Behavior Research Methods
2010, 42 (1), 292-301
doi:10.3758/BRM.42.1.292

Video context-dependent recall

Steven M. Smith and Isabel Manzano
Texas A&M University, College Station, Texas

In two experiments, we used an effective new method for experimentally manipulating local and global contexts to examine context-dependent recall. The method included video-recorded scenes of real environments, with target words superimposed over the scenes. In Experiment 1, we used a within-subjects manipulation of video contexts and compared the effects of reinstatement of a global context (15 words per context) with effects of less overloaded context cues (1 and 3 words per context) on recall. The size of the reinstatement effects in Experiment 1 show how potently video contexts can cue recall. A strong effect of cue overload was also found; reinstatement effects were smaller, but still quite robust, in the 15 words per context condition. The powerful reinstatement effect was replicated for local contexts in Experiment 2, which included a no-contexts-reinstated group, a control condition used to determine whether reinstatement of half of the cues caused biased output interference for uncued targets. The video context method is a potent way to investigate context-dependent memory.

Many theories of episodic memory include mechanisms involving contextual processes, such as contextual associations and contextual cuing (e.g., J. M. Eich, 1982; Gillund & Shiffrin, 1984; Hintzman, 1988; Humphreys, Bain, & Pike, 1989; Murdock, 1993; Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997; Tulving, 1983; Tulving & Thomson, 1973). *Contextual drift*—the fluctuation of mental contexts over time—has been theorized to influence encoding and retrieval in ways that affect such phenomena as discrimination, repetition effects, recency, output interference, reminiscence, directed forgetting, and false memories (e.g., Anderson & Bower, 1972; Glenberg, 1979; Glenberg, Bradley, Kraus, & Renzaglia, 1983; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988; Sahakyan & Kelley, 2002; Shiffrin & Steyvers, 1997; S. M. Smith & Vela, 2001). Clearly, the contexts in which events occur play an important role in human memory.

Unfortunately, there are few, if any, laboratory methods for manipulating environmental contexts in ways in which they powerfully influence recollection. Reinstating the same environmental context at test that was present at encoding has been shown to benefit memory for many types of contexts, including contexts operationally defined in terms of the experiment room (e.g., S. M. Smith, 1979; S. M. Smith, Glenberg, & Bjork, 1978), natural environments (e.g., Godden & Baddeley, 1975), ambient odors (e.g., Herz, 1997; D. G. Smith, Standing, & de Man, 1992), and background music (e.g., Balch, Bowman, & Mohler, 1992; S. M. Smith, 1985). Although context reinstatement usually benefits episodic memory (S. M. Smith & Vela, 2001), there have been many studies in which such effects were not found, including not only failures associated with recognition memory tests (e.g., Godden & Baddeley, 1980; Jacoby, 1983; S. M. Smith et al., 1978), but also studies in which recall tests were used (e.g., E. Eich, 1985; Fernandez & Glenberg, 1985). Summing across 93 published effect sizes, S. M. Smith and Vela concluded that incidental environmental contexts have a small but reliable effect on memory, with an average effect size of $d = 0.28$.

A common way to study context reinstatement effects has been through physical reinstatement, in which participants are physically immersed in varying environmental contexts. In these studies, the participants encode material in one environmental context and are then tested either in the same environmental context or in a different context (e.g., Godden & Baddeley, 1975; S. M. Smith, 1979, 1986). With this method, one can examine the effects of global contexts (Glenberg, 1979)—that is, cases in which a single context is associated with many memory targets. In one classic study, scuba divers heard a list of words while they were either underwater or on dry land; recall was better when the divers were tested in the same environment in which encoding occurred than when they were tested in a different environment (Godden & Baddeley, 1975). A number of researchers have used room manipulations to examine the effects of context reinstatement (e.g., Fernandez & Glenberg, 1985; Isarida & Isarida, 2004; S. M. Smith, 1979; S. M. Smith et al., 1978). In these experiments, the participants studied lists of words in one laboratory room and recalled the words either in another perceptually distinct room or in the same room in which encoding took place. Although large effects have been reported occasionally for room reinstatement effects, most effects have been modest or nonexistent (see S. M. Smith & Vela, 2001).

S. M. Smith, stevesmith@tamu.edu

Limitations of typical context manipulations include (1) the possibility that global context cues may be overloaded (e.g., Watkins & Watkins, 1975) (that is, if too many target words are associated with one context cue, that overloaded cue may have only a weak effect of evoking a memory; e.g., Isarida & Isarida, 2007; Rutherford, 2004); (2) incidental contexts might not be well encoded; and (3) the participants may spontaneously try to mentally reinstate the study context when tests take place in a new environmental context, thereby weakening or eliminating the effects of experimental manipulations of context (e.g., S. M. Smith, 1979, 1984).

A method commonly used for examining context-dependent memory has been to present words at study and test on a screen, using one simple context or another at study and test. Simple contexts manipulate features such as font color, background screen color, or screen location (e.g., Dulsky, 1935; Isarida & Isarida, 2007; Macken, 2002; Murnane & Phelps, 1993, 1994, 1995; Murnane, Phelps, & Malmberg, 1999; Rutherford, 2004; Weiss & Margolius, 1954). With such methods, local contexts—that is, contexts that can change from one event to the next—are usually examined (Glenberg, 1979; Rutherford, 2004). In a series of studies of context-dependent recognition, Murnane and colleagues, using simple local contexts, failed to show any context-dependent discrimination effects (i.e., an increase in hit rate greater than the increase in false alarms when simple contexts are reinstated at test), although their results often showed increases in both hits and false alarms when simple contexts were reinstated at test (Murnane & Phelps, 1993, 1994, 1995; Murnane et al., 1999). Only when screen contexts were *perceptually rich*—that is, when they were line drawings of physical settings—did Murnane et al. find context-dependent discrimination effects. Although others (e.g., Isarida & Isarida, 2007; Macken, 2002; Rutherford, 2004) have found context-dependent recognition memory with simple contexts, the effects have been rather small. For example, Rutherford found context-dependent discrimination effect sizes ranging from small ($d = 0.01$) to moderate ($d = 0.34$), and Macken's largest context-dependent discrimination effect size was moderate ($d = 0.56$).

