Journal of Experimental Psychology:
Learning, Memory, and Cognition, in press

Remembering Remembering

Michelle M. Arnold and D. Stephen Lindsay
University of Victoria

We developed a laboratory analogue of the "forgot-it-all-along" effect that Schooler, Bendiksen, and Ambadar (1997) proposed for cases of "recovered memories" in which individuals had forgotten episodes of talking about the abuse when they were supposedly amnestic for it. In Experiment 1, participants studied homographs with disambiguating context words; in Test 1 they received studied- or other-context words as cues, and in Test 2 they received studied-context cues and judged whether they had recalled each item during Test 1. Experiment 2 manipulated retrieval cues on both tests. In Experiment 3, both the studied- and other-context cues corresponded to the same meaning of each homograph. In Experiment 4, Test 1 was free recall and studied- versus other-context cues were presented in Test 2. Participants more often forgot that they had previously recalled an item if they were cued to think of it differently on the two tests.

The controversy regarding reported "recovered memories" of childhood sexual abuse has highlighted a number of questions of interest to memory researchers (see Lindsay & Read, in press, for a recent enumeration of some of these). Among these is the question of how people make judgments about whether or not they had previously recollected a particular episode that they currently recollect.

Schooler and his coworkers (Schooler, 1999, in press; Schooler Ambadar, & Bendiksen, 1997; Schooler, Bendiksen, & Ambadar, 1997) described two fascinating cases in which women reported full-blown recovered-memory experiences. What makes these two cases particularly interesting is that in each a close confidant of the woman involved reported that the woman had talked about the abuse during the period of supposed amnesia. Schooler and his coauthors speculated that during the recovered-memory experience the women remembered the abuse in a different way than they had previously (e.g., more completely, more episodically, or with a qualitatively different interpretation), such that the experience of remembering was very emotionally intense and qualitatively different from their previous recollections of the abuse, and that this in turn gave rise to what they termed a "forgot-it-all-along effect" (in reference to Fischhoff's, 1977, knew-it-all-along effect). That is, recollecting an event in manner X may cause one to forget having previously recollected it in manner Y.

To the best of our knowledge, the only published research on remembering prior instances of recollection is a study by Parks (1999), in which undergraduate participants were first asked to describe various events from their pasts and shortly thereafter asked how recently they had recollected those and other events. Parks found that participants often failed to report their Phase-1 reminiscences a few minutes later, when making the Phase-2 judgments. Relatedly, in a recent study by Padilla-Walker and Poole (in press), participants first studied a list of sentences, then in Test 1 they free-recalled the sentences, and in Test 2 they were given either a recognition test, cued recall test, or free recall test and were asked to judge whether they had recalled each sentence in Test 1. Participants more often failed to remember their Test 1 recall if Test 2 involved a recognition or cued recall task than if Test 2 was free recall. Although there are other potential explanations, this pattern is consistent with the notion that a change in the way individuals think about a past event across different episodes of remembering (in this case, from free to cued recall or recognition) increases the likelihood that individuals will forget a prior instance of remembering (see also Joslyn, Loftus, McNoughton, & Powers, in press).

We developed a laboratory analog of the reinterpretation process that Schooler (in press) proposed contributes to the forgot-it-all-along (FIA) effect. The cognitive processes involved in recollecting childhood abuse may differ from those involved in recalling laboratory events. However, if a change in the way a past event is thought about on different occasions can lead to forgetting of previous episodes of recalling abuse, then a change in the way neutral laboratory events are thought about on different tests should also result in the forgetting of previous recollection. In our paradigm, participants first study some verbal materials, then attempt to recall those materials on two occasions; during the second recall occasion, participants are also asked if they had recalled each item during the first recall attempt. Three of the experiments reported here (Experiment 1, 2, and 4) tested for the FIA effect by using context words to manipulate the meaning of studied homographs between the two recall attempts. The other experiment (Experiment 3) tested for the FIA effect by manipulating the context (but not the meaning) of studied words across the two recall attempts. We predicted that participants would more often fail to remember, at Test 2, that they had recalled a word on Test 1 if they were led to think of the recalled words differently on the two tests.

