APPLIED COGNITIVE PSYCHOLOGY, VOL. 6, 125–139 (1992)

Environmental Context-dependent Eyewitness Recognition

STEVEN M. SMITH* and EDWARD VELA†

**Texas A&M University; and †California State University Chico, U.S.A.*

SUMMARY

The present study helped resolve the apparent conflict between many laboratory list-learning studies, which have not found environmental context-dependent recognition memory, and staged field studies (e.g. Malpass and Devine, 1981), whose results with 'guided memory' techniques suggest that eyewitness face recognition should depend upon environmental context reinstatement. It was found in two different experiments that, relative to testing in a new place, returning participants to the environment where a live staged event had occurred improved performance on identification of a confederate's face (i.e., hit rate). Although physical reinstatement improved identification performance in Experiment 1, mental reinstatement instructions to subjects tested in a new environment did not improve identification performance over an uninstructed group. The environmental reinstatement effect did not interact with test delay or confederate. In Experiment 2 it was found that environmental reinstatement improved accuracy (hit rate and foil identification rate) when the correct target was present in the test line-up, and that false identifications were not significantly affected by contextual manipulations when the correct target was absent from the line-up. The results provide an empirical basis for the hypothesis that returning to the scene of an event improves eyewitness face recognition.

Does returning to the scene of an event enhance a witness' memory? The experimental literature related to this issue does not offer a clear answer about the mnemonic effects of returning to the scene of an event. List-learning studies of contextual reinstatement disagree, in general, with what might be predicted from the results of staged field studies concerning the effect of physical reinstatement of an environment on recognition memory. Studies using lists of words or other verbal materials (e.g. Smith, Glenberg, and Bjork, 1978; Godden and Baddeley, 1980) have typically found no effect of environmental manipulations on recognition memory. In apparent contrast, staged field studies have shown face recognition improvements from techniques which involve a composite of mental reinstatement strategies (e.g. Malpass and Devine, 1981; Krafka and Penrod, 1985), although these field studies have not directly tested environmental resintatement effects on eyewitness recognition. The present study was designed to provide empirical evidence relevant to the question of environmental context-dependent recognition memory, using live events staged and tested in live environmental contexts.

The efficacy of context reinstatement procedures for improving eyewitness memory accuracy is an important issue for criminal investigations. Investigative leads for finding criminals and accurate identification of suspects are commonly believed to

0888–4080/92/020125–15/$07.50

Received 4 August 1989

be enhanced by returning eyewitnesses to the scene of the witnessed crime. Some investigators routinely bring victims of crimes to the scene of the events in question, claiming thereby to evoke important memories. Such putative enhancement of memory might affect recall by improving access to otherwise inaccessible memories. Identification, however, may not be similarly affected if it depends more upon the familiarity of the target than upon target access. Would someone who had been seen by an eyewitness be better identified if viewed in the appropriate setting rather than at the police station? The usefulness of context cues and eyewitness performance in criminal investigations, as well as other issues relating to the practicality of context-dependent memory findings, has been discussed more fully by Davies and Thomson (1988).

Although previous investigations have not examined the memory effects of returning a witness to the scene of the target event, there are some studies which are relevant to the issue. Malpass and Devine (1981) studied eyewitness recognition of a person who committed a staged vandalism. On a later recognition test some subjects were given 'guided memory' instructions prior to making recognition judgements, and some were not. The guided memory procedure consisted of a series of instructions to image the target environment (e.g. Smith, 1979), and to try to re-create the feelings and thoughts associated with the event. The guided memory procedure also provided subjects with detailed information about the event and its circumstances, such as where the instructor was sitting and what he was doing, which door the vandal entered from, and what the vandal did to make the instructor yell at him. All subjects were tested away from the classroom where the event occurred. Eyewitnesses given the guided memory procedure were more accurate in their recognition judgements than the control subjects.

A similar study reported by Krafka and Penrod (1985) also examined contextual reinstatement and eyewitness recognition. In their study a customer–confederate entered a small store and purchased a small article with a $10 traveller's cheque. Clerks' recognition memory of the confederate was tested by another confederate posing as a 'low intern'. Before asking clerks to choose the correct photo from a line-up of six photos, the 'intern' gave contextual resinstatement instructions to half of the subjects. Reinstatement involved giving subjects (clerks) a copy of a signed cheque used by the suspect, a copy of an identification card used by the customer–confederate (the card had no photo of the confederate), and instructions to try to image the previous customer transaction. All subjects were tested in the context in which the incident had been staged. The reinstatement procedure was effective for improving eyewitness recognition, as in the Malpass and Devine (1981) study. It is not clear from these studies whether the mental reinstatement procedures facilitated recognition memory because they induced subjects to image the event context, or because the procedures provided subjects with specific event information, such as the cheque and ID (Krafka and Penrod, 1985) or the detailed event information (Malpass and Devine).

Sanders (1984), who studied effects of context cues on eyewitness identification responses, found that 'reinstating the physical environment in which the crime was witnessed had no discernible effects' (p. 386). It must be noted, however, that the criminal incidents and line-ups Sanders used were not live, but were black-and-white videotaped scenes lasting for brief durations. The fact that the videotaped materials rather than live events were used limits conclusions about the effectiveness of return-

ing to a scene to revive memories of live events. For example, in the live event participants move themselves into the original and test environments over a period of time, getting physically settled and 'induced' into a situational set in a way similar to the induction procedures used in state- and mood-dependent studies. When a videotape is viewed for a period lasting only seconds, however, it is not clear what contextual elements are 'setting' the participant; is it the videotaped environment, the setting in which the videotape is shown, or some combination? It is not clear, therefore, that the Sanders (1984) study can directly address the effects of returning to the physical setting of an event.

Cutler, Penrod, and Martens (1987) found benefits in identification performance as a function of their 'cognitive interview'. Geiselman, Fisher, MacKinnon, and Holland (1985), who designed the 'cognitive interview' used by Cutler *et al.*, found that it improved eyewitness recall relative to a standard interview. Among other mnemonic instructions the procedure involved directing participants to imagine the context of the event in question, showing them snapshots of the environment and victim, and having them read their own description of the robber, which had been written immediately after the event. Reinstatement effects on identification occurred in Cutler *et al.*'s study only for long (2 weeks) rather than short (2 days) intervals, for subjects who did not pre-view mugshots prior to the line-up, and for targets who appeared disguised in the line-up. As with Sanders' (1984) study, the Cutler *et al.* study used videotaped events rather than live ones, thus limiting the degree to which the study is relevant to the question of returning to the scene of a live event.

In the present experiments unexpected events were staged before numerous eyewitnesses. The participant's ability to identify the confederate who staged the event was tested using a photo line-up. The test was given where the events had occurred, or in a new place. It was predicted that reinstatement of the appropriate environmental context would improve identification performance.

EXPERIMENT 1

In Experiment 1 an event was staged in class by a confederate. Volunteers from the class who were naive as to the purpose of the experiment were tested either in their classroom, where the event had been staged, or in a different room in a different building. Half of those tested in the different environment were first given instructions to image their classroom and list several objects which were in that room. No details about the event or the environment were provided for the subjects, as in the Malpass and Devine (1981) and Krafka and Penrod (1985) studies. All subjects were shown 10 photographs of faces three times apiece, and were asked to choose, on the third viewing, which of the 10 faces was the confederate from the staged event. The correct target was present in all conditions of Experiment 1. Subjects gave confidence rating to their identification judgement.

Method

Participants

The 212 Texas A&M University students who served as volunteers in Experiment 1 fulfilled part of a course requirement for the introductory psychology classes. Volunteer sign-up sheets were posted to recruit subjects for the nine test sessions

of the experiment. Participants were free to sign up for any (one) session they liked. The numbers of subjects who signed up for and attended each of the test sessions are shown in Table 1. Some of the Participants in each of the nine groups had seen one confederate (Bob), and the rest had seen the other confederate (Jed). Participants were told beforehand only that the experiment would test some cognitive abilities.

Design, procedure, and materials

In naturally occurring cases in which eyewitness memory is of interest, eyewitnesses are usually not prepared or forewarned of the events in question. Such prepration might induce a set or bias which could detract from the incidental nature of the witnessed events. Therefore, to avoid such bias and reactivity in the present study, participants did not attend prearranged experimental sessions for the target event. Instead, the event was staged in a more 'naturally occurring' setting, the students' introductory psychology class.

An incident was staged at the beginning of each of eight introductory psychology classes (there were different students in each class). All classes met in the same lecture hall. A confederate entered the classroom and announced that it was a (fictitious) person's birthday. To increase the generality of the results two different male confederates (Bob and Jed) were used; each staged the incident in half of the introductory psychology classes. Both confederates were dressed casually but wore a white laboratory coat over their regular clothing. They also carried a 'birthday gift'; a bottle wrapped in aluminium foil that held a plastic rod with a red and silver foil balloon attached to the end.

When the confederate entered the front of their classroom he asked for everyone's attention. He then asked if a fictitious student was present because he had a birthday gift to deliver to her. When no-one responded, he left, saying, 'I guess she must be absent today'. From start to finish the incident lasted approximately 1 minute or less.

A day, 2 days, or 1 week later, participants were tested for recognition of the confederate. The test took place either in the classroom where the incident took place, or in a new room in a different building. The classroom in which the event was staged (i.e. the SC test room) was a large, open, carpeted, well-lit classroom with 250 permanent seats, a stage with a podium, chalkboards, and viewing screens. The SC room was in the Harrington Education Center, a modern classroom building on the Texas A&M University campus. The DC test room was an old classroom with 75 movable wooden desks, a tile floor, windows on the side, and a blackboard at the front of the room. The DC test room was located in the Academic Building, an old historically preserved office/classroom building on the Texas A&M University campus. The participants expressed no suspicion that the staged incident was part of an experiment and were naive about the purpose of the present experiment.

At the test participants viewed 10 color mugshots (front view of shoulders and head) of males presented on slides. At each test session were some subjects who had seen one confederate and some who had seen the other confederate. Participants were strictly instructed to stay quiet during the viewings of the photo line-up so that they would not influence the responses of other subjects. One confederate's photo was in position 3 of the photospread, and the other was in position 8. Each slide was presented for 12 seconds with a 3-second inter-trial interval.

One-third of the groups were tested in their regular classroom (same context,

or SC). They were asked to sit where they normally sit in class. The remaining two-thirds were tested in a new room (different context, or DC). Before the second viewing of the slides one of the DC groups (DCI) was asked to image and mentally reinstate their classroom to facilitate recognition of the photos. The DCI subjects were asked to image themselves in their regular seats, to image the others who had been around them, and to think about how they felt at the time of the incident. To help ensure that some attempt at mental resinstatement was initiated, subjects were asked to list three things found in the room. A highly similar procedure used by Smith (1979, 1984) has been found to substantially increase recall of word lists for subjects tested in DC conditions.

The photospread was shown to subjects three times in the same order, although participants did not write down their responses until their third viewing of the slides. The purpose of having three viewings of the slides was to ensure that subjects could compare all of the photos and be more certain of their responses. Subjects were not reminded of the staged incident before the first viewing; they were simply asked to name any of the photospread people they knew (none were identified by any participants on their first viewing). The purpose of this procedure was to familiarize subjects with all of the photospread faces, and to determine whether or not faces in the photospread were known to subjects outside of the experiment (none were). After this familiarization viewing subjects were reminded of the classroom delivery incident, and they were told that a slide of the person who had come into their class was in the photospread. Their task was to identify the confederate's picture in the photospread. They were told that they would get two more viewings of the photos, and that they should not write down their responses until the final viewing. During the second and final viewings subjects were to indicate for each photo how certain they were that the photo was or was not the confederate. A response of +3 indicated they were certain that it was the confederate, +2 moderately certain, and +1 indicated that they were only minimally certain. A −1 indicated that they were minimally certain that the photo was not the confederate, −2 moderately certain, and −3 indicated that they were certain that the photo was not the confederate. Unexpectedly, a number of subjects gave positive confidence judgements for more than one photo, some chose exactly one (as we had expected), and some gave no positive confidence judgements for any photo (the numbers of subjects of these three types are listed in Table 3). In the analyses a hit was counted as a positive confidence reported for the correct confederate, and a foil identification (e.g. Wells and Turtle, 1986) was a positive confidence given for any of the nine incorrect photos. No subject could get more than one hit, but it was possible to get more than one foil identification per subject. Foil rejects were counted as negative confidence judgements for any of the nine incorrect photos.

RESULTS AND DISCUSSION

Accuracy

Two different measures related to accuracy, correct identifications (hits) and foil identifications, were used to compute separate $3 \times 3 \times 2$ (context × interval × confederate) ANOVAs for all subjects.

The ANOVA computed for hits found a significant effect of context,

$F(2,195) = 3.28$, $p < .05$, MSe = .237. Subjects had the highest hit rate in the SC condition (mean = 66.2 per cent) as compared with the DC group (mean = 50.0 per cent) and the DCI group (mean = 47.3 per cent). Newman–Keuls pairwise comparisons indicated that the DC and DCI hit rates did not significantly differ from each other ($\alpha = .05$; for $r = 2$ critical difference = 16.6 per cent). That this environmental context-dependent recognition memory effect is both significant and sizeable is at odds with most laboratory findings (e.g. Smith *et al.*, 1978; Godden and Baddeley, 1980; Jacoby, 1983; Eich, 1985), which have used not single faces, but rather lists of words as targets. On the other hand, the finding of environmental context-dependent eyewitness recognition of a face is predictable from the results of previous eyewitness memory studies (e.g. Malpass and Devine, 1981; Krafka and Penrod, 1985; Cutler *et al.*, 1987) which found beneficial effects on recognition using a variety of reinstatement procedures.

There was also an effect of confederate $F(1,195) = 10.35$, $p < .01$, MSe = .237, with Bob remembered better than Jed. Although hits decreased numerically as a function of increased retention interval, the effect of interval was not significant, $F(2,195) = 2.05$, $p = .13$, MSe = .237. Forgetting over a longer retention interval than 1 week may have proven more statistically reliable. There were no interactions among any of the variables.

The ANOVA for foil identifications found a non-significant effect of context, $F(2,195) = 2.55$, $p = .08$, MSe = 1.102. The fewest foil identifications were made in the SC condition (Table 1). This suggests that the increase in hits for the SC condition occurred not from a response bias to identify any faces from the line-up, because such a bias would have concomitantly increased foil identifications. Although the effect was not significant, the trend in the results suggests that SC subjects were more able than DC subjects to distinguish the target from foils in the line-up.

The effect of confederate on foil identifications was not significant, $F(1,195) = 2.99$, $p = .085$, MSe = 1.102. There was no effect of interval, $F(2,195) = .53$. No interactions approached significance.

The previous analyses indicate that recognition of a face improved as a function of environmental reinstatement, regardless of the confederate or the retention interval. Imaginal reinstatement instructions had no obvious beneficial effect on recognition for subjects who did not get physical environmental reinstatement. A potential problem with this conclusion is that the same context test room and the different context test room were different, since the witnessed events were all staged in the regular lecture room. A review of the experimental literature on context-dependent memory, however, indicates that 43 experiments have found significant environmental context effects and examined test room counterbalancing effects; of those 43, 42 failed to find any effect of room counterbalancings. Therefore, it was considered very unlikely that the test room (the classroom in the Harrington Classroom building) was somehow an especially good environment for remembering things, accounting for the present findings.

Although there was exactly one correct target for each subject, subjects were not explicitly instructed to give a positive response to exactly one photo. They were simply asked to rate their confidence in their recognition memory judgement for each face. Therefore, although we had expected that almost everyone would select a single target, some subjects positively identified zero faces, some chose one face, and some chose more than one.

Table 1. Hit rate, foil identification rate, standardized confidence, and *n* as a function of context and interval in Experiment 1

	Context		
	SC	DC	DCI
Interval			
One Day			
Hit rate	.76	.65	.45
Foil identification rate	.06	.07	.19
Standardized confidence	2.07	1.55	2.51
n	25	17	11
Two Days			
Hit rate	.73	.47	.52
Foil identification rate	.06	.08	.05
Standardized confidence	2.03	1.03	1.27
n	11	19	26
One week			
Hit rate	.57	.44	.44
Foil identification rate	.06	.10	.12
Standardized confidence	1.52	1.01	.78
n	35	32	36

Note: Hit rate refers to the probability of a correct identification; foil identification rate refers to the proportion of positive identifications for any of the nine incorrect line-up faces (maximum = 9). Standardized confidence is the *z*-score of a subject's confidence for the correct photo calculated from the subject's 10 confidence ratings.

The numbers of subjects in each group for SC, DC, and DCI conditions are shown in Table 2. Although the proportions of SC and DC subjects in each group are about equal, there were proportionately more subjects in the DCI group who chose zero faces, and more who chose multiple faces than in the SC or DC groups.

Two separate 2 × 3 (number identified × context) ANOVAs were computed using

Table 2. Hit rate, foil identification rate, and *n* as a function of context and number identified in Experiment 1

	Context		
	SC	DC	DCI
Number Identified			
One			
Hit rate	.78	.60	.71
Foil identification rate	.02	.04	.03
n	45	42	28
Multiple			
Hit rate	.80	.45	.54
Foil identification rate	.21	.19	.25
n	15	20	28
Zero			
n	11	6	17

Note: Hit rate refers to the probability of a correct identification; foil identification rate refers to the probability of positive identifications of the nine incorrect line-up faces.

hits and foil identifications as dependent measures. Number identified was either one or multiple faces. Because those who identified zero faces had zero hits and zero foil identifications, there were not included in this analysis. There were significantly more hits overall in the SC conditions than in the DC or DCI conditions, $F(2,172) = 3.98$, $p < .05$, MSe = .22. The number identified was not predictive of hit rate, $F(1,172) = 1.76$, $p = .19$, MSe = .22 (Table 2). Context and number identified did not interact, $F(2,172) < 1.0$. This lack of an interaction may also be characterized as evidence that environmental reinstatement enhanced recognition for participants who chose one face from the line-up as well as for those who identified more than one face.

As expected, more foil identifications were made by the multiple group than the one group, $F(1,172) = 199.79$, $p < .05$, MSe = .55. There was, however no effect of context, $F(2,172) = 1.42$, $p = .25$, nor was there an interaction of context and number identified, $F(2,172) < 1.0$.

Confidence

Confidence ratings, which ranged from −3 to +3 (with no zero choices allowed), were first translated to a scale of 1 (for −3) to 6 (for +3), since the confidence rating scales contained six equally spaced choice points. The distribution of 10 confidence judgements for each subject was used to compute a *z*-score for the subject's judged confidence of the correct target photo. This metric, which we will refer to as 'standardized confidence', provided another measure of how well a subject could distinguish the correct target face from the other faces in the line-up.

A 3 × 3 × 2 (context × interval × confederate) ANOVA was computed using standardized confidence as a dependent measure. This analysis found a significant effect of context, $F(2,195) = 3.57$, $p < .05$, MSe = 2.00, with the SC groups again performing better than the DC or DCI groups (Table 1). This again agrees with analyses of hit rate.

Although not found to be statistically reliable with other measures of recognition accuracy, a significant effect of interval on standardized confidence was found, $F(2,195) = 8.19$, $p < .05$, MSe = 2.00. Standardized confidence decreased from 1 day to 2 days to 1 week.

There was a marginal but non-significant effect of confederate, $F(1,195 = 2.73$, $p = .07$, MSe = 2.00. No significant interactions were found.

Calibration of confidence and accuracy

Accuracy, as measured independently by correct identifications (i.e. hits) and foil rejections, was correlated with confidence judgements across all subjects, and again for each context condition (Table 3). These types of correlations between accuracy and confidence refer to what may be called calibration. Better calibration occurs the more predictive accuracy is from judged confidence.

A correct identification was scored as 1, and a foil identification was scored as a 0. Confidence, in this analysis, was either 4 (not certain), 5 (fairly certain), or 6 (very certain). For correct identifications (hits) × confidence there was a negligible correlation, $r = .104$. The correlation was highest in the DCI condition, although calibration was not significant for any context group (Table 3).

Table 3. Correlations of confidence × accuracy as a function of context in Experiment 1

	Context			
Measure	SC	DC	DCI	All
Hits	.078	.059	.128	.104
Foil rejections	−.061	−.011	.307**	.156*

Note: * indicates a significant correlation at $p < .05$; ** indicates a significant correlation at $p < .01$. All tests were two-tailed t-tests.

The correlation between foil rejections (foil rejection = 1, miss = 0) and confidence judgements (1 = not certain, 2 = fairly certain, 3 = very certain) was significant, $r = .156$ (Table 3), $t(211) = 2.29$, $p < .05$. Again, the correlation was highest in the DCI condition ($r = .307$), which was significant, $t(72) = 2.74$, $p < .01$. Calibration was not significant in the SC or DC conditions. Interestingly, although imagery reinstatement instructions failed to benefit accuracy, the DCI group had the only significantly calibrated recognition memories.

EXPERIMENT 2

Although the results of Experiment 1 demonstrate that reinstatement of the original environmental context improves eyewitness identification accuracy when the correct target appears in the line-up, it does not address the question of accuracy in cases in which the correct target is absent from the line-up. Correct rejection of lures in the line-up is critically important whether or not the correct target appears in a line-up. Since it cannot be known in naturally occurring cases of eyewitness memory whether the target is actually in the line-up, it is necessary to assess the effects of contextual reinstatement in a target-absent condition. Effects of contextual reinstatement on eyewitness recognition were examined for both target-present and target-absent line-up conditions in Experiment 2.

Method

Participants

The 83 Texas A&M University students who served as volunteers for Experiment 2 fulfilled part of an introductory psychology course requirement by participating. All had attended a mass testing session of introductory psychology students during the first week of classes. Volunteer sign-up sheets were posted to recruit participants for the experiment; 52 signed up for the sessions which became the target-absent conditions, with 31 in the target-present conditions. The numbers of participants enrolling in each of the four treatment groups are shown in Table 4.

Design and procedure

During the first week of spring semester classes about 1000 introductory psychology students attended a 'pre-screening' session in which they filled out a succession of forms and questionnaires related to psychological research projects. The pre-screening occurred in a large lecture hall in Heldenfels Hall, a building different from

those used in Experiment 1. Five minutes into the pre-screening session a confederate, carrying a pizza box in hand, walked into the front of the auditorium and loudly stated that he was delivering a pizza to 'Kendra Arista'. After a moment most of the students laughed, and the confederate looked briefly across the audience before leaving through the front of the auditorium. No mention of the incident was ever made prior to the eyewitness identification test, done 4 days after the staged pizza delivery.

The test session was conducted either in the pre-screening room in Heldenfels Classroom Building (SC), or in a different classroom in the Harrington Education Center (DC). As previously noted, it was remotely possible that the same context test room was an especially good environment for remembering things, thus causing the context-dependent findings in Experiment 1. To generalize the results to other test rooms, the same context test room from Experiment 1 was used as the different context test room in Experiment 2.

At the test session participants were shown six faces via slide projectors, and they were asked to identify the person who had staged the pizza delivery at the pre-testing session earlier in the week. The six photos were shown simultaneously for a duration of 1 minute. Subjects were informed that the confederate might not be in any of the photos. They were to choose either one face in the photospread or to respond that none was the correct target. The six faces were labelled A through F, and 'N' the indicated response for 'none of the above'.

After the first viewing of the photospread all participants were given instructions to try to mentally reinstate the pre-testing session. They were asked to remember how they got to the pre-testing session, what they saw there, and what happened there. As a manipulation check, participants were asked to list three things they remembered seeing at the pre-testing session. Following this mental reinstatement procedure a second viewing of the photospread was given. Subjects were asked to use their memory of the pre-testing session to help them identify the pizza delivery man (the confederate). They saw the same line-up and were given the same instructions they had for their first viewing, to identify the correct target, or to indicate that the target was not among the six photos. The correct target was included in the photos for the target-present groups, but not in the target-absent groups.

Line-ups

The photo of the correct confederate was placed in position 2 of the six line-up photos for the target-present condition. A similar-looking lure was placed in position 2 for the target-absent condition. The confederate and the lures were Caucasian males in their 20s with light brown hair and no facial hair. Three faces were pictured on each slide. Two such slides were displayed simultaneously on slide projectors to allow subjects the opportunity to compare the faces, should they wish to do so.

Results and discussion

Selection of the photo in position 2 was a correct identification (hit) in the target-present condition, and in the target-absent condition it was termed a false identification (after Wells and Turtle, 1986). A $2 \times 2 \times 2$ (context $\times$ present/absent $\times$ line-up) ANOVA was computed using responses for position 2 as the dependent measure.

Context was SC or DC, present/absent was target-present or target-absent, and line-up was either first viewing or second viewing. The ANOVA found a significant effect of present/absent, $F(1,79) = 12.73$, $p < .05$, MSe = .18, indicating that the correct target was identified more often than the lure used in the same position in the target-absent condition (Table 4). This ANOVA also found a significant effect of context, $F(1,79) = 6.53$, $p < .05$, MSe = .18. This result indicates that the position 2 photo was more likely to be selected in the SC condition that in the DC condition. Planned one-tailed comparisons were computed contrasting performance (hits) in the SC/target-present condition with that of the DC/target-present condition for both first and second viewings of the line-up. Although the trend for the first viewing favored the SC condition, the effect was not significant, $t(29) = 1.16$, $p > .05$. The finding of greater hits in the SC condition was significant only for the second viewing, $t(29) = 2.00$, $p < .05$. Planned one-tailed comparison were also computed contrasting performance (false identifications) in the SC/target-absent condition with that of the DC/target-absent condition for both first and second viewings of the line-up. Neither comparison showed a significant effect. No other main effects of interactions were significant.

Table 4. Hit rate, foil identification rate, and false identification rate as a function of context and line-up conditions in Experiment 2

	Context	
Line-up	SC	DC
Target present		
First viewing		
Hit rate	.44	.27
Foil identification rate	.11	.15
Second viewing		
Hit rate	.44	.13
Foil identification rate	.11	.17
n	16	15
Target absent		
First viewing		
False identification rate	.17	.00
Foil identification rate	.16	.19
Second viewing		
False identifications	.09	.04
Foil identification rate	.15	.17
n	23	29

Note: The target was in position 2 in the target-present condition; false identification rate was the proportion of the subjects who incorrectly identified the position 2 lure; foil identification rate was the proportion of subjects in a group who selected any lures *exclusive* of those in position 2 (maximum = 5).

The results support the findings of Experiment 1 in showing improved hit rates in the SC condition relative to the DC condition, although the effect was significant only after imagined reinstatement and a second viewing of the photospread. This pattern of results extends the findings of Experiment 1 by showing that false identifications for the lure in position 2 of the target-absent condition were not increased

by testing in SC rather that DC conditions, a finding which held for both the first and second viewings of the photospread.

Another $2 \times 2 \times 2$ (context × present/absent × line-up) ANOVA was computed using frequency of foil identifications (positive identification or lures not in position 2 of the line-up) as the dependent measure. Context was SC or DC, present/absent was target-present or target-absent, and line-up was either first or second viewing. This ANOVA found a significant effect of context, $F(1,79) = 4.56$, $p < .05$, MSe = .27. In Table 4 it can be seen that there were fewer foil identifications in the SC condition than in the DC condition. No other main effects or interactions were significant.

The effect of mental contextual reinstatement was not tested in Experiment 2 because the effects of the procedure were confounded with first vs. second viewing of the photospread. It is clear from Table 4, however, that a second viewing of the photospread following imagined reinstatement did not improve hits, foil identifications, or false identification scores.

GENERAL DISCUSSION

Ellis (1984), in a chapter on the topic of practical aspects of face memory, stated: 'There appears to be an absence of critical experimental evidence on the question of context and identification accuracy. What is required is an investigation of face recognition when the context changes totally, compared with performance when no alterations in context occur' (p. 33). We believe that the present study makes a clear statement concerning the efficacy of context reinstatement techniques in facilitating eyewitness face recognition.

The hypothesis most consistently supported by the present results is that physical reinstatement of the environmental context of a live event improves eyewitness identification of a person seen at the event. With the SC conditions in Experiment 1 subjects had the highest hit rates, the highest standardized confidence, and the lowest foil identification rates. Superior performance could be seen for both difficult and easy-to-recognize confederates (as defined by how well identified the two confederates were), and at 1-day, 2-day, and 1-week intervals following the event.

In Experiment 2 the beneficial effect of context reinstatement was significant on the second viewing, after witnesses had been asked to try to sharpen their memories of the staged event using an imagined reinstatement procedure. Foil identifications were also improved (decreased) by testing in the SC environment. Importantly, there was no deleterious effect of context reinstatement on false identification rates in the target-absent conditions in Experiment 2. This indicates that the improvement in identification rates caused by context reinstatement is not accompanied by an increase in false identification rates; i.e., reinstatement appears to improve memory, not merely cause a shift in the witness' decision criterion.

Why did the present study find such robust dependence of recognition on environmental reinstatement when most laboratory tests of environmental context-dependent recognition either fail to show effects (e.g. Smith *et al.*, 1978; Godden and Baddeley, 1980; Jacoby, 1983; Eich, 1985; Fernandez and Glenberg, 1985), or find relatively small effects under unusual circumstances (e.g. Smith, 1986; Canas and Nelson, 1986)? There are several possible hypotheses.

Foremost of these hypotheses, we believe, is the outshining hypothesis, (e.g. Smith,

1986), which states that the effectiveness of a cue is diminished when other cues are used. The reason that environmental context-dependent recognition typically fails to occur on list-learning recognition tests, according to this explanation, is that such tests provide 'other cues', including the test words themselves, and especially associative cues (e.g. other test words). This hypothesis, therefore, predicts that when a recognition test consists of a single target, the target will have no inter-item, intra-list semantic associations which can later be used to aid recognition memory judgements. Whether the single target is a single face, as in the present study, or a single word, this hypothesis predicts an environmental context-dependent recognition memory effect. The results of the present study were predicted by the outshining hypothesis, but do not constitute a critical test of the hypothesis.

Another possible hypothesis is that recognition of faces may be more likely than recognition of words to be context-dependent. One of the only mood-dependent recognition memory findings ever reported (Gage and Safer, 1985) tested recognition of facial expressions. Other eyewitness memory studies which tested face recognition (e.g. Malpass and Devine, 1981; Krafka and Penrod, 1985) have found benefits of various reinstatement instructions (see Davies, 1988, for a review of context effects on face recognition). The present findings of environmental context-dependent face recognition appear to be consistent with this pattern. Again, critical tests remain to be conducted.

Another factor, suggested by anecdotal accounts, relates to the issue of emotional experiences. It may be that environmental reinstatement affects memory in naturalistic settings at least partly by reviving memories of emotional experiences, which, in turn, help redintegrate the original memories of faces and events. The events staged for the present experiment appeared to be met with surprise, disappointment (that the requested student missed out on her delivered gift), and amusement. This may be part of the reason we found environmental context-dependent memory. On the other hand, Davies (1988), in reviewing the literature on face memory and emotional states, has argued that the evidence does not support the hypothesis that face recognition is affected by mood states.

Another possible consideration is that the degree of context change in the present study may have been greater than in many previous context-dependent recognition studies. In the present study the change was not only environmental, but also situational, with the event occurring in class or at a pre-screening session, and with the test in an experiment. As pointed out by Canas and Nelson (1986), typical environmental changes in context-dependent memory studies fail to change the situational context of the experiment. In Canas and Nelson's (1986) study, as well as in our own present study, the DC test situation was altered substantially from the original situation in which the events were staged. Both studies showed context-dependent recognition memory.

Although the environmental reinstatement technique had a robust beneficial effect on eyewitness recognition, it is important to point out that even in the best conditions, numerous errors (both misses and foil identifications) were committed. In Experiment 1, for example, in the SC one day group, the hit rate was 76 per cent, and foil identifications were committed for about 6 per cent of the distractor faces in the line-up. After a week the hit rate in the SC group was only 57 per cent. In Experiment 2 the SC hit rate was only 44 per cent. Reinstatement appears to benefit eyewitness face recognition, but it remains far from perfect.

It is also clear that, even in the best conditions, calibration of subjects' memories was very low. That is, for most groups confidence of recognition judgements were unrelated to their accuracy. This poor calibration is consistent with previous studies. Wells and Murray (1984), in reviewing the literature on eyewitness confidence, concluded: 'Yet, the empirical evidence does not support the idea that eyewitness confidence is a valid measure of eyewitness accuracy under ecologically valid conditions' (pp. 168–169). Our results give further support to this conclusion.

Why was performance in the DCI condition in Experiment 1 not improved over the DC condition? It might be reasoned on the basis of the guided memory type of studies (e.g. Malpass and Devine, 1981; Krafka and Penrod, 1985) that environmental imagery should benefit recognition of a face, yet our results do not support this conclusion. It is important to note that the previous studies not only gave imagery instructions, but also supplied participants with critical information related to the original events. A logical next step in the present line of research would be to further examine the interactive effects of imaged and physical reinstatement techiques. Contextual reinstatement apprears to be an effective method for improving eyewitness recognition, but more extensive research in this area will be needed to improve the effectiveness of these memory-enhancement techniques.

ACKNOWLEDGEMENTS

This research was supported by NIMH grant number 1 R01 MH39977–01 awarded to Steven M. Smith. Thanks are due to Jed Friend, Robert Reinhardt, and John Butemeyer, who served as confederates, and to Tom Ward, Wayne Shebilske, Graham Davies, Pamela Kenealy, and anonymous reviewers for their helpful comments on an earlier version of this manuscript. Correspondence should be addressed to Steven Smith, Psychology Department, Texas A&M University, College Station, Texas 77843, U.S.A.

REFERENCES

Canas, J. J. and Nelson, D. L. (1986). Recognition and environmental context: the effect of testing by phone. *Bulletin of the Psychonomic Society*, **24**, 407–409.

Cutler, B. L., Penrod, S. D. and Martens, T. K. (1987). Improving the reliability of eyewitness identification: putting context into context. *Journal of Applied Psychology*, **72**, 629–637.

Davies, G. (1988). Faces and places: laboratory research on context and face recognition. In G. M. Davies and D. M. Thomson (Eds), *Memory in context: context in memory* (pp. 35–53). Chichester: John Wiley & Sons.

Davies, G. and Thomson, D. H. (1988). Context in context. In G. M. Davies and D. M. Thomson (Eds), *Memory in context: context in memory* (pp. 335–345). Chichester: John Wiley & Sons.

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

Ellis, H. D. (1984). Practical aspects of face memory. *Eyewitness testimony: psychological perspectives* (pp. 12–37). New York: Cambridge University Press.

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

Gage, D. F. and Safer, M. A. (1985). Hemisphere differences in the mood state-dependent effect for recognition of emotional faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **11**, 752–763.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P. and Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, **70**, 401–412.

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

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

Krafka, C. and Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Applied Psychology*, **49**, 58–69.

Malpass, R. S. and Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology*, **66**, 343–350.

Sanders, G. L. (1984). The effects of context cues on eyewitness identification responses. *Journal of Applied Social Psychology*, **14**, 386–397.

Smith S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 460–471.

Smith, S. M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory and Cognition*, **12**, 477–482.

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

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies and D. M. Thomson (Eds), *Memory in context: context in memory* (pp. 13–34). Chichester: John Wiley & Sons.

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

Wells, G. L. and Murray, D. M. (1984). Eyewitness confidence. In G. Wells and E. F. Loftus (Eds), *Eyewitness testimony: psychological perspectives* (pp. 155–170). New York: Cambridge University Press.

Wells, G. L. and Turtle, J. W. (1986). Eyewitness identification: the importance of line-up models. *Psychological Bulletin*, **99**, 320–329.