Although there is no question that global encoding and test contexts influence episodic memory, it has been difficult to experimentally study variables that might moderate the effects, because the standard effect is so often small. For example, Cousins and Hanley (1996) studied whether relational processing at encoding would nullify the effects of reinstated environmental context cues at test, a prediction of the outshining hypothesis[1] (e.g., S. M. Smith, 1988; S. M. Smith & Vela, 2001). In discussing the results of their Experiment 1, Cousins and Hanley stated,

> The results show quite clearly that the number of words recalled was not affected by whether or not subjects attempted to remember the list items in the room in which they were learned. Even subjects who performed an individual item processing task at encoding did not benefit from contextual reinstatement. The recall results, therefore, have provided no support for the outshining hypothesis. (pp. 83–84)

Given an unreliable method for evoking the standard context-dependent recall effect—room manipulations—we cannot easily examine factors that might modulate these weak effects.

In the present experiments, we tested a new method for manipulating contexts, presenting target words on a screen superimposed over videotaped segments of real environments and cuing recall with the same video scenes. These video scenes were brief movies of places that were not likely to be familiar to the participants but that showed situations that were likely to be familiar to the participants. For example, the video scenes included brief footage from a softball game, a restaurant, a carousel ride, windmills, an elevator, a parking lot, a mountain stream, a commuter train, or a roomful of people on exercise machines. In addition to the visual richness to these scenes, there was movement and action, as well as accompanying sounds, making these contexts perceptually multimodal. Although the scenes were examples of types of situational contexts familiar to our experimental participants, the specific places were videotaped in geographic locations far from the participants' campus.

Because the scenes were digitized, they could be presented and changed much more rapidly and fluently than physical environments can be changed. In most of the treatment conditions of the two experiments reported here, a single target word was presented with each video context; the words came into view without transition, using the *Movie Title* function of a software package for editing video. Whereas past experiments have often used entire lists of words associated with a single global environmental context, possibly leading to overloaded context cues, in the present experiments, we compared a global context cue with a 15:1 target:context ratio with local context cues, including 3:1 and 1:1 target:context ratios, which lessened the overloading of context cues. Furthermore, the present method made spontaneous mental reinstatement of contexts less likely when numerous contexts were used at encoding, making spontaneous recollection of contexts difficult (e.g., S. M. Smith, 1979). Finally, the participants in the present experiments were instructed to attend to both target words and scenes, thereby increasing the likelihood that contexts were encoded. Thus, the three major obstacles for observing context-dependent memory effects—cue overload, spontaneous mental reinstatement of contexts, and failures of context encoding—were addressed by our video context method.

Our approach is not radically different from other context manipulations in which target words were associated with contexts at study and in which the targets were cued at test (e.g., Dulsky, 1935; Isarida & Isarida, 2004; Murnane & Phelps, 1994; Rutherford, 2004; Weiss & Margolius, 1954). Providing memory cues at test, as was done in these studies, is an experimental method with a long history (e.g., Calkins, 1896; Thomson & Tulving, 1970; Tulving & Osler, 1968). In cued recall, stimuli, such as paired associates, category cues, extralist semantic associates of target words, or pictures, might be provided or withheld at test in order to determine the effectiveness of the cues. The methods used in the present study resemble the cued recall methods used in investigations of *encoding specificity*, the theory that memory cues for events can be effective only if those cues were originally encoded in relation to the target events (e.g., Thomson & Tulving, 1970; Tulving & Osler, 1968). For example, in some conditions of Tulving and Osler's study, each target word was studied along with a weakly related word cue, and during the recall test, either the cue words that were present at study or different unstudied cue words were provided. In the experiments reported here, the method was similar to this cued recall procedure, except that the cues were multimodal video scenes, and in some treatment conditions, multiple target words were associated with each video context. Thus, our method begins to bridge the gap between methods in which ecologically realistic environmental cues, such as rooms or outdoor settings (e.g., Godden & Baddeley, 1975; S. M. Smith et al., 1978), are used and methods that involve more highly controlled laboratory stimuli. In the present experiments, we tried to create the most powerful context-dependent memory effect that we could, so that subsequent researchers can better determine what factors moderate the effect.

Contextually rich and perceptually multimodal contexts in the form of video scenes were shown, each with a corresponding word superimposed in red letters over the scene. The scenes were shown for at least 5 sec each—sufficient time for context encoding, according to Malmberg and Shiffrin's (2005) one-shot model. Furthermore, context reinstatement at test was manipulated within subjects, making context familiarity equivalent for reinstated and nonreinstated scenes. In both of the reported experiments, memory was tested with recall, a test that has shown high susceptibility to context-dependent memory effects in published studies (e.g., S. M. Smith, 1979, 1982, 1984, 1986; S. M. Smith et al., 1978; S. M. Smith & Vela, 2001). It was predicted that studied words corresponding to video scenes reinstated at test would be recalled better than words corresponding to nonreinstated scenes.

EXPERIMENT 1

In Experiment 1, participants studied a single list of 30 words, followed by a recall test. At study, the words were superimposed over movie scenes, and the scenes selected were unrelated to the target words in any obvious way. These movie scenes were amateur videos and included no plot or dialogue. They included background sounds and were scenes of events that one might encounter on a daily basis (e.g., cars on a highway, a person walking down a sidewalk, people playing baseball on a field). At the time of the recall test, half of the studied movie scenes were shown without the target words. The participants were asked to recall all of the list words, not just those that had been associated with the videos shown at test. Half of the studied scenes were randomly selected to be test cues for one counterbalancing condition, and the other half were shown at test for a second counterbalancing condition.