Prior research on memory for events supports the prediction that a change across tests in the way a past event is thought about will contribute to the forgot-it-all-along effect. For instance, recognition failure occurs when participants fail to recognize studied target words paired at test with context words different from those with which targets had been studied, yet recall the target words when cued with the studied context words (Tulving & Thomson, 1973). Recognition failure has been demonstrated both for high-frequency words with multiple meanings (Tulving & Thomson, 1973) and for single-meaning words (Tulving & Watkins, 1977). More generally, memory for past events is usually (although not always) better when study and test occur in the same context, mood, or state (see Bouton, Nelson, & Rosas, 1999; Eich, 1995; Light & Carter-Soboll, 1970; cf. Franks, Bilbrey, Lien, & McNamara, 2000). The predominant explanation for such effects centers around Tulving's (1984) concept of encoding specificity: the closer the match between the encoding conditions and retrieval conditions, the more likely it is that the event will be successfully recalled/recognized. Recalling prior episodes of recollection is likely similar to recalling other types of past episodes. Therefore, as in the case of recognition failure, it is reasonable to hypothesize that

Michelle M. Arnold and D. Stephen Lindsay, Department of Psychology, University of Victoria, Victoria, British Columbia, Canada.

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through a postgraduate scholarship to Michelle M. Arnold and an NSERC research grant to D. Stephen Lindsay. Portions of this research were reported at the annual meeting of the Northwest Cognition and Memory Society, Bellingham, Washington, June 2000. We thank Michael E. J. Masson and J. Don Read for their suggestions on an earlier draft of this manuscript.

Correspondence concerning this article should be addressed to Michelle M. Arnold, Department of Psychology, University of Victoria, P.O. Box 3050, Victoria, British Columbia, V8W 3P5, Canada. Email: arnoldm@uvic.ca.

participants in our studies would more often forget prior remembering when they had been cued to think of the items differently on the two tests.

Experiment 1

Participants studied a list of homographic target words paired with disambiguating context words. On the first cued recall test participants were tested on two thirds of the target words, of which half were cued with the studied-context words and half were cued with other-context words. For example, a participant who studied the target word "palm" with the context word "hand" could be cued for this target word with the context word "hand" (*studied-context* condition) or with the context word "tree" (*other-context* condition). Another third of the items were not tested for on Test 1 (*not-tested* condition). In a second cued recall test participants were tested on all of the target words, and the target words were always cued with the studied-context words. Additionally, the second test required participants to judge whether they remembered having recalled each item in Test 1.

Method

Participants. Twenty six University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course. Two participants failed to follow the directions for the judgment task and their data were excluded from the analyses.

Materials. A set of 113 homographs with two dominant meanings (e.g., "palm" in the type-of-tree sense and in the part-of-hand sense) was constructed from various pools of homograph norms (e.g., Azuma, 1996; Gawlick-Grendell & Woltz, 1994; Twilley, Dixon, Taylor, & Clark, 1994). Four of the items were used as primacy buffers and four as recency buffers. The target words were randomly divided into three lists of 35 words (*test list* factor), with each list appearing equally often across subjects in the studied-context, other-context, and not-tested within-subject conditions of Test 1. A context word was assigned to each of the two different meanings of the target words (e.g., "tree" and "hand" for the target "palm"). Two study lists were constructed (*study list* factor) to counterbalance the meanings of the target words between subjects (e.g., "palm" studied with "tree" for half of the participants and with "hand" for the other half of the participants). A sentence was constructed for each context-target word pair for the study phase, containing the context word and a row of asterisk symbols for the target word (e.g., "He collected coconuts from the *** tree on the beach" and "He used the *** of his hand to swat the fly").

Procedure. Participants were tested individually on an IBM-compatible personal computer using Schneider's Micro Experimental Laboratory Professional software package (Schneider, 1988). Participants were seated in front of the computer monitor, with the experimenter sitting off to one side. The experimenter read the instructions aloud as they appeared on the screen for each phase. Participants were told that for each study trial a context word and target word would be displayed on the screen for approximately 2 s, and that their task was to repeat the words aloud in preparation for a subsequent memory test. The two words were then removed from the screen and a sentence containing the context word and a row of asterisk symbols for the target word was presented for 3.5 s. Participants were instructed to read the sentence aloud, filling in the row of asterisk symbols with the target word. After the 3.5 s elapsed, the target word appeared above the sentence for one second, after which the computer screen went blank for one second before the next trial began.

