higher false alarm rate for consistent items (.32) than for inconsistent items (.24). Also, the false alarm rate was higher after 1 day (.37) than immediately (.19), $F(1, 57) = 11.38, M_{SE} = 0.08$.

Experiment 2

The principal result of Experiment 1, that items inconsistent with expectations were better recalled and recognized than items consistent with expectations, is discrepant with the result reported by Brewer and Treyens (1981), that schema expectancy was positively correlated with recall and recognition. There are several methodological differences between the Brewer and Treyens' study and our own. The purpose of Experiment 2 is to test if any of these specific methodological differences can account for the difference in results between the two studies.

One methodological difference between the present study and that of Brewer and Treyens is the ratio of inconsistent to consistent items available in the room studied. In Brewer and Treyens' study there were 61 objects in the graduate student's office, including "a few items not consistent with the office schema" (p. 210). Thus, the ratio of inconsistent items to...
consistent items was very small, significantly less than in other comparable studies. If the consistency effect results from encoding differences between consistent and inconsistent items, perhaps the inconsistent items in the Brewer and Treyens study were simply not noticed or not attended to.

In Experiment 2, subjects viewed the graduate student’s office with the 8 consistent and 8 inconsistent items used in Experiment 1. To test if the difference between the results in this study and those of Brewer and Treyens can be accounted for by the difference in the ratio of inconsistent to consistent items in the two studies, Experiment 2 included a second graduate student’s office. The second office was a normal, very full office with dozens of consistent items in it. Thus, the ratio of inconsistent to consistent items was smaller in the full room than in the simplified room. Half of the subjects viewed the 8 consistent and 8 inconsistent items in this full office, and half viewed the same items in the simplified office from the previous experiments. Only memory for the 8 consistent and 8 inconsistent target items was tested.

A second methodological difference between the present study and that of Brewer and Treyens is the instructions given to subjects prior to the study phase. Whereas we used intentional instructions (“I want you to study this environment carefully as I am going to ask you some questions afterward about this room and your reaction to it.”), Brewer and Treyens used incidental instructions (the experimenter simply told the subjects, “Wait in my office”). If the consistency effect results from encoding differences between consistent and inconsistent items, subjects may be less likely to attend to the inconsistent items in the incidental instructions condition. This is plausible, given that in the Brewer and Treyens study subjects were in the room for only 35 s, as compared with 1 min in the present study.

In a related study Salmaso, Baroni, Job, and Peron (1983) reported that the encoding instructions given to subjects significantly varied their memory for types of objects in a scene. Subjects were walked through an office area and then tested on their memory for the physical appearance of room frame objects (e.g., floor, walls, ceiling light) and schema consistent objects (e.g., chair, waste basket, card index). With incidental instructions, room frame objects were better recognized than schema consistent objects; however, the result reversed with intentional instructions.

In Experiment 2 of the present study, half of the subjects received the incidental instructions of Brewer and Treyens, and half received our intentional instructions. The 35-s exposure condition was used in all conditions in Experiment 2 for comparability with the Brewer and Treyens study.

A third methodological difference between these two studies involves the type of recognition memory test used. The recognition test used by Brewer and Treyens was a verbal rather than a visual recognition test. That is, their subjects received a list of words and checked off the items they remembered from the experimental setting. This test thus assessed inventory information, whereas the test used in the present study assessed descriptive information. Brewer and Treyens did not assess whether subjects who recognized that they had seen an office chair, for example, could recognize the physical features of that office chair.

In addition to the present study, two other studies of the consistency effect have examined recognition memory for descriptive information in pictures of scenes using token, changed, distractor items (Friedman, 1979; Goodman, 1980), and each of these studies has also reported that subjects more accurately detect changes in unexpected items than in expected items. Thus discrepancies between the recognition memory results of Brewer and Treyens and these other studies may in part be due to differences in the type of recognition memory test utilized. To test this hypothesis, subjects in the present study received both a verbal recognition memory test and a visual recognition memory test. Experiment 2 did not include a recall test.

**Method**

**Subjects and Design**

The subjects were 80 undergraduates who volunteered from classes at the Claremont Colleges. Subjects viewed either the full graduate student’s office or the simplified graduate student’s office used in Experiment 1, with intentional or incidental instructions. They were tested immediately afterward (no delay test) with both a verbal recognition test and a visual recognition test. Each room contained the same 8 consistent and 8 inconsistent items, and each test assessed memory for only these 16 items. The design was thus a 2 (room) x 2 (instructions) x 2 (consistency) mixed factorial design, with the first two factors manipulated between subjects and the third factor manipulated within subjects.