The number of video contexts used for the 30-word list was also varied. In the most overloaded condition, 15 words were studied superimposed over a single video context, and the second 15 words were shown over a second context. In a less overloaded condition, each video clip was associated with 3 list words, and in the least overloaded condition, each context was associated with only 1 word. At test, half of the studied scenes were shown as recall cues: only 1 video (of the original 2) in the most overloaded condition, 5 videos (of the original 10) in the less overloaded condition, and 15 (of the original 30) in the least overloaded condition. It was predicted that reinstatement effects (i.e., differences in recall between words corresponding to reinstated scenes vs. words corresponding to nonreinstated scenes) would be greater for less overloaded contexts.

Method

Participants. A total of 324 Texas A&M University undergraduate students participated in this experiment in return for partial course credit. Participation was voluntary, and other options were available to earn equal credit. The participants self-enrolled in the two different counterbalancing conditions. The number of participants in each experimental session depended on the random enrollment of the participants, and varied from 10 to 20 participants per session. There were 108 participants in the 1-word condition, 108 in the 3-word condition, and 108 in the 15-word condition.

Design and Materials. The experiment was a 2 × 2 × 3 mixed design. Reinstatement (scenes reinstated vs. no reinstatement) served as the within-subjects variable; word order (Order X vs. Order Y) and context load (1 word vs. 3 words vs. 15 words) served as the between-subjects variables. The proportion correct on the recall test was the dependent variable.

Thirty words were derived from the MRC Psycholinguistic Database with written frequencies ranging from 50 to 100 (Kučera–Francis frequency norms). The words were all concrete nouns, and word length varied from five to nine letters. The 30 background movie scenes associated with the words were randomly paired, although obvious relationships were avoided. The scenes were simple everyday scenes (e.g., a park, a stairway, a kitchen, driving down the highway). Each of the 30 target words was shown 1 at a time, in red letters superimposed over the movie clips. The target words were displayed using the *Movie Title* function of the CyberLink PowerDirector software package for editing video recordings; each word appeared, without transition, in large letters in the center of the screen, with the video clips showing in the background.

Procedure. The participants were tested in groups of 10–20 people, depending on participant enrollment, and were seated in front of a large video screen. They were told that they would study a list of words superimposed on background movie scenes and that they should try to remember both the words and the movie scenes for a later memory test. Because we manipulated study context load, the number of video contexts used for the 30-word list varied. At study, one group saw the 30 words superimposed over 30 different movie scenes, 1 word at a time. The words and movie scenes were shown for 5 sec each. A second group saw the same 30 words superimposed over only 10 movie scenes, such that 3 words appeared over each movie scene, 1 word at a time. In this case, each of the 10 scenes appeared for 15 sec, and each of the 3 words appeared over the movie clips for 5 sec. Finally, a third group saw the same 30-word list, but only 2 movie scenes, so that 15 words corresponded to 1 movie scene, and the remaining 15 words corresponded to a 2nd movie scene; this was the most overloaded condition. In this condition, each of the 2 scenes appeared on the screen for a total of 75 sec, and each of the words appeared on the screen for 5 sec. In addition to the manipulation of total word load on each context, we counterbalanced the word order; half of the participants in each of the three load conditions received the words in one order (Order X), and the other half of the participants saw the words in a different randomly determined order (Order *Y*). The movie scenes always appeared in the same order in each word order condition.

After the study phase, the participants were given an immediate memory test. The participants were given a blank sheet of paper and were told that they would see some of the movie scenes viewed earlier. The participants were told that while the movie scenes were repeating, they were to write down as many words as they could remember from the study list.

At test, the three groups were shown only half of the studied scenes as recall cues. The participants in the most overloaded condition (15 words per scene) were shown only 1 of the 2 scenes that they were shown at study. The participants in a less overloaded condition (3 words per scene) were shown only 5 of the 10 scenes that they saw at study. Finally, those participants in the least overloaded condition (1 word per scene) were shown 15 of the 30 scenes that they saw at study. The recall test lasted 2.5 min, the amount of time that it took for two complete repetitions of the movie clips selected for each of the three conditions. At test, the movie scenes were shown in a single fixed order for all conditions. Half of the participants saw one set of scenes (arbitrarily designated to be the *A scenes*), and half saw a different set of scenes (arbitrarily designated to be the *B scenes*). The A scenes were the scenes from the first half of the presentation list, and the B scenes were from the second half. These reinstated test scenes were shown in the same order as they had been viewed in the original presentation.

Results

An ANOVA was used to test the effects of reinstatement, the load on each context (i.e., the number of words per context), and word counterbalancing condition on recall. A $2 \times 2 \times 3$ mixed ANOVA was performed, using proportion recalled as the dependent measure. The number of words per context (1 vs. 3 vs. 15) and word order (*X* vs. *Y*) were between-subjects variables, and reinstatement (reinstated scenes vs. nonreinstated scenes) was a within-subjects variable. The analysis showed a main effect of reinstatement [$F(1,318) = 648.19$, $p = .000$, $\eta^2 = .67$]; words corresponding to reinstated scenes were recalled far more often than words corresponding to nonreinstated scenes (see Figure 1). The effect size for this reinstatement effect, using Cohen's *d* (Cohen, 1992) was $d = 1.11$. Cohen characterized *d* values of 0.2 as small effects, 0.5 as medium effects, and 0.8 or greater as large effects, so the observed effect was clearly a large one.

There was also a significant main effect of context load [$F(2,318) = 27.02$, $p = .000$, $\eta^2 = .15$], showing that the fewer words associated with each context, the greater was the proportion recalled (see Figure 1). There was no main effect of word order [$F(1,318) = 0.28$, $p = .60$, $\eta^2 = .00$].

The interaction of reinstatement with context load was significant [$F(2,318) = 60.77$, $p = .000$, $\eta^2 = .28$]; reinstatement effects diminished as more words were associated with a context (Figure 1). The interaction of reinstatement with word order was not significant [$F(1,318) < 1$], nor was there a significant three-way interaction of reinstatement, context load, and word order [$F(2,318) < 1$].

A priori pairwise comparisons, using a familywise corrected significance level of $p < .017$,[2] showed that context reinstatement effects were significant for all three context load conditions. For the heavily loaded condition (i.e., 15 words per context) [$t(108) = 7.24$, $SE = .02$, $p = .000$], for the moderately loaded condition (i.e., 3 words per context) [$t(108) = 14.47$, $SE = .03$, $p = .000$], and for the least loaded condition (i.e., 1 word per context) [$t(108) = 21.87$, $SE = .02$, $p = .000$]. Estimates of the size of reinstatement effects observed for each level of cue load are shown in Table 1.

Discussion

The context reinstatement effects found in Experiment 1 were large, according to Cohen's (1992) guidelines for effect sizes. Even in the global context condition, in which there were 15 words in each of two video contexts, the words associated with the one reinstated context were

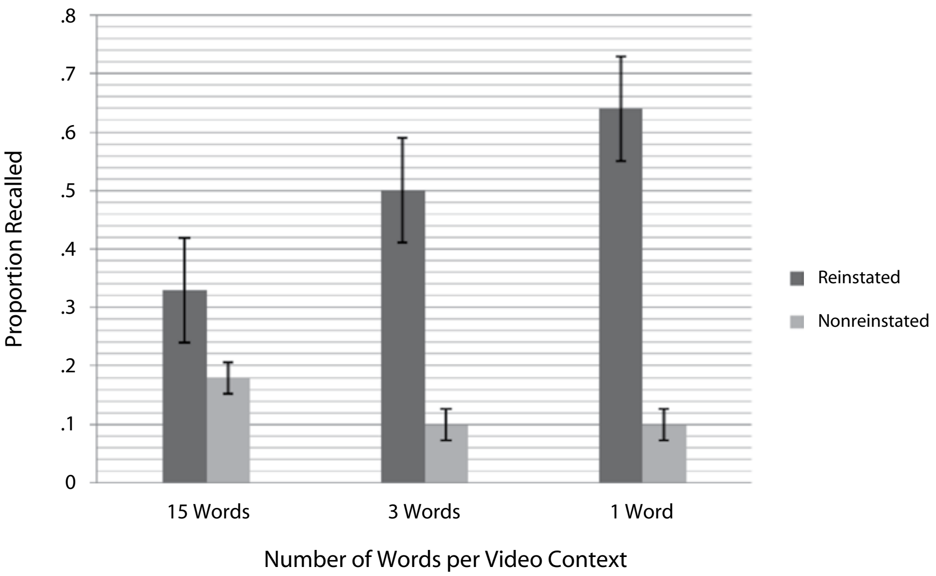

Figure 1. Mean proportions recalled in Experiment 1 as a function of test scene reinstatement and number of words per scene. Error bars denote standard errors of the means.

Table 1
Cohen's *d* for Context Reinstatement Effects in Experiment 1 As a Function of Contextual Load

Contextual Load	Cohen's *d*
1 word per scene	3.02
3 words per scene	2.18
15 words per scene	1.00

recalled nearly twice as often as the words corresponding to the nonreinstated scene. When there were 3 words per context, this increase was fivefold, and when there was 1 word per scene, the increase was more than sixfold. The corresponding effect sizes, in terms of Cohen's *d*, are shown in Table 1. The modulation of context reinstatement effects by context cue load was quite clear in the results of Experiment 1; more overloaded context cues produced smaller reinstatement effects, as was predicted. This result is consistent with similar findings by Rutherford (2004), who used recognition memory tests. Although recall was less affected by context reinstatement in the most overloaded condition, the effect in that condition was nevertheless large ($d = 1.00$).

One concern about the interpretation of the results of Experiment 1 is that the participants may not have understood that they should recall both words associated with nonreinstated scenes and words linked to reinstated ones. The participants were instructed to recall all of the words that they had seen, but there might be a question about the clarity of the recall instructions. A second concern is that the reinstatement effects might be attributed entirely to output interference; that is, the participants may have recalled words associated with reinstated scenes first, before going on to recall words linked to nonreinstated scenes, causing greater output interference for the nonreinstated condition. These concerns were addressed in Experiment 2.

EXPERIMENT 2

In Experiment 2, as in Experiment 1, participants studied a single list of 30 words (the same study list used in Experiment 1) superimposed over movie scenes, followed by a free recall test, during which half of the scenes were reinstated. There were, however, a few changes from Experiment 1. One change was that in Experiment 2, we used only a local context condition—that is, a condition in which there was 1 word per video context at encoding. A second change was the selection of test contexts. In Experiment 1, the videos shown at test were either from the first half of the original presentation order or from the second half. This was done to make context cuing more similar in the 30-context than in the 2-context conditions, because in the 2-context condition (i.e., 15 words per context), only a single video (either the first or the second encoding context) was reinstated at test. In Experiment 2, test videos were selected as every other scene from the original presentation; in one counterbalancing condition, the odd-numbered scenes were reinstated at test, and in the other counterbalancing condition, the even-numbered scenes were reinstated. A third change was the use of an additional instruction at study that asked the participants to think of a relation or association between the words and movie clips. This was done in an attempt to strengthen the relation between the context and target words in order to get a more robust reinstatement effect. Another change included an instruction at test that repeated and emphasized that the participants should recall both words corresponding to nonreinstated scenes and those that were linked to reinstated scenes. Finally, a third condition was added to the design, in which no scenes were reinstated at test. This no-contexts-reinstated condition was used because having no test scenes should avoid the potential output order bias that the participants might experience if they were given half of the scenes at test.

As in Experiment 1, it was predicted that target words that had been associated with test scenes would be recalled better than words whose scenes were not reinstated—a context reinstatement effect. The potential effects of biased output orders (and therefore more output interference) were also examined.

Method

Participants. A total of 78 Texas A&M University undergraduate students participated in this experiment in return for partial course credit. Participation was voluntary, and other options were available to earn equal credit. There were 26 participants in the A-scenes-reinstated condition, 26 in the B-scenes-reinstated condition, and 26 in the no-scenes-reinstated condition.

Design and Materials. The experiment was a 2×3 mixed design. Word subset (A words vs. B words) was a within-subjects variable, and reinstatement (A scenes reinstated vs. B scenes reinstated vs. no scenes reinstated) was a between-subjects variable. The proportion correct on the recall test was the dependent variable.

The word list was the same list used in Experiment 1, and the 30 background movie scenes were randomly selected. The scenes were simple everyday scenes; most were the same as those used in Experiment 1, and a few of the scenes were new. The 30 video scenes used in Experiment 2 can be downloaded from www.tamu.edu/faculty/stevesmith/BRM_Files/. Each of the 30 target words was shown 1 at a time in red and was superimposed over a movie scene. That is, there was a 1:1 ratio of words to movie scenes, with each scene paired with 1, and only 1, word in all conditions.

Procedure. The participants were tested in small groups of 10–20 people that varied on the basis of student enrollment and were seated in front of a large video screen. They were told that they would study several lists of words superimposed on background movie scenes and were asked to try to remember the words and movie scenes for a later memory test. Additionally, the participants were instructed to think of a relation or association between the word and the scene and that making these relations might help on the subsequent memory test. The words and movie scenes were presented for 5 sec each, and all of the participants saw all 30 background movie scenes and words.

After the study phase, the participants were given an immediate recall test. They were told that they would see some of the movie scenes that they had seen earlier and were given a blank sheet of paper. They were asked to write down as many words as they could remember in any order, and they were provided with an additional instruction: "Make sure that you try to recall ALL of the words, not just the ones corresponding to the scenes that you are about to see." During the recall test, the participants in each condition saw 15 randomly chosen movie scenes (e.g., either Set A or Set B), and each of the 15 movie scenes was played for 3 sec. The participants were given a total of 3 min for this recall test, the equivalent of four

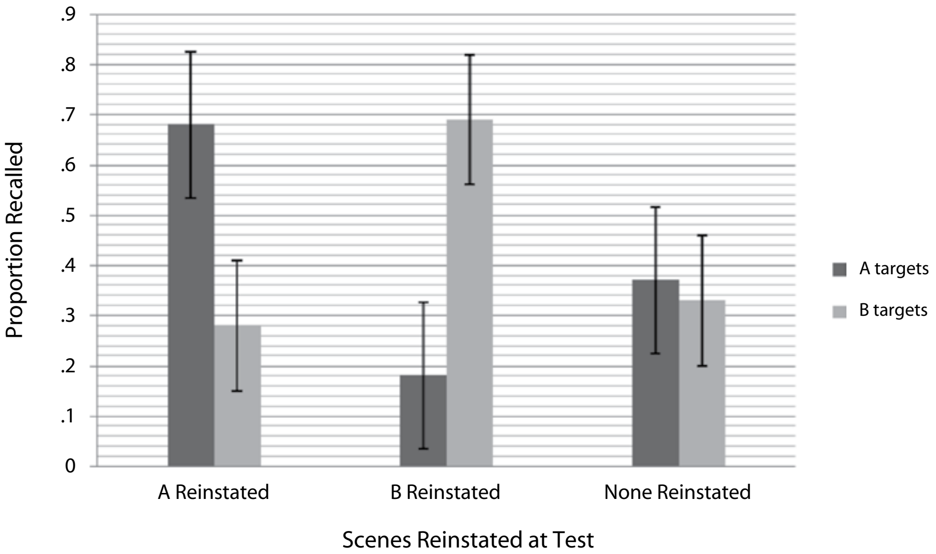

Figure 2. Mean proportions recalled in Experiment 2 as a function of scenes reinstated at test and word counterbalancing condition. Error bars denote standard errors of the means.

complete repetitions of the 15 reinstated movie scenes, and the presentation order of these scenes was the same in all four repetitions. The participants in the no-contexts-reinstated condition saw a blank screen during the recall test and were given 3 min to try to recall as many words as they could remember.

Results

An ANOVA was used to test the effects of reinstatement and word counterbalancing condition on recall. A 2 × 3 mixed ANOVA was performed, using proportion recalled as the dependent measure. Reinstatement (A scenes reinstated vs. B scenes reinstated vs. no scenes reinstated) was a between-subjects variable, and word counterbalancing condition (Counterbalancing Condition A vs. B) was a within-subjects variable. The analysis showed a main effect of reinstatement [$F(1,75) = 7.21$, $p = .001$, $\eta^2 = .16$], apparently due to better recall for words corresponding to reinstated scenes than for those corresponding to nonreinstated items (Figure 2). There was no main effect of word counterbalancing condition [$F(1,75) = 1.10$, $p = .299$, $\eta^2 = .01$], showing that recall performance was no different for Counterbalancing Condition A or B words. The interaction of counterbalancing condition and reinstatement was significant [$F(1,75) = 156.70$, $p = .000$, $\eta^2 = .81$]. When A scenes were reinstated, the participants recalled more Counterbalancing Condition A words than Counterbalancing Condition B words, whereas when B scenes were reinstated, the participants recalled more Counterbalancing Condition B words than Counterbalancing Condition A words. When no scenes were reinstated, recall for Counterbalancing Condition A words and Counterbalancing Condition B words did not differ.

Six a priori pairwise comparisons were computed, using a familywise corrected significance level of $p < .008$ (see note 2). Recall of Counterbalancing Condition A words was significantly better when A scenes were reinstated than when B scenes were reinstated [$t(40.94) = 11.43$, $SE = .04$, $p = .000$] (see Table 2 for corresponding effect size). Recall of Counterbalancing Condition A words was also significantly better when A scenes were reinstated than when no scenes were shown at test [$t(50) = 6.51$, $SE = .09$, $p = .000$] (see Table 2). Finally, recall of Counterbalancing Condition A words was significantly better when no scenes were shown at test than when B scenes were reinstated at test [$t(50) = 5.84$, $SE = .04$, $p = .000$].

Recall of Counterbalancing Condition B words was significantly better when B scenes were reinstated than when A scenes were reinstated [$t(50) = 9.05$, $SE = .04$, $p = .000$] (see Table 2 for corresponding effect size). Recall of Counterbalancing Condition B words was also significantly better when B scenes were reinstated than when no scenes were shown at test [$t(50) = 8.14$, $SE = .04$, $p = .000$] (see Table 2). Finally, recall of Counterbalancing Condition B words did not differ between when no scenes were shown at test and when A scenes were reinstated at test [$t(50) = 1.35$, $SE = .04$, $p = .183$].

Discussion

The results from Experiment 2 demonstrate a robust context reinstatement effect; words associated with rein-

Table 2
Cohen's *d* for Context Reinstatement Effects in Experiment 2 As a Function of Word Counterbalancing Condition and Type of Comparison

Counterbalancing Condition	Comparison	Cohen's *d*
A	A reinstated vs. B reinstated	2.34
	A reinstated vs. none reinstated	1.84
B	B reinstated vs. A reinstated	3.40
	B reinstated vs. none reinstated	2.30

stated scenes were recalled at more than double the rate of recall of items that were not associated with the scenes shown at test. The effect sizes for these reinstatement effects ($d = 2.34$ for Counterbalancing Condition A words, $d = 3.40$ for Counterbalancing Condition B words; see Table 2) were even larger than the corresponding effects found in Experiment 1.

The potential contributions of three artifacts to the large context effects seen in Experiment 1 were examined in Experiment 2. First, an instruction included in Experiment 2, but not in Experiment 1, explicitly and emphatically directed the participants to recall all of the presented words, not merely the ones associated with the video scenes provided at test. Clearly, this added instruction did not diminish the context reinstatement effect sizes in Experiment 2. Second, the use of context cues from either the first half or the second half of the presentation list in Experiment 1 was clearly not an essential feature in the procedure; the cues in Experiment 2 were drawn from every other context–word pair, spanning all 30 presentation positions.

Third, the contribution of output interference to the reinstatement effects, potentially caused by a cue-biased output order, was apparently minor. Recall levels for words associated with nonreinstated scenes were examined; we compared the control condition (in which no scenes were reinstated) with conditions in which half of the scenes (i.e., either the A scenes or the B scenes) were reinstated. Although the difference was significant for one counterbalancing condition (indicating a contribution of output interference to the reinstatement effect), it was not significant for the other counterbalancing condition (indicating little effect of cue-biased output interference). Furthermore, recall of words associated with reinstated scenes in both counterbalancing conditions was far greater than recall of the same words in the no-contexts-reinstated condition; this difference was 30% in Counterbalancing Condition A, and 35% in Counterbalancing Condition B. Finally, the small benefit for uncued recall in the no-contexts-reinstated group may not be due to decreased output interference, but to mental reinstatement of nonprovided video contexts at test. Mental reinstatement of video contexts could be easier if one sees no video contexts at test, because watching videos during the test might interfere with mental reinstatement of contexts. It is not clear which of these explanations better accounts for this small effect.

Several factors in Experiment 2 probably contributed to the context reinstatement effects that were found. We used local contexts in a 1:1 context:target ratio that precluded cue overload effects, thereby enhancing the power of context cues to evoke associated words. The use of many (30) different contexts likely made it difficult for the participants to recall or mentally reinstate many of the nonprovided contexts; the lack of either physically or mentally reinstated context cues contributed to the low recall levels of words associated with nonreinstated contexts. In addition, the video contexts were salient and perceptually rich, and the participants were encouraged to encode the video contexts. It is not clear which of these factors caused such large reinstatement effects. It is clear, however, that the video contexts method used in the present experiments is ideal for producing quite powerful context-dependent memory effects.

GENERAL DISCUSSION

The video context method used in the two experiments reported here produced consistent and robust context-dependent memory effects. In both Experiments 1 and 2, context reinstatement effect sizes greater than $d = 2.0$ were found, and even the global context condition of Experiment 1 produced a large effect size of $d = 1.0$. Words associated with video scenes that were provided at test were recalled several times more often than words associated with nonreinstated contexts, and the effect was seen across all list serial positions.

This method, which involves superimposing verbal stimuli over videotaped movie scenes of everyday settings, may have produced such robust effects for several reasons. One reason may be that the contexts, which, in most conditions, were manipulated in a 1:1 ratio with to-be-remembered words, were less overloaded as cues (e.g., Watkins & Watkins, 1975) than is the case for most global context-dependent memory studies. A second reason for these powerful effects may be that the great number of contexts that were used made it difficult for the participants to mentally reinstate all of the context cues. Spontaneous mental reinstatement of contexts can nullify the effects of experimental manipulations of contextual stimuli (e.g., S. M. Smith, 1979). That is, participants, uninstructed, might decide to recall the nonreinstated contexts as a strategy to aid recall. If the participants were successful at such a self-generated context cuing strategy, it would nullify the effect of experimenter-provided context cues. S. M. Smith (1979) showed that when many environmental contexts had been experienced, it was difficult for participants to generate their own context cues from memory. Except for the global cue condition in Experiment 1, the conditions in Experiments 1 and 2 in the present study involved the use of many encoding contexts. Furthermore, it may be that generating one's own context cues from memory could have been especially difficult while the participants were viewing half of the video scenes at test. That is, perceiving some video contexts could have supplanted memories of other contexts, because perception and memory of environments appear to use a common pool of cognitive resources (e.g., Glenberg, 1997). A third reason for the large effects may have been due to the instructions to attend to the video scenes at encoding; if the instructions had not directed attention to the contexts, as in studies of incidental context, the effects might have been weaker.