The first cued recall test immediately followed the study phase, with the items presented in random order. Participants were informed that they would be tested on half of the target words (they were actually tested on two thirds of the items, but it was easier to explain the task in terms of half of the items) and that this would be done by presenting a context word along with the first and last letters of a target word (e.g., "hand - p_ _ m" or "tree - p _ _ m"). Further, participants were instructed that for half of the trials the context words would correspond to the context words presented with the targets during the study phase, and that for the remaining trials the context words would not be the same as at study but that nonetheless the context words would be related to the target words, and they were given an example. The test instructions also warned participants only to respond with answers they *remembered* seeing during the study phase: Participants were instructed to say "Pass" if they did not remember the answer or if the answer they came up with was a guess (e.g., filling in the blanks instead of remembering the word from the study phase). The computer gave participants item-by-item feedback in the form of a tone for incorrect answers or responses of "Pass," and the phrase "Correct Response" for correct answers. After completing the first test participants were given a 5-min break during which they conversed with the experimenter before moving on to the second cued recall test.

All 105 critical target words were tested in random order on the second cued recall task. The second test was similar in format to the first test. For each trial participants were given a context word with the first and last letters of a target word separated by dashes and asked to recall the target word from the study phase. Participants were informed that all of the context words on Test 2 corresponded with the context words presented with the target words during the Study phase. As in Test 1, participants were told to respond with an answer only if they remembered seeing the target word during the Study phase. After each recall attempt the screen went blank, and whenever a participant gave an incorrect answer or said "Pass" the experimenter supplied the correct target word. Participants were then required to judge if they remembered recalling the target word during Test 1. Participants were explicitly instructed that their judgments should not be based on whether or not they had seen the Test 2 context word during Test 1, but rather on whether or not they remembered recalling the target word on Test 1. Participants were also reminded that many of the study items had not been tested for (and hence could not have been recalled) on Test 1. The experimenter emphasized that the task was not to judge if the context word had changed between the two tests, nor whether the target word had merely been tested on Test 1 (e.g., "If you remember that the computer beeped at you for a particular trial during the first test because you said 'Pass' or gave the wrong answer, then you should say 'No')." During the test, participants were stopped two or three times and reminded of the instructions for the judgment task.

Results and Discussion

Our interest focused on participants' judgments about prior remembering. Nonetheless, in this and the subsequent Results sections we first report analyses of recall performance on Test 1 and Test 2 (collapsed across the counterbalancing factors of study list and test list[1])

---

[1]The analyses of recall performance for all four experiments were initially performed using omnibus ANOVAs that included the study list and test list counterbalancing factors. In Experiments 2-4 there were sometimes significant effects of the study list and/or test list factors (although the pattern was not consistent across experiments). The assignment of the meanings of the target words to either of the two study lists, as well as the assignment of the items to a test list, was random; it appears that flukes of random assignment sometimes led more memorable senses of the target words to be assigned to one study list than the other, and sometimes led one version of the test list to be easier than the other. It is also possible that some of these effects were Type I errors. In any case, differences in recall performance due to the counterbalancing of the study and tests lists are irrelevant to our hypotheses regarding judgments of prior recollection, so we do not report these analyses here; they may be requested from either author.

before turning to analyses of the prior-remembering judgment data.

Recall performance. Proportion correctly recalled on Test 1 was significantly higher for items cued with studied-context words ($M = .91$) than for items cued with the other-context words ($M = .77$), $t(23) = 9.62$, $p < .0001$. A within-subjects ANOVA was performed on the proportion correctly recalled on Test 2, with context on Test 1 (studied, other, or not-tested) as the within-subjects factor. Performance was near ceiling (as intended), and there was no significant difference in proportion of items recalled for the studied-context ($M = .94$), other-context ($M = .92$), and not-tested ($M = .92$) conditions, $F(2, 46) = 1.36$, $MSE = .002$, $p > .27$.

Judgment of previous recollection. The proportion of items correctly judged on Test 2 as recalled on Test 1 is shown in Table 1. The analyses reported here were performed on the judgment data for items correctly recalled on both Test 1 and Test 2 (shown in bold in Table 1), although the same pattern of results was found when analyses were contingent only on correct recall on Test 1. In an initial omnibus ANOVA, no reliable effects of the counterbalancing factors of study list or test list were found (all *F*s < 1), and therefore the data were collapsed across these variables. A within-subjects ANOVA was performed on the proportion of correct "Yes" judgments, with context on Test 1 (studied vs. other) as the within-subjects factor. Participants were significantly more likely to forget that they had recalled an item on Test 1 if it had been cued with the other-context word on Test 1 than if it had been cued with the studied-context word, $F(1, 23) = 51.09$, $MSE = .02$, $p < .0001$. Participants rarely erred on not-tested items by saying that they had remembered those items on Test 1.

As predicted, participants were dramatically more likely to forget that they had recalled a word on Test 1 if Test 1 recall of that word had been cued with the other-context word than if it had been cued with the studied-context word. This finding is consistent with the hypothesis that changes in the way an event is remembered can produce a forgot-it-all-along effect. The results of Experiment 1 are also amenable, however, to a slightly different interpretation: It may be that cuing with the other-context word on Test 1 produced weaker or less complete recollection of the target word than did cuing with the studied-context word.[2] If so, then the experience of remembering during Test 1 might be less memorable for words in the other-context condition than for words in the studied-context condition. This mechanism could play a role in real-world cases like those described by Schooler (in press); that is, the women's pre-memory-recovery recollections of the abuse may have been vague or incomplete and hence simply not memorable.

Experiment 2

Experiment 2 was designed to assess the two accounts of the results of Experiment 1 mentioned above by manipulating context on both Test 1 and Test 2. If the tendency to forget prior remembering of words in the other-context condition of Experiment 1 was due only to weaker Test 1 recollections of those words, then in Experiment 2 that effect should be obtained only for items cued with other-context words during Test 1 and studied-context words on Test 2. If, on the other hand, shifts in how participants think about the target words on the two tests contribute to the effect, then it should also be observed for words cued on Test 1 with studied-context words but cued on Test 2 with other-context words.

The second experiment employed the same basic procedure as Experiment 1, but context was manipulated on both the first and the second recall tests. Six within-subject conditions were created, with target items: (a) tested with the studied-context word on Test 1 and Test 2 (*studied/studied* condition), (b) tested with the other-context word on Test 1 but the studied-context word on Test 2 (*other/studied* condition), ©) not tested on Test 1 and tested with the studied-context word on Test 2 (*not-tested/studied* condition), (d) tested with the studied-context word on Test 1 but with the other-context word on Test 2 (*studied/other* condition), (e) tested with the other-context word on Test 1 and Test 2 (*other/other* condition), and (f) not tested on Test 1 and tested with the other-context word on Test 2 (*not-tested/other* condition).

Method

Participants. Twenty seven University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course or a $10 payment. Three participants recalled fewer than 50% of the items in at least one of the conditions on Test 1 and/or Test 2 and their data were excluded from the analyses.

Materials. Three additional homographs were added to the set of homographs used in Experiment 1, resulting in a set of 116 target words. Four of the items were used as primacy buffers and four were used as recency buffers. The target words were randomly divided into six lists of 18 words (*test list* factor), with each list appearing equally often across participants in the studied/studied, other/studied, not-tested/studied, studied/other, other/other, and not-tested/other within-subject conditions of Test 1 and Test 2. Two study lists were constructed (*study list* factor) to counterbalance the meanings of the target words between subjects.

Procedure. The basic procedure of Experiment 1 was used, with four modifications. First, instead of viewing the context–target word pairs and sentences during the Study phase, participants heard the study materials read aloud. Participants were told that for each study trial the experimenter would read aloud a context word and a target word, followed by a sentence containing those two words. Participants were instructed to repeat the sentence aloud and then write down the target word on a sheet provided by the experimenter. This change to the way the study list was presented was intended to make it easier for participants, during Test 2, to differentiate between memories of studying a target word (hearing the context word and target word) and memories of recalling it on Test 1 (seeing a context word with the first and last letters of the target, and responding with the target word).

A second modification involved replacing the dashes between the first and last letters of the target word on Test 2 with asterisk symbols (e.g., "tree - p**m"). This difference was incorporated to assist participants in differentiating between Test 1 and Test 2.

As a third modification, participants were instructed on Test 2 that, as in Test 1, half of the context words would be the same as those presented with the target words during the study phase, and half of them would be different. Participants were also warned that the Test 2 context words presented with the items that had been tested on Test 1 could be the same as or different from the context words used to test for those items on Test 1.

The fourth change to the procedure was established to ensure that participants understood the difference between (a) remembering only being *tested on* a target word in Test 1 (being cued for a target word, but not necessarily recalling that target) and (b) remembering *recalling* the target word. On Test 1 participants were given a hand-held microphone and informed that their responses would be tape-recorded. The tape recorder was set so that participants could hear their voices come through the speakers. The instructions for the judgment task in Test 2 were the same as in Experiment 1, with an additional sentence that stated "Another way to think of the judgment task is, if the tape recording from Test 1 was played back, would you hear your voice saying the target word?" Participants were instructed to say "Yes" only if they remembered recalling and saying the target word on Test 1.

Results and Discussion

Recall performance. The proportion correctly recalled for each

[2] We thank Michael E. J. Masson for bringing this alternative explanation to our attention.

condition of Test 1 and Test 2 is shown in Table 2. Proportion correctly recalled on Test 1 was higher for items cued with the studied-context words (M = .88) than with other-context words (M = .77), $F(1, 23) = 57.76$, $MSE = .003$ $p < .0001$. The proportions of items correctly recalled on Test 2 were analyzed in a 3 (Test 1: studied-context, other-context, and not-tested) x 2 (Test 2: studied-context vs. other-context) within-subjects ANOVA. Correct recall on Test 2 was higher for items cued with the studied-context words (M = .90) than with the other-context words (M = .80), $F(1, 23) = 26.29$, $MSE = .01$, $p < .0001$. Correct recall on Test 2 was also influenced by context on Test 1, $F(2, 46) = 3.34$, $MSE = .01$, $p = .04$: Test 2 recall performance was slightly poorer for items not-tested on Test 1 (M = .83) than for items tested in the studied-context condition (M = .87), $t(23) = 2.19$, $p = .04$, or other-context condition (M = .86), $t(23) = 2.53$, $p = .02$. There was no interaction between context on Test 1 and Test 2, $F < 1$.

Judgment of previous recollection. The proportions of items on Test 2 correctly judged as recalled on Test 1 are shown in Table 3. The subsequent analyses were performed on the judgment data for items correctly recalled on both Test 1 and Test 2 (shown in bold in Table 3), although the same pattern of results was found when analyses were contingent only on correct recall on Test 1. No reliable effect of the test list counterbalancing factor for the judgment task was found in an initial omnibus ANOVA (all Fs < 1.73, ps ≥ .20), and therefore analyses collapsed across this manipulation.

The proportions of correct "Yes" judgments were analyzed in a 2 (Test 1: studied-context vs. other-context) x 2 (Test 2: studied-context vs. other-context) x 2 (Study list: study list 1 vs. study list 2) mixed factorial ANOVA. As expected, there was a significant interaction between context on Test 1 and context on Test 2, $F(1, 22) = 53.99$, $MSE = .01$, $p < .0001$. This effect was moderated by a Test 1 Context x Test 2 Context x Study List interaction, $F(1, 22) = 10.07$, $MSE = .01$, $p = .004$. Despite the fact that there was an effect of study list (i.e., the sense with which the target words were studied did influence the judgment task), the pattern of judgments for both study list 1 and study list 2 was the same: A significant interaction between context on Test 1 and context on Test 2 was found for participants in the study list 1 group, $F(1, 11) = 6.98$, $MSE = .01$, $p = .02$, and for participants in the study list 2 group, $F(1, 11) = 73.65$, $MSE = .01$, $p < .0001$. Planned comparisons (collapsing across study list) showed that for items that had been cued with studied-context words on Test 1 participants less often remembered their previous recall when they were cued with the other-context words on Test 2 (studied/other) than when they were cued with studied-context words on Test 2 (studied/studied), $t(23) = 4.33$, $p < .001$. Conversely, for items that had been cued with other-context words on Test 1, participants less often remembered their prior recall when they were cued with studied-context words on Test 2 (other/studied) than when they were cued with other-context words on Test 2 (other/other), $t(23) = 4.65$, $p < .0001$.[3] Additionally, there was a significant main effect of Test 1 context, $F(1,22) = 11.43$, $MSE = .010$, $p = .003$, with correct memory for prior recall being greater for items that had been cued with studied-context words on Test 1 (M = .76) than for items that had been cued with other-context words on Test 1 (M = .70). This result is consistent with the possibility that correct recall during Test 1 was more complete for items cued with studied-context words than for those cued with other-context words. There was not a reliable main effect of context on Test 2 for correct "Yes" judgments, $F < 1$. Finally, as in Experiment 1, participants rarely falsely reported they had remembered items that had not been tested on Test 1.

Experiment 3

In Experiments 1 and 2, the context words were intended to cue distinctly different meanings of each target word. A participant who recalled "palm" on Test 1 and thought of it as part of a hand and who later recalled "palm" on Test 2 and thought of it as a tree would be correct, in a sense, to deny having previously recalled "palm" (tree) (see Martin, 1975, for an analogous argument regarding recognition failure paradigms that use homographs). Of course, it may be that real-world cases of the forgot-it-all-along effect, such as those described by Schooler (in press), also involve fundamental changes in the meaning of the remembered event (e.g., from fondling to rape). Does the forgot-it-all-along effect occur only when a past event is thought about in a dramatically different way, or can more subtle shifts also lead to forgetting of prior remembering? Experiment 3 was designed to explore this possibility.[4]

Experiment 3 was designed to test for the forgot-it-all-along effect by manipulating the context of the target items, but not their sense or meaning (e.g., "palm" always studied and tested as "part of hand," but in different contexts). For comparability across studies, we used the same target words as in Experiment 2, with two context sentences prepared for each target word. The design was otherwise similar to that used in Experiment 1.

Method

Participants. Twelve University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course.

Materials. The materials used in Experiment 2 were adapted by creating two context sentences for each target (e.g., "He swatted the fly with the palm of his hand" and "The fortune teller traced the lifeline on the palm of his hand"). Four of the items were used as primacy buffers and four as recency buffers. On both Test 1 and Test 2, the cues used to test for the target words were the context sentences. For Test 1, participants were cued with a studied- or other-context sentence containing a row of asterisk symbols for the target word and the first letter of the target word (e.g., "He swatted the fly with the *** of his hand" - "p"). On Test 2, participants were cued with studied-context sentences containing the first letter of the target word (e.g., "He swatted the fly with the p?? of his hand"). The target words were randomly

[3] It could be argued that, due to the reliable 3-way interaction of Test 1 Context x Test 2 Context x Study List, these planned comparisons should be carried out separately for the two study lists. Separate directional analyses demonstrated that the same overall patterns for the judgment task were also found for both study list groups. That is, for study list 1, participants less often remembered their previous recall of items cued with studied-context words on Test 1 when they were cued with other-context words on Test 2 (studied/other; M = .69) than when they were cued with studied-context words on Test 2 (studied/studied; M = .78), $t(11) = 1.92$, $p < .05$. Conversely, for study list 1, the items cued with other-context words on Test 1 were less often correctly judged as previously recalled when they were cued with studied-context words on Test 2 (other/studied; M = .63) than when they were cued with other-context words on Test 2 (other/other; M = .72), $t(11) = 2.03$, $p < .05$. For study list 2, participants were also less likely to remember their previous recall of target items cued with studied-context words on Test 1 if they were tested with other-context words on Test 2 (studied/other; M = .70) than if they were cued with studied-context words on Test 2 (studied/studied; M = .89), $t(11) = 4.55$, $p = .001$. For items that had been cued with other-context words on Test 1, study list 2 participants less often remembered their previous recall when they were cued with studied-context words on Test 2 (other/studied; M = .58) than when they were cued with other-context words on Test 2 (other/other; M = .85), $t(11) = 5.04$, $p < .0001$.

[4] We thank Bruce W. A. Whittlesea for raising this issue.

divided into three lists of 36 words (*test list* factor), with each list appearing equally often across participants in the studied-context, other-context, and not-tested within-subject conditions of Test 1. Two study lists were constructed (*study list* factor) to counterbalance the contexts of studied words between subjects.

Procedure. The procedure was basically the same as in Experiment 1, but it was adjusted to include three of the modifications from Experiment 2: (a) auditory study, (b) tape recording of Test 1 responses, and ©) changing the symbols of the to-be-recalled target word from "*" on Test 1 to "?" on Test 2.

Results and Discussion

Recall performance. Proportion correctly recalled on Test 1 was significantly higher for items in the studied-context condition (M = .94) than items in the other-context-condition (M = .81), $t(11) = 5.02$, $p < .001$. A within-subjects ANOVA was performed on the proportion correctly recalled on Test 2, with context on Test 1 (studied, other, or not-tested) as the within-subjects factor. There was no significant difference in proportion of items recalled for the studied-context (M = .96), other-context (M = .94), and not-tested (M = .93) conditions, $F < 1$.

Judgment of previous recollection. The proportions of items on Test 2 correctly judged as recalled on Test 1 are shown in Table 4. The analyses reported here were performed on the judgment data for target items correctly recalled on both Test 1 and Test 2 (shown in bold in Table 4), although the same pattern of results was found when analyses were contingent only on correct recall on Test 1. No reliable effects of the study list or test list factors for the judgment task were found in an initial omnibus ANOVA (all Fs < 2.17, ps ≥ .20), and therefore the data were collapsed across these variables. A within-subjects ANOVA was performed on the proportion of "Yes" judgments, with context on Test 1 (studied vs. other) as the within-subjects factor. Participants were significantly more likely to forget that they had recalled an item on Test 1 if it had been cued with the other-context sentence on Test 1 than if it had been cued with the studied-context sentence, $F(1, 11) = 48.54$, $MSE = .01$, $p < .0001$. Participants rarely erred on not-tested items by saying that they had remembered those items on Test 1.

Experiment 4

The results of Experiments 1 through 3 support the hypothesis that changes in the way an individual thinks about an event when it is recalled on different occasions can lead to forgetting of the prior recollections. A critic might argue, however, that participants in our experiments based their judgments of prior recollection not on whether they remembered recalling a target item on Test 1, but rather on whether they remembered encountering the Test 2 retrieval cue on Test 1. We explicitly and emphatically instructed participants that their judgments should be based on whether they remembered recalling the target, rather than on whether they remembered encountering the cue, but it is possible that participants did not always follow this instruction. We eliminated this alternative explanation in Experiment 4 by changing Test 1 from cued to free recall (with the assumption that during free recall participants would tend to think about the words in the way they had been biased to do during study). As in Experiments 1-3, Test 2 was cued recall, with some items cued by studied-context words and others by other-context words. Because the context cues were not encountered until Test 2, participants could not possibly base their judgments of prior recollection on whether or not they remembered encountering those cues on Test 1. We expected that the FIA effect would be smaller in this procedure than in Experiments 1-3, because free recall is relatively memorable (e.g., the items successfully recalled are the ones that participants are able to bring to mind on their own) and because the familiarity of the cues themselves could not contribute to the FIA effect.

Participants studied 5 lists of 16 homographic target words presented in disambiguating context sentences. After each study list was presented, participants attempted free recall for the target words (Test 1). To explore the relationship between the phenomenology of Test 1 recall and subsequent judgments of prior remembering participants were asked to make a "Remember -Know" judgment for each recalled word (e.g., Gardiner & Java, 1990). Later, participants were given a cue to recall each of the target words (Test 2): half of the items were cued with context words relating to the studied sense of the targets (*studied-context* condition) and half were cued with context words that did not relate in sense to the sentences presented with the target items at study (*other-context* condition). Additionally, the second test required participants to judge whether they remembered recalling the word on the free-recall test.

Method

Participants. Twenty seven University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course. Three participants recalled fewer than eight items in either the studied-context or other-context conditions on the cued recall test that they had also previously recalled in free recall; their data were dropped from the analyses.

Materials. A list of 90 target words was constructed from the set of homographic target words used in the previous three experiments. Ten of the items were used as a practice list. Two study lists were constructed to counterbalance the meanings of the target words between subjects (*study list* factor). For each participant, the computer randomly divided the 80 target words into five lists of 16 items. The target words were randomly partitioned into two lists of 40 words (*test list* factor), with each list appearing equally often across participants in the studied-context and other-context within-subject conditions of Test 2. The two context words associated with each target word were the same as those used in the first two experiments, but the sentences in which the target items were presented during study were modified. In Experiments 1 and 2, the sentences in which the target words were presented during the study phase always included both the target word and one of its two context words. In this experiment, the senses of the study sentences were not changed from those used in the first two experiments, but the context words themselves were dropped from the sentences. For example, in Experiments 1 and 2 the target word "palm" was sometimes studied with the sentence "He used the palm of his hand to swat the fly," and recall of "palm" was sometimes cued with the context word "hand." In this experiment, some subjects studied "palm" with the sentence "He used his palm to swat the fly," and in Test 2 "palm" was cued with either "hand" or "tree." Therefore, in both the studied-context and other-context conditions, participants never heard the context words until the final cued-recall test.

Procedure. Participants were told that on each study trial they would be read a sentence, and that one of the words would be verbally emphasized as a target word. Their task was to repeat the sentence aloud, verbally emphasizing the target word. In the event that a participant did not repeat a sentence or failed to emphasize the target word, the experimenter repeated the sentence (although participants were told it was important to pay close attention to each sentence so that it would not need to be repeated, and participants almost always succeeded). Participants were informed that after each list of 16 sentences was presented, they would be required to recall the target words. The experimenter stressed to participants that they should write down a word only if they were confident that they heard that item as a target word in the study list. Participants were also required to make a Remember-Know judgment for each target word they recalled. Participants were instructed to write an "R" beside the target word if they could actually recollect something about the experience of having studied that target word, and to place a "K" beside the target word if they knew the item was in the study list but could not recollect any specific details of their encounter with it (as in Gardiner & Java, 1990).

For each list, participants were given 1.5-min to write down the target words and corresponding Remember-Know judgments. Participants were given the practice list and performed the free-recall task for that list to ensure that they understood the instructions and were comfortable with the tasks. They then completed the five study/recall cycles. Thereafter, participants were given a 10-min break during which they conversed with the experimenter before moving on to the final cued-recall test.

All 80 of the critical target words were tested for in random order with cued recall on Test 2. On each trial, a context word and the first and last letters of a target word were presented on the computer screen, and participants were asked to recall the corresponding studied target word. Participants were told that for half of the trials the context word would be closely related to the sentence in which the to-be-recalled word had been studied, and that for the other half of the trials the context word would be different from the sentence presented with the target during study but that the context word would nonetheless be related to the target word. To clarify these instructions, participants were shown an example of a studied- and other-context word that could be used to test for a target word (e.g., "You studied 'palm' with 'He used his palm to swat the fly.' We could test you for this target word by presenting the related context word 'hand' or by presenting the context word 'tree,' which is not related to the sentence with which you studied "palm" but is related to the target 'palm'"). Participants were instructed to respond with an answer only if they remembered hearing that target word during study. After each item was recalled (or, if recall failed, the experimenter announced the target word) participants judged whether they remembered having recalled the target word during Test 1. Participants were told that they should respond with a "Yes" only if they remembered recalling and writing down the target word during the first phase of the experiment.

Results and Discussion

Recall performance. