The ability of familiarity, disruption, and the relative strength of nonenvironmental context cues to explain unreliable environmental-context-dependent memory effects in free recall

A. RUTHERFORD
Keele University, Keele, England

The ability of environmental-context (EC) familiarity, movement disruption, and the relative strength of memory cues to explain unreliable EC-dependent free-recall memory effects was examined in two experiments. Experiment 1 replicated Smith's (1979, Experiment 1) results confirming that familiarity and disruption cannot account for free-recall EC-reinstatement effects. In Experiment 2, a level of processing manipulation varied stimulus item memory cue strengths, and memory was again assessed by free recall. Contrary to Murnane and Phelps's (1995) and Dougal and Rotello's (1999) recognition findings, an EC-reinstatement effect was observed with low, but not high, levels of processing. However, comparisons across the two experiments revealed inconsistencies with the relative cue strength hypothesis. Consequently, a variant of the relative cue strength hypothesis that highlights the role of retrieval processes was proposed to explain the interaction between the levels of processing and the EC-reinstatement effect.

Environmental context (EC) refers to information available in the physical environment in which a stimulus is presented that is incidental to task performance (Bjork & Richardson-Klavehn, 1989). EC-dependent memory effects have a long history (Smith, 1988), and EC is employed in many theoretical accounts of memory (e.g., Eich, 1985; Gillund & Shiffrin, 1984; Hintzman, 1988; Humphreys, Bain, & Pike, 1989; Morton, Hammersley, & Bekerian, 1985; Murdock, 1993). However, a number of reviews (e.g., Bjork & Richardson-Klavehn, 1989; Smith, 1988; Wickens, 1987) have addressed the reputation that EC-dependent memory effects have acquired as unreliable phenomena (e.g., Fernandez & Glenberg, 1985). Since most memory theories employing EC predict EC-dependent memory effects, the inconsistency of these effects poses problems for these theoretical accounts.

Although a whole range of EC-dependent memory effects have been reported (e.g., Smith, 1988), most attempts to explain the inconsistency of EC-dependent memory have focused on the EC-reinstatement effect. In a typical EC-reinstatement experiment, subjects are presented with material in one EC, and then their memory for this material in the same or in a different EC is tested. An EC-reinstatement effect is observed when memory tested in the same EC is superior to memory tested in a different

EC. In the present article, two experiments are reported that examine variables that may influence the reliability of this EC-dependent memory effect. In Experiment 1, the influence of the match between the presentation and test EC is compared with the influence of test-EC familiarity on the free-recall EC-reinstatement effect when the disruption between stimulus presentation and memory test is controlled. The aim of this experiment is to determine if any part of the EC-reinstatement effect is an artifact of disruption and/or test-EC familiarity. In Experiment 2, a level of processing (LOP) manipulation at encoding is employed to examine the influence of stimulus item cue strength on EC-dependent free-recall memory performance. The aim of this experiment is to determine whether greater stimulus item cue strength eliminates the EC-reinstatement effect when individual memory records are retrieved, as in free recall.

Disruption and Familiarity

Bilodeau and Schlosberg (1951) and Greenspoon and Ranyard (1957) found that retroactive interference (RI) was reduced when subjects learned two lists in different rooms. However, Strand (1970) noted that subjects presented with the two lists in different ECs moved from one EC to the other after learning the first list, while subjects in the same-EC condition remained in the one EC throughout the experiment. Therefore, RI reductions may have been due to the disruption caused by moving between rooms, rather than a reduction of the associations between words and the test EC. To test this hypothesis, Strand var-

Correspondence should be addressed to A. Rutherford, Psychology Department, Keele University, Keele, Staffordshire ST5 5BG, England (e-mail: a.rutherford@psy.keele.ac.uk).

ied EC match and disruption between first and second list learning. She found RI differences only between conditions distinguished by disruption and not between those distinguished by EC match. Although Strand observed greater consistency of recall order for the second list when the EC was changed, she concluded that disruption was the real reason for RI reduction when EC was changed.

Godden and Baddeley (1975) also examined the disruption hypothesis. After detecting an EC-reinstatement effect when divers' free recall was tested on land or under water, after presentation on land or underwater, they checked to see if a similar reduction in recall could be obtained simply as a consequence of moving between land and water (and vice versa) in the different-EC conditions. However, no support was found for the disruption hypothesis.

Whereas Godden and Baddeley (1975) carried out a post hoc assessment of the tenability of the disruption hypothesis as an explanation of their EC-reinstatement effect, Smith, Glenberg, and Bjork (1978, Experiment 2) controlled disruption in an experiment that manipulated the match between the presentation and test ECs by disrupting all subjects equally, irrespective of condition. Subjects were presented with a list of words in one EC and then were presented with another word list in another EC. Later, subjects' memory for both lists was tested in one of these ECs. An EC-reinstatement effect was observed; the words remembered best were those that had been presented in the test EC.

In fact, Smith et al. (1978, Experiment 2) not only controlled disruption, but were the first to control the familiarity of the test EC. In these EC-reinstatement experiments, subjects in the same-EC conditions are tested in the presentation EC. Since these subjects spend the duration of the stimulus presentation phase in this EC, they become familiar with what will be their test EC. In contrast, subjects in the different-EC condition are tested in an EC that is different from the EC in which the stimuli were presented, and so this EC will be relatively unfamiliar to these subjects. Since subjects in Smith et al. (Experiment 2) are presented with word lists in both ECs, they are already familiar with whichever EC is used as the test EC.

Smith (1979, Experiment 1) also conducted the first experiment that controlled disruption while comparing the relative contribution of the match between the presentation and test ECs and the familiarity of the test EC on the free-recall EC-reinstatement effect. Subjects in same-EC conditions become familiar with the test EC because they spend the presentation phase of the experiment in this EC. Therefore, Smith employed a different-EC condition in which, prior to the memory test, subjects were familiarized with the test EC by spending the same amount of time in the test EC as they had just spent in the presentation EC. To ensure that these subjects would become familiar with the test EC, they were required to draw sketches of the test EC. In another different-EC condition, the usual lack of familiarity with the test-EC was main-

tained by having subjects enter and draw an "irrelevant" room. Consequently, three experimental conditions were involved. Subjects in the same and familiar (SF) condition memorized words in Room A before moving to a waiting area. Next, they entered Room B and sketched this EC, before returning to the waiting area. Finally, they went back to Room A for the memory test. Subjects in the different and familiar (DF) condition memorized words in Room A, before moving to the waiting area. Next, they entered Room B and sketched this EC, and then returned to the waiting area. Finally, they went back to Room B for the memory test. Subjects in the different and unfamiliar (DU) condition memorized words in Room A and then moved to the waiting area. Next, they entered the "irrelevant" Room C and sketched this EC, and then returned to the waiting area. Finally, they went back to Room B for the memory test. Since all subjects, irrespective of condition, moved to the waiting area between phases, all subjects were disrupted equally and so disruption was controlled.

If the match between the presentation and test EC influences memory performance, as predicted by the EC-dependent memory hypothesis, then free recall in SF conditions should be greater than that in DF and DU conditions. If EC familiarity determines memory performance, then memory performance in SF and DF conditions will be equal, but superior to that observed in DU conditions. Since Smith (1979) detected a significant advantage for free recall in SF conditions over DF conditions, but found no significant difference between DF and DU conditions, he concluded that the EC-reinstatement effect was due to the match between the presentation and test ECs and could not be attributed to EC familiarity or disruption.

Nevertheless, given the reputation for unreliability that EC-dependent memory effects have acquired, it is surprising that the literature contains no other attempts to establish that the EC-reinstatement effect is not an artifact of disruption effects or EC familiarity effects, or disruption and familiarity effects. Smith (1979, Experiment 1) remains the only experiment to have controlled disruption and examined the contribution of the presentation and test EC match and the familiarity of the test EC to the EC-reinstatement effect. Subsequent experimental research has tended to ignore familiarity and disruption variables, believing them to be of no consequence for EC-dependent memory effects. However, the unreliability of EC-dependent memory effects raises the concern that perhaps Smith's studies (Smith, 1979, Experiment 1; Smith et al., 1978, Experiment 2) were misleading and that EC-dependent memory effects appear only when EC conditions are distinguished sufficiently in terms of familiarity or disruption, or familiarity and disruption. Clearly, a replication of Smith's results would provide considerable reassurance as to the existence of a genuine memory-based EC-reinstatement effect, rather than an essentially artifactual, general-performance-based EC-reinstatement effect, or even EC-dependent memory effects reliant on Type I

errors. Therefore, the first experiment reported is a replication of Smith (1979, Experiment 1).

Stimulus Item Cue Strength

Another account of EC-dependent memory effects attributes their unreliability to memory retrieval, which is governed primarily by strong stimulus item cues that conceal any influence of EC retrieval cues (e.g., Dalton, 1993; Geiselman & Bjork, 1980; Nixon & Kanak, 1985). Smith (e.g., 1986) poetically labeled this the *outshining hypothesis*, suggesting that the way a bright light outshines and masks any changes in a dim light is analogous to the way high-strength stimulus item cues mask any influence of lower strength EC cues.

Murnane and Phelps (1995) examined the influence of stimulus item cue strength on EC-dependent recognition memory in relation to their general context model (Murnane & Phelps, 1994) prediction that EC-reinstatement benefits recognition memory performance. The general context model encapsulates formal model global activation accounts of recognition performance (e.g., CHARM, Eich, 1985; SAM, Gillund, & Shiffrin, 1984; Minerva 2, Hintzman, 1988; Matrix Model, Humphreys et al., 1989; TODAM, Murdock, 1993) with respect to EC and other context (e.g., semantic, physiological) changes.

In contrast to many other accounts, global activation models assume that at recognition all of the representations in long-term memory that share common features with the test cues constituting the memory probe are activated. The degree of activation of each memory representation is determined by the nature of the memory representation and its match with the memory probe. Recognition decisions are based on global activation levels, given by the sum of the scaled memory representation activations,

$$M = \sum_{j=1}^{K} f(I_j, C),$$ (1)

where M is global activation, I is the strength of the match between the item information in the cue and memory (i.e., the cuing strength of stimulus items), C is the strength of the match between the context information in the cue and memory (i.e., the cuing strength of the context), K is the number of items activated, and f is the activation function.

As a consequence of the way different formal model mechanisms activate items in memory, when the issue of stimulus cue strength is addressed, two variants of the general context model emerge (Dougal & Rotello, 1999; Murnane & Phelps, 1995). The multiplicative model variant (e.g., SAM, Minerva 2) predicts that greater stimulus item cue strengths will increase EC-dependent recognition memory effects, while the additive model variant (e.g., CHARM, Matrix Model, TODAM) predicts that greater stimulus item cue strengths will have no influence on the magnitude of EC-dependent recognition memory effects. However, there is no prediction concordant with the outshining hypothesis. Over three experiments, Mur-

nane and Phelps (1995) manipulated stimulus item cue strength by varying the number of repetitions of stimulus items, the amount of study time per stimulus item, and the LOP of stimulus items, while the cue strength of EC (defined as the specific combination of foreground and background colors, and spatial locations of stimulus words on a computer screen) was held constant. Since subjects' EC-dependent recognition performance increased with stimulus repetitions and was unaffected by increases in stimulus study time or the LOP of stimulus items, the experimental results supported the general context models' accounts, at the expense of the outshining hypothesis.