Finally, the size of the reinstatement effects may be due to aspects of the video contexts that were used. It may be that because the video contexts were perceptually rich (Murnane et al., 1999), they provided a good mnemonic basis for encoding. The participants may have intentionally encoded contexts, rather than processing them as incidental backgrounds. Furthermore, these multimodal action scenes may have encouraged the participants to encode context–word ensembles (e.g., Murnane et al., 1999), making context reinstatement particularly effective for mnemonic enhancement (e.g., Bower, 1970). Finally, the encoding time given for even the briefest video contexts was 5 sec—more than enough time for complete context encoding (Malmberg & Shiffrin, 2005).

Several factors other than context reinstatement were ruled out as possible causes of these effects. Variations in encoding and test instructions did not appear to modulate the effects observed. Whether the participants were told to think of a relation between each list word and its background scene at encoding (Experiment 2) or told nothing about encoding such associations (Experiment 1), large reinstatement effects were seen. Furthermore, the possibility that the participants might not have tried to recall words associated with nonreinstated scenes is doubtful, because the extra instruction, given in Experiment 2 (but not in Experiment 1), admonished the participants to recall all of the presented words, not just the ones corresponding to the scenes shown at test; this admonishment did nothing to decrease the effect size that was observed. Whether context cues were drawn from the first or second half of the presentation list (Experiment 1) or from serial positions spanning the entire list (Experiment 2), very large effect sizes were found.

A control condition (with no scenes reinstated at test) was implemented in Experiment 2 to examine the contribution of cue-biased output interference (i.e., extra output interference due to initial recall of contextually cued words) to the observed reinstatement effects. In lieu of test context cues, it was assumed that output interference in this condition was not biased by either A or B scenes. The analyses indicated, however, that cue-biased output interference contributed little, if anything, to the robust context reinstatement effects that we observed. The participants in the no-contexts-reinstated control condition recalled only marginally more noncued words than did the participants who had half of the video contexts reinstated at test. Furthermore, the small advantage of uncued words in the no-contexts-reinstated condition might not have been attributable to decreased output interference in that condition, but rather to the possibility that mental reinstatement of nonprovided video contexts might have been easier for the participants who saw no videos at test and more difficult for those who were watching videos during the test. Furthermore, the level of recall in the no-contexts-reinstated condition was far less than recall of words corresponding to reinstated scenes. Whatever explanation is best for this small effect, it clearly does not account for the large context reinstatement effects observed here.

The participants in both of our reported experiments were asked to pay attention to the video contexts and the target words, and in Experiment 2, they were asked at encoding to think of relations between target words and video contexts. These instructions were given to maximize the likelihood that video contexts would serve as effective recall cues at test. The present experiments, therefore, differ from studies in which contexts were presented incidentally. Thus, it remains unknown whether video contexts cue memory as effectively if they are incidentally encoded.

Some may see the present method as similar to the method of paired associates, another highly controlled laboratory method for testing recall. There are, however, important differences between our video context method and the method of paired associates. One difference is that the participant in a paired associates study must recall the correct response term for a given stimulus term, whereas in the present method, and in the cued recall methods used by Tulving and Osler (1968), Thomson and Tulving (1970), and others, the participants were instructed to recall any and all target words and were not asked to link recall targets with appropriate stimuli. A second important difference is that our video context cues were quite effective even when many (e.g., 15) target words were associated with a single context cue; in paired associates, there is a one-to-one mapping of stimulus terms with response terms. Although our method differs from paired associates learning, it is quite similar to the cued recall method used in encoding specificity experiments (e.g., Thomson & Tulving, 1970; Tulving & Osler, 1968). Whereas in those experiments, words were typically used as cues, in the present study we used video recordings of real environments as cues. Our intention was to develop a method that links investigations of ecologically realistic environmental cues with methods that involve more highly controlled laboratory stimuli.

We classify prospective questions that can be addressed with the movie context method reported here as methodological and theoretical types of investigations. Methodological questions must examine the features of the video contexts, such as their distinctiveness, familiarity, continuity, integrality, emotionality, multimodality, or content and how those features contribute to context dependence. Functional relationships, such as numbers, types, or relationships of target events with video contexts, temporal factors, or effects of repetitions of contexts can be examined with this method. Another set of methodological considerations includes questions about context dependency on various types of memory tests. Our use of recall in the present experiments was based on the observation that free recall is particularly sensitive to context-dependent memory effects (e.g., S. M. Smith et al., 1978; S. M. Smith & Vela, 2001). It remains to be seen whether other tests, such as recognition or indirect memory tests (e.g., word fragment completion, repetition priming), are also sensitive to manipulations of video contexts. Finally, it is not clear how well incidentally encoded global contexts cue memory; the present method, in which we used intentionally encoded local contexts, leaves these questions open to investigation.

Theoretical questions that remain untested can also be addressed with these video contexts. The most traditional questions that have been investigated with global contextual manipulations have been contextual cuing of episodic memories (e.g., Dulsky, 1935; S. M. Smith et al., 1978) and interference reduction (e.g., Bilodeau & Schlosberg, 1951; Greenspoon & Ranyard, 1957). Theoretical mechanisms concerned with contextual cuing, such as outshining (reduced effectiveness of context cues due to increased use of noncontextual cues; e.g., S. M. Smith

& Vela, 2001), overshadowing (reduced effectiveness of context cues due to decreased encoding of context; e.g., S. M. Smith, 1988), suppression of context (e.g., S. M. Smith & Vela, 2001), and the effects of context similarity on cuing (e.g., McGeoch, 1942) can be examined with the video-context methodology. In addition, questions about source monitoring can be investigated in paradigms in which memory for contexts associated with events is tested. In source monitoring studies, the use of a small set of differentiable sources, such as a male versus female voice or a small set of colored backgrounds, leads to the problem that when the participants identify an event's source, they could be guessing. Using many differentiable video contexts as sources would provide a way to test source memory by recalling specific contexts, such as a soccer game or a grocery aisle, mitigating the need to take guessing into account. There are also theoretical mechanisms involving temporal/contextual fluctuation (e.g., Glenberg, 1979; Howard & Kahana, 1999, 2002; Mensink & Raaijmakers, 1988) that have implications for such phenomena as spacing of repetitions effects (e.g., Glenberg, 1977, 1979), long-term recency (e.g., Glenberg et al., 1983; Glenberg & Swanson, 1986; Howard, Kahana, & Wingfield, 2006), and false memories (e.g., Kimball, Smith, & Kahana, 2007). The video context methodology provides the means for experimentally manipulating contextual fluctuation.

In conclusion, the video context method reported in the present study produced very large effect sizes. These effects were found in spite of variations in the instructions and in the context cues used. The effects could not be attributed to cue-biased differences in output interference. This method provides opportunities for testing numerous methodological and theoretical questions concerned with context-dependent memory.

AUTHOR NOTE

Correspondence concerning this article should be addressed to S. M. Smith, Department of Psychology, Texas A&M University, College Station, TX 77843 (e-mail: stevesmith@tamu.edu).

REFERENCES

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, **79**, 97-123.

Balch, W. R., Bowman, K., & Mohler, L. A. (1992). Music-dependent memory in immediate and delayed word recall. *Memory & Cognition*, **20**, 21-28.

Bilodeau, I. M., & Schlosberg, H. (1951). Similarity in stimulating conditions as a variable in retroactive inhibition. *Journal of Experimental Psychology*, **41**, 199-204.

Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning & Verbal Behavior*, **9**, 529-533.

Calkins, M. W. (1896). Association: An essay analytic and experimental. *Psychological Review Monograph Supplements*, **2**.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155-159.

Cousins, R., & Hanley, R. (1996). The effect of environmental context on recall and category clustering scores following relational and individual item processing: A test of the outshining hypothesis. *Memory*, **4**, 79-90.

Dulsky, S. G. (1935). The effect of a change of background on recall and relearning. *Journal of Experimental Psychology*, **18**, 725-740.

Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 764-770.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, **89**, 627-661.

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, **13**, 333-345.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Glenberg, A. M. (1977). Influences of retrieval processes on the spacing effect in free recall. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 282-294.

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. (1997). What memory is for. *Behavioral & Brain Sciences*, **20**, 1-55.

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, & Cognition*, **9**, 231-255.

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 3-15.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: Land and underwater. *British Journal of Psychology*, **66**, 325-331.

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology*, **71**, 99-104.

Greenspoon, J., & Ranyard, R. (1957). Stimulus conditions and retroactive inhibition. *Journal of Experimental Psychology*, **53**, 55-59.

Herz, R. S. (1997). The effects of cue distinctiveness on odor-based context-dependent memory. *Memory & Cognition*, **25**, 375-380.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 923-941.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, **46**, 269-299.

Howard, M. W., Kahana, M. J., & Wingfield, A. (2006). Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin & Review*, **13**, 439-445.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, **96**, 208-233.

Isarida, T., & Isarida, T. K. (2004). Effects of environmental context manipulated by the combination of place and task on free recall. *Memory*, **12**, 376-384.

Isarida, T., & Isarida, T. K. (2007). Environmental context effects of background color in free recall. *Memory & Cognition*, **35**, 1620-1629.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 21-38.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, **114**, 954-993.

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 153-161.

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 322-336.

McGeoch, J. A. (1942). *The psychology of human learning*. New York: Longmans.

Mensink, G., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, **95**, 434-455.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, **100**, 183-203.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 882-894.

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition*, **22**, 584-590.

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 158-172.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, **128**, 403-415.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, **88**, 93-134.

Rutherford, A. (2004). Environmental context-dependent recognition memory effects: An examination of ICE model and cue-overload hypotheses. *Quarterly Journal of Experimental Psychology*, **57A**, 107-127.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1064-1072.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.

Smith, D. G., Standing, L., & de Man, A. (1992). Verbal memory elicited by ambient odor. *Perceptual & Motor Skills*, **74**, 339-343.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning & 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. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory & Cognition*, **12**, 477-482.

Smith, S. M. (1985). Environmental context and recognition memory reconsidered. *Bulletin of the Psychonomic Society*, **23**, 173-176.

Smith, S. M. (1986). Environmental context-dependent recognition memory using a short-term memory task for input. *Memory & Cognition*, **14**, 347-354.

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13-34). New York: Wiley.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, **6**, 342-353.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and a meta-analysis. *Psychonomic Bulletin & Review*, **8**, 203-220.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, **86**, 255-262.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, **77**, 593-601.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning & Memory*, **1**, 442-452.

Weiss, W., & Margolius, G. (1954). The effect of context stimuli on learning and retention. *Journal of Experimental Psychology*, **48**, 318-322.

NOTES

1. The outshining hypothesis states that "the use at test of noncontextual cues, such as cues that make use of interitem associations, can diminish the subject's use of ambient contextual cues, thereby decreasing the influence of environmental manipulations" (S. M. Smith & Vela, 2001, p. 206).

2. Recall was compared for reinstated and nonreinstated targets three times, once for each context load condition. The family of six pairwise comparisons included comparisons among all three test conditions on recall of Counterbalancing Condition A words and among all three conditions for recall of Counterbalancing Condition B words.

(Manuscript received March 26, 2009;
revision accepted for publication June 23, 2009.)