**Stimulus Items and Settings**

The same 8 consistent and 8 inconsistent items from Experiment 1 were used in Experiment 2. Half of the subjects viewed these 16 items placed in the graduate student’s office used in Experiment 1. The other half of the subjects viewed these 16 items placed in a different graduate student’s office, approximately the same size as and next door to the first office. This second office was, in fact, a real office shared by three graduate students. It included three desks with chairs, one tall book case, two four-drawer file cabinets, two windows, and a built-in wooden cabinet across the back wall. All of these surfaces were covered with the normal impedimenta of a graduate student’s office—computer printouts, pads, bookbags, personal pictures, lunch bags, pens and pencils, stacking trays, clip boards, tape cassettes, research materials, sweaters, and so forth. No items that duplicated the target, consistent or inconsistent items, were included in the room. For the presentation phase, in each of the two rooms the same 16 target items were placed throughout the room, primarily on room frame objects, in the same position for each subject.

**Procedure**

**Presentation phase.** Subjects participated in small groups as in the first experiment. Before entering the room, subjects were given either the intentional instructions used in Experiment 1, or they were given incidental instructions. The incidental instructions mentioned that a previous group of subjects had not quite finished their task. Thus, the subjects were instructed, “Please have a seat in the next office and wait for just a minute.” They were then seated in the experimental room in which chairs were lined up for the subjects just inside the doorway. They were told not to talk to each other because
the subjects in the next room could hear them. After 35 s, the
experimenter entered the room and asked the subjects to come into
an adjacent room and sit at the table.

Test phase. The test phase consisted of a verbal recognition test
followed by a visual recognition test. In the verbal recognition test,
subjects were given a response sheet with the names of 32 items
written on it. Half of the items had been in the room, and half were
new. Half of the new items were items that would be consistent with
expectations about what would be in a graduate student's office, and
half were inconsistent with expectations. Subjects were instructed to
decide if each of the 32 test items was old or new and to circle the
corresponding response.

The new items were primarily drawn from items that were rated
in Experiment 1 but were not presented in Experiments 1 or 2. The
inconsistent new items were items that also were consistent with
expectations about what would be in a preschool classroom, because
this is how the inconsistent target items were selected in Experiment
1.

As in Experiment 1, while subjects completed the first test a second
experimenter entered the experimental room through another door
and substituted token changed items for 4 consistent and 4 inconsist-
ent target items. They also placed a numbered card, 1 to 16, in front
of the test items. When subjects completed the verbal recognition
test, they were led back into the room for the visual recognition test.
In the test room were 16 items; half were old and half were token
changed new items. Subjects were instructed that they would see 16
items with a number in front of each. They were to look at each item,
and substitute token changed items for 4 consistent and 4 inconsist-
ent new items. The only other significant effect on the false alarm
rate for inconsistent items was similar when subjects had viewed the
simplified room (p = .28) and the full room (p = .23). However, the false alarm
rate for consistent items was greater when subjects had viewed the
visual recognition memory test (F = 1.86).

Table 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Simplified room</th>
<th>Full room</th>
<th>Visual recognition d'</th>
<th>Intentional</th>
<th>Incidental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>2.71</td>
<td>2.99</td>
<td>2.30</td>
<td>0.77</td>
<td>1.48</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>1.44</td>
<td>1.73</td>
<td>2.53</td>
<td>2.63</td>
<td>3.87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Simplified room</th>
<th>Full room</th>
<th>Verbal recognition d'</th>
<th>Intentional</th>
<th>Incidental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>2.56</td>
<td>2.51</td>
<td>0.94</td>
<td>1.63</td>
<td>0.92</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>0.94</td>
<td>2.10</td>
<td>0.57</td>
<td>2.10</td>
<td>0.90</td>
</tr>
</tbody>
</table>

The interactions between instructions and consistency, room type and consistency, and Room Type × Instructions × Consistency address whether differences in the results be-
tween the present study and that of Brewer and Treyens (1981) are based upon methodological differences between
the two studies. None of the interactions involving consistency approached significance. Thus, the superior recognition of
inconsistent over consistent items was not affected by either instructions or the ratio of inconsistent to consistent items in
the room.

Separate 2 (room) × 2 (instructions) × 2 (consistency) analyses of variance were also performed on the hit rate and
false alarm rate data. Only instructions produced a significant
effect in the analysis of the hit rate data. Hit rates were higher
with intentional (p = .84) than with incidental instructions
(p = .74), F(1, 76) = 5.61, MS_e = 0.07. As in Experiment 1,
consistent items were slightly better recognized when they had been seen in the simplified room (d' = 1.50); however, the names of consistent items were recognized
far better when they had been seen in the simplified room (d' = 1.75) than in the full room (d' = 0.08).

Separate 2 (room type) × 2 (instructions) × 2 (consistency) analyses of variance were also performed on the hit rate and
false alarm rate data. As on the visual recognition test, verbal recognition hit rates did not significantly differ as a function of item consistency \((F = 1.26)\). However, there were significant effects of room type, \(F(1, 76) = 8.42, MSe = 0.03\), and instruction, \(F(1, 76) = 23.76\), and the three-way interaction of Room Type \(\times\) Instructions \(\times\) Consistency was also significant, \(F(1, 76) = 13.23, MSe = 0.02\).

The analysis of the false alarm rate data yielded a significant main effect of consistency, \(F(1, 76) = 77.25, MSe = 0.03\). False alarm rates were greater to the names of consistent \((p = .45)\) than inconsistent items \((p = .21)\). The interaction of Consistency \(\times\) Room Type was also significant, \(F(1, 76) = 41.36, MSe = 0.03\). Whereas the false alarm rate to the names of inconsistent items was similar when subjects had viewed the simplified room \((p = .19)\) and the full room \((p = .23)\), the false alarm rate to the names of consistent items was far greater when subjects had viewed the full room \((p = .64)\) than the simplified room \((p = .26)\). The false alarm rate was also greater when subjects had seen the full room \((p = .44)\) than the simplified room \((p = .23)\), \(F(1, 76) = 35.93, MSe = 0.05\), and the false alarm rate was greater with incidental instructions \((p = .42)\) than with intentional instructions \((p = .24)\), \(F(1, 76) = 24.14\).

The results of Experiment 2 bolster the conclusion that inconsistent items are better recognized than consistent items, regardless of whether the test is a same-changed visual recognition test or an old-new verbal recognition test. The effect of consistency was significant on the \(d'\) measure for both types of test. Further, this effect was not influenced by methodological factors that differed between the present study and that of Brewer and Treyens (1981). On the \(d'\) measure the consistency factor did not significantly interact with either room type or instructions. And the effect of consistency was obtained with the 35-s viewing time common to Experiment 2 and the study by Brewer and Treyens.

Although there were several differences between the results on the visual recognition test and the verbal recognition test, in terms of the effect of consistency, similar patterns were obtained on both tests. First, both tests yielded significantly better \(d'\) performance for inconsistent items than for consistent items. Second, neither test produced a significant difference in hit rates between the inconsistent and the consistent items. Third, on both tests the false alarm rate was higher for consistent than for inconsistent items, and this effect interacted with room type. That is, whereas the false alarm rates for inconsistent items were similar in the simplified and the full room, the false alarm rates for consistent items were far greater in the full room, more comparable to that used by Brewer and Treyens, than in the simplified room. This last result is consistent with the Brewer and Treyens finding that in their study, the simple correlation was a very impressive +.75 between verbal recognition \((1 = \text{certainly new}; 6 = \text{certainly old})\) and schema expectancy for test items that had not been in the study room.

How can we account for the difference between the results of the present study and those of Brewer and Treyens (1981)? With several methodological explanations for this difference tested and rejected in Experiment 2, a compelling explanation is that the method of data analysis used by Breyer and Treyens may not have been sensitive to subjects’ performance on the inconsistent items. Their conclusion that “schema expectancy was positively correlated with recall and recognition” (p. 207) is based on correlations computed across 43 items (for the recall measure) and 61 items (for the recognition measure). Because only “a few” of the items in the room were inconsistent, it is possible that mean recall and recognition performance could have been greater for the inconsistent items than for the consistent items, and yet the overall correlation across the remaining majority of the items could still have been negative.

Although mean performance levels were not presented in the Brewer and Treyens study, these values were generously offered by the first author of that study. W. F. Brewer (personal communication, March 6, 1988) reported the median recall rank for the 9 objects with the lowest expectancy scores was approximately 18 (out of 61 objects). Thus, these schema-inconsistent items were recalled fairly well; they fell toward the top—not the bottom—of the distribution. Also, the mean verbal recognition performance \((1 = \text{certainly new}; 6 = \text{certainly old})\) was 3.78 for the 9 objects with the lowest schema-expectancy scores and 3.54 for the 45 remaining objects (excluding the room frame objects). Although these means are not significantly different, the difference is in the same direction reported in the present study, not in the opposite direction. Further, W. F. Brewer suggested that the mean recognition performance for the 45 objects was likely inflated relative to the comparable group in the present study because these 45 objects included both schema consistent and schema irrelevant objects. Thus, the difference between the results of these two studies is only an apparent difference resulting from the insufficient sensitivity of the Brewer and Treyens correlational analysis to detect the relative memorability of inconsistent and consistent items.

General Discussion

The principal purpose of this study was to test the generalizability of the consistency effect to real-world settings. The consistency effect predicts that items inconsistent with expectations will be recalled and recognized better than items consistent with expectations. In both experiments in the present study, the consistency effect was affirmed; items inconsistent with expectations were significantly better recalled and recognized than items consistent with expectations.

Why are inconsistent items better recalled and recognized than consistent items? Three interpretations have been discussed in the literature. One interpretation is that the consistency effect is a general form of the von Restorff effect (Koffka, 1935), in which an item that stands out against its background is better remembered than items that are similar to each other (e.g., the word peach in a list of vegetables). The von Restorff effect has been explained as (a) the novelty of the isolated item creating a stronger memory trace, and similarly, (b) less interference from related items for the isolated item. However, the von Restorff interpretation does not convincingly handle the present results. Koffka (1935) reported that the more isolated items there are in a list—that is, the larger the ratio of isolated to similar items—the smaller the memory advan-
tage for the isolated items. In Experiment 1 in the present study, both rooms included an equal number of consistent and inconsistent items, a condition that does not produce the von Restorff effect. Also, the ratio of consistent to consistent items was directly varied in Experiment 2, and this manipulation did not influence the effect of consistency.

Another interpretation of the consistency effect is that it results from a retrieval advantage for inconsistent items. Accordingly, inconsistent items are difficult to comprehend and are thus retained in working memory longer than consistent items. During this additional time in working memory, additional items are also being processed, and some of these become related to the inconsistent items in an attempt to comprehend the inconsistent item. As a result, the inconsistent items become associated with more items than do consistent items, and these additional associations subsequently can serve as retrieval cues. This retrieval interpretation has been offered by Srull, Lichtenstein, and Rothbart (1985) as an extension of the model developed by Hastie (1980) for explaining why recall is better for information incongruent with rather than congruent with expectations in person memory tasks.

However, the retrieval interpretation predicts differences between inconsistent and consistent items in recall but not recognition. Although Srull (1981) did not report a recognition memory difference between inconsistent and consistent behavioral statements about people, in the present study, as well as in others (Goodman, 1980, with pictures; Marmurek, 1984, with disposition-based categories), inconsistent items have been significantly better recognized than consistent items. A recognition advantage as well as a recall advantage for inconsistent items would be difficult to explain if the consistency effect were localized at retrieval. Also, Brewer and Treyens (1981) used a retrieval interpretation to explain why they reported a recall advantage for consistent items over inconsistent items. According to their application of the retrieval interpretation, the connections are stronger between the central schema and consistent items than between the central schema and inconsistent items. Thus, the consistent items are more likely retrieved than are the inconsistent items when the central schema is activated.

A third interpretation for the consistency effect is that it results from encoding differences for inconsistent and consistent items. Loftus and Mackworth (1978) reported that when subjects view a picture that includes a consistent item (e.g., a tractor in a farm scene) as compared with an inconsistent item (e.g., an octopus in a farm scene), they fixate earlier, more often, and with longer durations on the inconsistent item. Friedman (1979) reported similar results when subjects looked at pictures of scenes that included expected and unexpected items. But further, Friedman (1979) reported qualitative differences between processing expected and unexpected items. In a recognition memory test, she found that subjects rarely noticed missing, new, or physical changes in consistent items, whereas they almost always noticed these changes in inconsistent items.

Friedman (1979) articulated a process model that relies on differentiating between the processes of feature detection and feature analysis. Accordingly, feature detection involves processing items in a manner than results in activation of the appropriate scene frame for processing the picture (i.e., the “top” in “top down processing”). Thus feature detection processes usually operate on schema consistent items. Schema consistent items are therefore processed only to the degree necessary to activate the correct frame for the scene as a whole. On the other hand, feature analysis involves processing items in a manner that results in distinguishing the current scene from other scenes. Thus, feature analysis processes usually operate on schema inconsistent items; for instance, in the example above, it is the octopus that distinguishes one farm scene from another one. Consequently, the physical appearance of schema inconsistent items is encoded in more detail than is the physical appearance of schema consistent items.

The encoding interpretation of the consistency effect predicts an advantage for inconsistent over consistent items on tests of recognition as well as recall and thus handles the results of the present study. Accordingly, when subjects looked at the graduate student office or the preschool classroom, they encoded the consistent items in a manner that activated the correct scene frame for that room. This would involve feature detection processes with little analysis of the physical features of the schema consistent items. However, once the correct scene frame was activated, feature analysis processes were applied to items that were necessary to distinguish the current scene from others with the same scene frame. This resulted in allocating any additional processing time available to encoding the physical features of the schema inconsistent items. The schema inconsistent items were thus processed for a longer period of time, and, more important, schema inconsistent items were encoded qualitatively differently than schema consistent items; that is, schema inconsistent items were encoded in a manner that was more likely to preserve their physical appearance. Consequently, for example, subjects were (a) more likely to recall that there was a stuffed bear than a telephone in the graduate student’s office and (b) more likely to recognize when a different (token changed) stuffed bear was substituted in the test phase than when a different telephone was substituted.

It is important to differentiate between the results of the present study and results of studies that have examined memory for items as a function of their schema relevance. Maki (1987) and Mandler (1984) have convincingly argued that consistency with expectation and schema relevance are orthogonal qualities. Goodman’s (1980) sample stimulus picture, for example, presents a picture of a girl reading in a library. The book case is a schema relevant item, and the potted plant is a schema irrelevant item. Both items are consistent with what one might expect in a library, but they vary in how relevant they are to the library scene frame. The general finding in the literature is that when schema relevance is manipulated, high-relevance items are better recalled than low-relevance items, but low-relevance items are better recognized. These were exactly the results reported by Goodman (1980). On the other hand, when consistency with expectation is manipulated, as in the present study, the general finding is that both recall and recognition are greater for inconsistent than for consistent items.
More broadly, the results of the present study suggest that environmental context can affect how much of the physical details a person is likely to remember for items in the environment. These results are relevant, for example, to eyewitness memory and suggest that witnesses would be able to describe more elaborately the unexpected items at the scene of a crime (e.g., the gun carried by someone in a convenience store) than the expected items (e.g., the shopping bag carried by someone in a convenience store). This study also asserts the generalizability of results from picture memory studies to memory for real-world scenes.

References


Appendix follows on next page
Appendix

Stimulus Items Used in Experiments 1 and 2 (and Mean Consistency Ratings, Range = 1–7)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Rating in preschool classroom</th>
<th>Rating in graduate student office</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items consistent with preschool classroom and inconsistent with graduate student office</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stuffed bear</td>
<td>2.00</td>
<td>6.13</td>
</tr>
<tr>
<td>Plastic ball</td>
<td>2.31</td>
<td>6.65</td>
</tr>
<tr>
<td>Toy truck</td>
<td>1.17</td>
<td>6.25</td>
</tr>
<tr>
<td>Blocks</td>
<td>1.54</td>
<td>6.67</td>
</tr>
<tr>
<td>Coloring books</td>
<td>1.69</td>
<td>6.42</td>
</tr>
<tr>
<td>Fingerpaints</td>
<td>1.39</td>
<td>5.75</td>
</tr>
<tr>
<td>Board game</td>
<td>1.31</td>
<td>5.83</td>
</tr>
<tr>
<td>Painted picture</td>
<td>3.07</td>
<td>6.83</td>
</tr>
</tbody>
</table>

| Items consistent with graduate student office and inconsistent with preschool classroom |                                |                                  |
| Clock radio                     | 6.00                          | 3.17                             |
| Hot pot                         | 5.69                          | 2.75                             |
| Textbooks                       | 6.46                          | 1.67                             |
| Ashtray                         | 4.35                          | 1.58                             |
| Desk lamp                       | 6.71                          | 3.30                             |
| Calculator                      | 6.58                          | 2.00                             |
| Typewriter                      | 6.08                          | 2.67                             |
| Beer bottle                     | 3.69                          | 1.50                             |

Received September 30, 1987
Revision received September 20, 1988
Accepted October 6, 1988