In an attempt to distinguish between the two variants of the general context model, Dougal and Rotello (1999) also investigated the influence of stimulus item cue strength on recognition memory. Defining EC in terms of computer screen foreground and background colors, they carried out two experiments that varied the frequency of presentation of stimulus items and one experiment that manipulated encoding LOP. All three experiments provided data consistent with an additive global matching model.

As well as global activation theories of recognition memory, there are theories that assume that each recognition decision is based on a separate assessment of the individual test item memory representation (e.g., Glanzer, Adams, Iverson, & Kim, 1993; Jacoby, 1991; Mandler, 1980; Tulving, 1983). Although Murnane and Phelps (1995) considered their experimental paradigm as one most likely to produce global-activation-based recognition performance, they did acknowledge that strong non-EC (i.e., stimulus item) cues might exert more influence on memory processes that retrieve individual memory representations. The nature of the free-recall task requires that individual items be provided. Theoretical accounts (e.g., Eich, 1985; Gillund & Shiffrin, 1984; Hintzman, 1988; Tulving, 1983) reflect this requirement by describing the retrieval of individual memory representations. Therefore, if strong non-EC cues have a powerful influence on the processes retrieving individual memory representations, then the effect of strong non-EC cues on EC-dependent memory should be most evident when free recall is used to test memory.

The outshining hypothesis was developed largely in response to failures to detect EC-dependent recognition memory (Smith, 1988). However, Murnane and Phelps (1995) and Dougal and Rotello (1999) have demonstrated that the outshining hypothesis does not apply to EC-dependent recognition memory. It is therefore somewhat ironic that support for the outshining hypothesis account of inconsistent EC-dependent memory effects is left to be sought with free recall. Certainly, failure to detect any influence of different stimulus item cue strengths on EC-dependent free recall would render the outshining hypothesis untenable.

The outshining hypothesis is tested in Experiment 2 by requiring free recall after a LOP manipulation at encoding has varied the relative strength of the stimulus item

retrieval cue information (e.g., Dougal & Rotello, 1999; Murnane & Phelps, 1995). In the low-LOP condition, low stimulus item cue strengths are fostered by instructing subjects to simply keep a tally of the number of words presented that contain three or more vowels. This task requires each word to be processed separately in a graphemic and possibly phonetic manner, and the maintenance of a running total of identified words. In the high-LOP condition, high stimulus item cue strengths are fostered by instructing subjects to construct a story from the words presented and to imagine the story events in their mind's eye. This task requires a substantial amount of semantic processing and should result in each word being processed in a semantic manner that links successive words. The LOP manipulation is described more accurately as a manipulation of processing type (Lockhart & Craik, 1990), with predominantly item-specific processing in the low-LOP condition and predominantly relational processing in the high-LOP condition. If there is any merit in the outshining hypothesis, then EC-dependent free recall should be observed after low LOP, but not after high LOP.

Recognition and Recall

It has been established for some time that forgetting rate is unaffected by processing type (e.g., McBride & Dosher, 1999; Slamecka & McElree, 1983) and that relational processing provides equal benefit to recognition and recall (e.g., Craik & Tulving, 1975; Hyde & Jenkins, 1973). For a while it was thought that item-specific processing conferred benefit only on recognition (e.g., Craik, 1981). However, it is known now that item-specific processing also facilitates recall, although to a lesser extent than recognition (Engelkamp, Biegelmann, & McDaniel, 1998; Roediger & Guynn, 1996). Therefore, given relational processing of a list of words, it follows that measures of subjects' overall recognition and recall performance (on mutually exclusive parts of the list) will be correlated. Similarly, given item-specific processing of a list, the same recognition and recall measures also will be correlated, but to a lesser extent. A variety of recognition and recall models can provide accounts of these effects (e.g., Gillund & Shiffrin, 1984; Haist, Shimamura, & Squire, 1992; Jacoby, 1991; Mandler, 1980). For current purposes, the important point is that all of these models are able to provide accounts of these effects because they assume a degree of overlap between recognition and recall processes that operate over a common set of memory representations.

While all of the formal models discussed provide accounts of recognition, only some describe recall (e.g., Eich, 1985; Gillund & Shiffrin, 1984; Hintzman, 1988; Metcalfe & Murdock, 1981). Compared with recognition in these models, a variety of additional processes are employed to produce individual items for recall. Recall memory probes, composed of the same kinds of information (although in different proportions) as recognition memory probes, are matched to the same memory representa-

tions as those assessed by the Phase 1 recognition processes. Since similar mechanisms employing similar information operate over the same memory representations in recognition and recall, it follows that greater stimulus item cue strength also is predicted to increase recall.

Equation 1 reveals that when context is held constant, the strength of the stimulus item cues is the only source of variation in global activation and so, subjects' recognition performance provides an overall index of stimulus item cue strengths. An immediate recognition test before the EC manipulation, as in Experiments 1 and 2, not only holds EC constant but also reduces the likelihood of other context changes (e.g., semantic, physiological) between the initial presentation and the encoding of the recognition items. According to the general context model, therefore, any variation in subjects' recognition performance reflects differences in the general level of stimulus item cue strength they have achieved at encoding. Consequently, recognition performance measures are well suited for the role of covariates in analysis of covariance (ANCOVA) to control for variation in stimulus item cue strength, due to encoding variation, across subjects.

EXPERIMENT 1

The aim of Experiment 1 was to replicate Smith (1979, Experiment 1). The purpose of carrying out this replication is to provide reassurance that EC-reinstatement effects are due to the match between presentation and test ECs, and not disruption and test-EC familiarity variables.

Method

Subjects. Forty-eight undergraduate students participated as subjects. Approximately equal numbers of males and females were paid for participating in the experiment.

Environmental contexts. Three adjacent rooms—A, B, and C—were employed in the experiment. These rooms were constructed to appear distinctly different. Room A was a long, thin, empty room with white walls and bright fluorescent strip lighting, and contained at the end opposite the entrance a small black wooden table where subjects sat. The stimuli were projected along the length of the room onto the wall facing the subjects as they sat at the table. In contrast, Room B was a small, square-shaped cluttered room with dim red lighting and two walls draped with curtains, containing a large white Formica-topped table where subjects sat. Stimuli were projected over subjects' heads onto the wall they faced as they sat at the table. Room C was a square-shaped room that was linked to a similar adjoining room through an open doorway. Subjects sat at three large tables pushed together from where they were able to see into the adjoining room, in which there was a large three-field tachistoscope. Room C also was illuminated by fluorescent strip lighting, but less brightly than Room A, and contained a couple of traffic cones, a bicycle tire, and on two shelves to the subjects' right, a variety of electronic cables and plugs. The waiting area was a large landing outside the rooms and contained two chairs on either side of a large storage radiator.

Stimuli. Eighty high-frequency (>50 occurrences per million words), logically unrelated nouns of 3–6 letters were selected from the Kučera and Francis (1967) word count. Two sets of 80 word slides were constructed: black print on a clear background and white print on a black background. (The latter were used in Room B

Table 1
Means and Standard Deviations of Recognition Hit Rate (HTR), Recognition
False Alarm Rate (FAR), Number of Words Recalled, and Adjusted Number
of Words Recalled (A Recall) by Experimental Condition

EC Familiarity and Matching	Presentation EC							
	HTR		FAR		Recall		A Recall	
	M	SD	M	SD	M	SD	M	SD
Same and Familiar								
A	0.88	0.11	0.05	0.08	27.13	9.25	26.22	6.12
B	0.76	0.13	0.03	0.05	26.28	6.28		
Different and Familiar								
A	0.76	0.18	0.09	0.17	21.75	9.63	21.87	6.00
B	0.79	0.14	0.03	0.05	20.38	8.18		
Different and Unfamiliar								
A	0.85	0.18	0.05	0.05	21.25	4.62	20.16	3.49
B	0.76	0.19	0.01	0.04	19.63	4.78		

Note—EC, environmental context.

to avoid extra illumination.) Each slide consisted of one centrally located word printed in uppercase. Seventy words were selected randomly as the initial presentation set. The presentation order of these 70 words also was selected randomly. The recognition test used 10 of the 70 words presented as targets and the remaining 10 words as distractors. The recognition targets were selected randomly given the constraint that they were distributed evenly over the 70 words presented. Recognition targets and distractors were randomly arranged and printed as a single column of uppercase words. Each subject received a slip of paper on which the recognition test was printed.

Apparatus. Two Kodak Carousel Projectors were used to present the word slides. A Kodak electronic interval timer controlled presentation rate.

Design. A two-factor (3 × 2) completely independent design was applied. The first factor was defined as the relationship between the presentation and the test EC and the familiarity of the test EC: same and familiar (SF), different and familiar (DF), and different and unfamiliar (DU). Although presentation and test ECs were counterbalanced within each SF, DF, and DU condition, to allow analysis of any room-specific EC effects, the second factor was defined by the presentation EC (Room A or B). Equal numbers of subjects were assigned randomly to each of the experimental conditions. Disruption was controlled by having all subjects move between the waiting area and the room ECs after each main period. Test-EC familiarity was manipulated across DF and DU conditions by having subjects draw an EC for the same amount of time as they had spent in the presentation EC. In DF conditions, the test EC was drawn, in DU conditions an "irrelevant" EC (Room C) was drawn. Presentation and test ECs are the same for subjects in the SF condition, so subjects are already familiar with the test EC. For half the subjects in the SF condition, the room–EC order was ABA, while the other half of the SF subjects experienced the presentation- and test-EC counterbalanced order, BAB. Similarly, half the DF subjects experienced the room–EC order ABB, while the other DF subjects experienced the counterbalanced room–EC order, BAA. Likewise, half the DU subjects experienced the room–EC order ACB, while the other DU subjects experienced the counterbalanced room–EC order, BCA.

Procedure. All subjects were tested individually. Each subject's experimental session comprised three phases, separated by two short intervals spent sitting in a waiting area. To discourage memory rehearsal during the drawing period, subjects were told they would be participating in two separate experiments—one on memory and then another on perception. In the first period, subjects viewed a presentation of the 70 stimulus words. Subjects were instructed to memorize these words and were informed that each word would be

presented for 3 sec with a 1.5-sec gap between words. Subjects initiated the presentation by switching on the power to the projector as soon as the experimenter left the room. (Subjects were left alone to carry out their tasks in each EC to try to prevent the experimenter from becoming part of the prevailing EC and later, a potential EC cue.) When the presentation was complete, subjects opened the door of the room EC and informed the experimenter, who returned and administered the recognition test. Subjects were given the recognition test slip and a pen and were asked to mark those words they recognized from the presentation. All subjects received 2.5 min to complete the recognition task and all finished the task before the time expired. In line with the experiment cover story, the impression was conveyed that this concluded the memory part of the study. The first period was complete within 8.5–9 min, by which time subjects were seated in the adjacent waiting area. After waiting for 1.5 min, subjects moved to the next EC, where they were told they would take part in a study of spatial representation. In the second period, subjects were asked to draw the room and its contents in detail. Subjects were asked to continue drawing until the experimenter returned, after 9 min. In the third period, subjects entered the appropriate test EC and were asked to write down all the words they could remember from the initial presentation, including any words from the recognition test. (Pilot work revealed that subjects could experience difficulties that disrupted their recall if they tried to discriminate between words appearing only in the initial presentation and words appearing in the recognition test.) Subjects were told they had 10 min to do this and were asked to start as soon as the experimenter left the room and to try to continue until the experimenter returned. Finally, all subjects returned to the waiting area, where they were paid and debriefed.

Scoring. Three classes of words could be recalled—distractor items from the recognition test, target items from the recognition test, and presented items that had not been used in the recognition test. To eliminate variance due to the different processing of target and distractor items during the recognition task and to provide distinct stimulus sets for recognition and recall, only subjects' recall of items that had not been used in the recognition test was analyzed. Henceforth, it is these scores that are termed *recall scores*. Recognition test performance hit rate (HTR) and false alarm rate (FAR) were calculated as recommended by Snodgrass and Corwin (1988) and Hautus (1995).

Results and Discussion

Table 1 presents the means and standard deviations of recognition HTR, FAR, recall, and adjusted recall by condition. A two-factor (3 × 2) ANOVA was applied to

subjects' recall scores. Checks of the ANOVA model assumptions suggested that all were tenable. The ANOVA revealed that only the EC match/familiarity factor exerted a significant effect on recall [$F(2,42) = 3.532$, $MS_e = 54.798$, $p = .038$]. The lack of effect of presentation EC [$F(1,42) = 0.342$] indicates that recall was not affected by the EC in which the stimulus words were presented. Moreover, the lack of an interaction between presentation EC and EC match/familiarity [$F(2,42) = 0.015$] indicates that recall did not benefit from any particular combination of presentation EC and memory test EC beyond that explained by the EC-match/familiarity factor.

Subsequent one-tailed comparisons revealed a significant difference between SF and DF conditions [$t(42) = 2.170$, $p = .018$, d (effect size) = 0.67], supporting an account of EC effects based on the benefit to memory of a match between presentation and test ECs, while the lack of a significant difference between DF and DU conditions [$t(42) = 0.263$, $p = .397$] undermines the familiarity account of EC effects. In short, Smith's (1979, Experiment 1) results are replicated, providing important support for a memory-based account of the EC-reinstatement effect.

A two-factor (3 × 2) ANCOVA using HTR and FAR as covariates was applied to the recall data. ANCOVA model assumption checks indicated all were tenable. HTR and FAR were employed as covariates to avoid the consequences of any inappropriate theoretical assumptions underlying detection theory sensitivity measures, such as the invariance of receiver operating characteristic (ROC) space or normal HTR and FAR distributions with equal standard deviations (Macmillan & Creelman, 1991; Ratcliffe, Sheu, & Gronlund, 1992; Swets, 1986a, 1986b; Swets & Picket, 1982). Preliminary investigation had revealed that employing d' and A' (e.g., Snodgrass & Corwin, 1988) as covariates, in separate ANCOVAs, tended to violate ANCOVA model assumptions, and remedy by dependent variable transformation provided ANCOVA models that accommodated less recall variance than those using HTR and FAR as covariates.

As in the ANOVA, presentation EC exerted no effect on recall, and consequently the data were collapsed across the presentation-EC factor levels to simplify analysis. The adjusted mean recall and standard deviations pertinent to the EC-match/familiarity effect are presented in Table 1. The single-factor ANCOVA revealed EC match/familiarity as exerting a significant effect on recall [$F(2,43) = 3.814$, $MS_e = 40.517$, $p = .030$]. Both covariates were significant predictors of recall [HTR, $F(1,43) = 7.827$, $r = 16.945$, $p = .008$; FAR, $F(1,43) = 5.662$, $r = -26.361$, $p = .022$]. ANCOVA model assumption checks indicated all were tenable. One-tailed comparisons revealed a significant difference between SF and DF conditions [$t(43) = 1.907$, $p = .032$, $d = 0.72$], but no significant difference between DF and DU conditions [$t(43) = 0.752$, $p = .228$].

The results of both the ANOVA and the ANCOVA replicate the results of Smith (1979, Experiment 1), sup-

porting a memory-based account of the EC-reinstatement effect, which depends on the match between the presentation and test ECs, and eliminating familiarity and disruption as candidates for explaining EC-reinstatement effects. This replication also suggests that it is most unlikely that Type I errors were responsible for other reports of EC-dependent memory effects.

EXPERIMENT 2

The purpose of Experiment 2 was to test the relative cue strength account of inconsistent EC-dependent memory effects with free recall. Subjects were presented with stimulus words, as in Experiment 1, but relative cue strength was varied by a LOP manipulation, as described in the introduction. In terms of cue strength, high LOP should result in stimulus item memory representations with high cue strengths, while low LOP should result in stimulus item memory representations with low cue strengths (see, e.g., Dougal & Rotello, 1999; Murnane & Phelps, 1995). If stimulus item cue strengths are high, then these non-EC memory cues should dominate retrieval processes, eliminating any effect of EC memory cues. In such circumstances, no EC-reinstatement effect is predicted. In contrast, if stimulus item cue strengths are low, then these non-EC memory cues should have little influence on retrieval processes. In these circumstances, the influence of any EC memory cues present in the test EC should be accentuated, and a large EC-reinstatement effect is predicted.

Method

Subjects. Sixty-four undergraduate students participated in the experiment. Approximately equal numbers of males and females completed the experiment in an appropriate manner and all were paid for participating.

Environmental contexts. As only SF and DF conditions were employed, only Rooms A and B were required for this experiment.

Stimuli. The stimuli, presentation times, and so on, were the same as those used in Experiment 1.

Apparatus. Slides were projected in Rooms A and B, as in Experiment 1, but this time by a computer-controlled slide presentation system. Slide presentation was initiated by one press on footswitches placed under the tables in Rooms A and B. A microphone in each of the two rooms was connected to a stereo cassette deck. A computer-operated buzzer, placed in the waiting area, informed the experimenter that the presentation was complete.

Design. A two-factor (2 × 2) completely independent design was applied. The first factor was defined by the LOP: high or low. The second factor was defined by the relationship between the presentation and test ECs: same or different. Equal numbers of subjects were assigned randomly to each of the experimental conditions. As in Experiment 1, presentation and test ECs were counterbalanced. Moreover, subjects were again asked to draw a room EC so that they would be occupied for 10 min. So, for the very little extra cost of returning subjects to the waiting area between each 10-min period, both disruption and test-EC familiarity were again controlled. Therefore, but for the exclusion of the DU conditions, room orders for Experiment 2 were identical to those of Experiment 1. The room orders for subjects in reinstated (SF) conditions were ABA, with the BAB counterbalance experienced by half these subjects. The

Table 2
Means and Standard Deviations of Recognition Hit Rate (HTR),
Recognition False Alarm Rate (FAR), Number of Words Recalled,
Transformed Number of Words Recalled (T Recall), and Adjusted Transformed
Number of Words Recalled (A T Recall) by Experimental Condition

LOP Condition	EC Match									
	HTR		FAR		Recall		T Recall		A T Recall	
	M	SD	M	SD	M	SD	M	SD	M	SD
High										
Same	0.79	0.11	0.04	0.05	19.25	8.21	8.72	1.79	8.77	1.49
Different	0.79	0.12	0.03	0.06	19.25	8.00	8.71	1.90	8.58	1.50
Low										
Same	0.54	0.17	0.09	0.04	6.69	4.22	5.08	1.73	5.66	1.61
Different	0.67	0.15	0.19	0.14	4.63	3.69	4.10	1.93	3.50	1.62

Note—M adjusted transformed recall (A T Recall) = transformed recall scores predicted from the levels-of-processing (LOP) condition covariate means.

room orders for subjects in changed (DF) conditions were ABB, with the BAA counterbalance experienced by half these subjects.

Procedure. The procedure was similar to that employed in Experiment 1. However, subjects in the low-LOP condition were asked to count the number of words containing three or more vowels, whereas subjects in the high-LOP condition were asked to try to form a story using the presented words and imagine the events of this story in their mind's eye as they created it. In the recall period, rather than write down the remembered words, subjects were instructed to say each word remembered out loud, so that their recalls could be recorded. They were asked to start as soon as the experimenter left the room and to try to continue until the experimenter returned. Due to the time required to convey the encoding instructions, the first period took 10 min to complete, and so the second drawing period was set at 10 min. Also, to more easily accommodate the time subjects needed to move to the waiting area from the room ECs, the waiting period was increased to 2 min. In general, therefore, the gap between the end of the presentation and start of subjects' recalls was about 2.5 min longer than in Experiment 1. After the 10-min recall period, all subjects returned to the waiting area. Here they were asked to complete a questionnaire, which comprised the open-ended questions, "As the words were presented to you, what did you do?" and, "When you were asked to recall the words, what did you do to remember them?" A rating of the similarity of the two rooms on a 1–100 scale (where a greater value indicated greater similarity) also was required. When the questionnaire was complete, subjects were paid and debriefed. Although subjects were naive regarding the purpose of the experiment, questionnaire responses indicated that 4 subjects in the low-LOP condition had not followed instructions and had tried to memorize the words. Since these subjects were identified by the questionnaire, there was no need to transcribe and score their recall protocols. They were replaced by subjects whose questionnaire responses did not indicate an attempt to memorize the stimulus words. (All subjects in the high-LOP condition indicated that they had employed the mnemonic strategies as instructed.) All subjects were tested individually.

Scoring. Scoring was identical to that in Experiment 1.

Results and Discussion

The median rating of the similarity of Rooms A and B was 28 ($M = 33.00$, $SD = 22.73$). Since low ratings indicated that the rooms appeared to be different, responses suggested that the attempt to make the two room ECs appear very dissimilar had been successful.

Table 2 presents means and standard deviations of subjects' HTR, FAR, recall, transformed recall, and adjusted transformed recall by experimental condition. A two-factor (2 × 2) ANOVA was applied to subjects' transformed recall scores [$(recall)^{0.5} + (recall + 1)^{0.5}$; Freeman & Tukey, 1950] to remedy heterogeneous error variance across LOP conditions. A significant effect of LOP was detected [$F(1,60) = 79.954$, $MS_e = 3.387$, $p < .001$], but no other effects were significant [EC match, $F(1,60) = 1.193$; LOP × EC–match interaction, $F(1,60) = 1.097$]. The size of the same-EC versus different-EC effect in low-LOP conditions was also calculated ($d = 0.54$).

The relative strength of stimulus item cues depends entirely on encoding processing. In such circumstances, any imbalance in encoding processing between same- and different-EC conditions will compromise the hypothesis test. This is particularly so for low-LOP same- or different-EC conditions, where, due to lower average recall, any switch to higher LOPs will have a greater proportionate effect than a converse switch in high-LOP conditions. In these circumstances, ANCOVA is most appropriate, because it assesses the relation between the covariate measures of encoding and the dependent variable and remedies the dependent variable for the consequences of any covariate imbalance across conditions (Senn, 1989).

Application of a heterogeneous regression ANCOVA in this experiment is motivated by the possibility of recognition–recall relationship differences due to the LOP manipulation discussed in the introduction. The LOP manipulation also affects the recognition covariate scores. Therefore, a hierarchical heterogeneous regression ANCOVA is an appropriate technique for analyzing such data (see Rutherford, 1992, for a detailed account of this analysis strategy). A two-factor (2 × 2) hierarchical heterogeneous regression ANCOVA was applied to subjects' transformed recall scores. The order in which the experimental factors and covariates are reported reflects their order of entry into the ANCOVA model.

Recall variance attributed to the LOP factor was removed first [$F(1,56) = 122.139$, $p < .001$]. Since all

variance attributable to LOP now has been removed, any variance subsequently attributed to HTR and FAR is distinct from that associated with the LOP manipulation. Variance attributed to the covariate HTR was removed [$F(1,56) = 18.830, p < .001$], followed by any additional variance attributable to fitting separate recall on HTR regressions within each LOP condition [LOP × HTR, $F(1,56) = 0.824, p = .368$; high LOP $r = 7.635$, low LOP $r = 6.916$]. Although HTR was a highly significant predictor, an insignificant amount of variance was attributable to fitting separate recall on HTR regressions within each LOP condition. (Prior to removing variance associated with LOP, significant variance was attributable to separate recall on HTR regressions within LOP conditions [$F(2,62) = 50.990, p < .001$]. These results support the hypothesis that the different relations between HTR and recall in high- and low-LOP conditions are due to the LOP manipulation.)

Next, variance attributable to the covariate FAR was removed [$F(1,56) = 1.339, p = .252$], followed by the variance attributable to the separate regressions of recall on FAR in the LOP conditions [LOP × FAR, $F(1,56) = 5.542, p = .022$; high LOP $r = -13.320$, low LOP $r = 3.364$]. In contrast to HTR, it is the separate recall on FAR regressions that accommodate significant variance. The significant LOP × FAR interaction is due to the lack of relationship between subjects' FAR and recall scores in low-LOP conditions [$F(1,56) = 1.652, p = .217$] (despite larger FAR mean and variance), and a significant relationship between subjects' FAR and recall scores in high-LOP conditions [$F(1,56) = 7.001, p = .011$]. Nevertheless, FAR scores are equal in high-LOP same- and different-EC conditions, so the only consequence of the LOP × FAR interaction is ANCOVA error term reduction.

Next, the EC-match factor was included in the AN-COVA model [$F(1,56) = 6.669, p = .012$], followed by the LOP × EC-match interaction [$F(1,56) = 5.962, MS_e = 2.217, p = .018$]. ANCOVA model assumption checks indicated all were tenable.

One-tailed comparisons examining the nature of the LOP × EC-match interaction revealed a significant difference between the adjusted means of low-LOP same- and different-EC conditions [$t(56) = 5.362, p < .001, d = 1.34$], but not between the adjusted means of high-LOP same- and different-EC conditions [$t(56) = 0.500, p = .310$]. These results are consistent with the hypothesis that high stimulus cue strengths dominate individual item retrieval processes masking or displacing any influence of EC memory cues.

Further analysis of the relationship between LOP and the cuing strength of stimulus items was carried out by comparing an ANCOVA model comprising only the variables HTR and FAR with an ANCOVA model comprising the variables LOP, HTR, and FAR. This revealed that a significant proportion of recall variance was attributable uniquely to LOP [$F(1,60) = 21.326, MS_e = 2.762, p < .001$]. This model comparison demonstrates that while there is considerable overlap in the recall variance accom-

modated by the LOP manipulation and the recognition performance measures (due to LOP affecting recognition), there are also distinct aspects of recall associated with the LOP manipulation and the recognition measures. In other words, stimulus item cue strength, as indexed by recognition measures, does not fully account for the LOP effect on recall.

GENERAL DISCUSSION

Two studies of EC-dependent memory were reported in which the roles of EC familiarity and the disruption caused by moving between two ECs and the relative strength of retrieval cues were examined for their ability to explain the unreliable nature of EC-dependent memory effects in free recall. Experiment 1 replicated Smith (1979, Experiment 1), providing important confirmation that familiarity and disruption cannot account for the EC-reinstatement effect. Indeed, Experiment 2 also controlled familiarity and disruption and also detected an EC-reinstatement effect. Consequently, despite notable failures to obtain EC-dependent memory effects, positive findings cannot be dismissed as artifacts of familiarity, or disruption variables, or as Type I errors. In Experiment 2, the relative cue strength account of inconsistent EC-dependent memory effects was tested. The results revealed that subjects' encoding strategies have consequences for EC-dependent memory effects, at least when a free-recall task is used to assess memory. Given that encoding processes usually are not controlled in EC-dependent memory studies, this may explain some of the apparent unreliability of EC-dependent memory effects. This point is emphasized when it is considered that despite a medium-sized EC effect in Experiment 2, it was declared statistically significant only after an ANCOVA increased analysis power by controlling encoding variation within LOP conditions over same- and different-EC conditions.

In Experiment 2, the LOP manipulation affected stimulus item cue strengths, as indexed by recognition test scores. An EC-reinstatement effect was observed with low stimulus item cue strengths, but was not observed with high stimulus item cue strengths. These results appear to support the outshining hypothesis, which holds that high stimulus item cue strengths are able to eliminate free-recall EC-reinstatement effects. However, there is a problem with this account. Since stimulus item cue strength in Experiment 1 is comparable with that observed in the high-LOP condition of Experiment 2 (Experiment 1, M HTR = 0.81, M FAR = 0.04; Experiment 2, High LOP M HTR = 0.79, M FAR = 0.04), no EC-reinstatement effect should be observed in Experiment 1. Therefore, something other than recognition test indexed stimulus item cue strength must be responsible for the pattern of EC-dependent memory effects observed over the two experiments.

The Experiment 2 data analysis revealed that LOP manipulation exerts an influence on subjects' recall scores beyond that reflected in their recognition scores and at-

tributed to stimulus item cue strength. This unique contribution of LOP may reflect the retrieval strategies available in high- but not low-LOP conditions. With semantically elaborated memory representations in high-LOP conditions, as a consequence of relational processing, retrieval of one item is likely to provide retrieval cues for other items. The item-specific processing required in low-LOP conditions is much less likely to provide this retrieval benefit, and consequently there may be little alternative but to make greater use of EC information in the memory retrieval attempts. Essentially, this is the outshining hypothesis again, but operating through some mechanism other than, or in addition to, stimulus item cue strength. The possibility that task demands in high-LOP conditions prevent EC encoding is unlikely, since recognition EC-reinstatement effects were detected by Murnane and Phelps (1995) and Dougal and Rotello (1999) after similar LOP manipulations.

Comparison across these experiments is problematic because they differ with respect to encoding instructions, the time between presentation and test, and the manner of subjects' recall (written or verbal). In spite of these differences, the observed pattern of EC-reinstatement effects favors the retrieval strategy account. However, further work is necessary to clarify the relations between EC-dependent memory effects, encoding processing, and free-recall retrieval strategies. In particular, understanding which encoding and especially which retrieval conditions support the use of EC information would be a significant advance.

REFERENCES

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

BJORK, R. A., & RICHARDSON-KLAVEHN, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition* (pp. 313-344). Hillsdale, NJ: Erlbaum.

CRAIK, F. I. M. (1981). Encoding and retrieval effects in human memory: A partial review. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 383-402). Hillsdale, NJ: Erlbaum.

CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268-294.

DALTON, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory & Cognition*, **21**, 223-234.

DOUGAL, S., & ROTELLO, C. M. (1999). Context effects in recognition memory. *American Journal of Psychology*, **112**, 277-295.

EICH, J. M. (1985). Levels of processing, encoding specificity, elaboration and CHARM. *Psychological Review*, **92**, 1-38.

ENGELKAMP, J., BIEGELMANN, U., & MCDANIEL, M. A. (1998). Relational and item-specific information: Trade-off and redundancy. *Memory*, **6**, 307-333.

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

FREEMAN, M. F., & TUKEY, J. W. (1950). Transformation related to the angular and square root. *Annals of Mathematical Statistics*, **21**, 607-611.

GEISELMAN, R. E., & BJORK, R. A. (1980). Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. *Cognitive Psychology*, **12**, 188-205.

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

GLANZER, M., ADAMS, J. K., IVERSON, G. I., & KIM, K. (1993). The regularities of recognition memory. *Psychological Review*, **100**, 546-567.

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

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

HAIST, F., SHIMAMURA, A. P., & SQUIRE, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 691-702.

HAUTUS, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers*, **27**, 46-51.

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

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.

HYDE, T. S., & JENKINS, J. J. (1973). Recall for words as a function of semantic, graphic and syntactic orienting tasks. *Journal of Verbal Learning & Verbal Behavior*, **12**, 471-480.

JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

KUČERA, H., & FRANCIS, W. N. (1967). *A computational analysis of present-day American English*. Providence: Brown University Press.

LOCKHART, R. S., & CRAIK, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology*, **44**, 87-112.

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

MANDLER, G. (1980). Recognizing: The judgement of previous occurrence. *Psychological Review*, **87**, 252-271.

MCBRIDE, D. M., & DOSHER, B. A. (1999). Forgetting rates are comparable in conscious and automatic memory: A process-dissociation study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 583-607.

METCALFE, J., & MURDOCK, B. B. (1981). An encoding and retrieval model for single-trial free recall. *Journal of Verbal Learning & Verbal Behavior*, **16**, 699-709.

MORTON, J., HAMMERSLEY, R. H., & BEKERIAN, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, **20**, 1-23.

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. (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*, **21**, 158-172.

NIXON, S. J., & KANAK, N. J. (1985). A theoretical account of the effects of environmental context upon cognitive processes. *Bulletin of the Psychonomic Society*, **23**, 139-142.

RATCLIFFE, R., SHEU, C. F., & GRONLUND, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.

ROEDIGER, H. L., III, & GUYNN, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (Handbook of perception and cognition, 2nd ed., pp. 197-236). San Diego: Academic Press.

RUTHERFORD, A. (1992). Alternatives to traditional analysis of covariance. *British Journal of Mathematical & Statistical Psychology*, **45**, 197-223.

SENN, S. J. (1989). Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine*, **8**, 467-475.

SLAMECKA, N. J., & MCELREE, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 384-397.

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

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. Thompson (Eds.), *Memory in context: Context in memory* (pp. 13-34). Chichester, U.K.: Wiley.

SMITH, S. M., GLENBERG, A. M., & BJORK, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, 6, 342-353.

SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics in measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.

STRAND, B. Z. (1970). Change of context and retroactive inhibition. *Journal of Verbal Learning & Verbal Behavior*, 9, 202-206.

SWETS, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, 99, 181-198.

SWETS, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, 99, 100-117.

SWETS, J. A., & PICKET, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.

TULVING, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.

WICKENS, D. D. (1987). The dual meanings of context: Implications for research, theory and applications. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus Centennial Conference* (pp. 135-152). Hillsdale, NJ: Erlbaum.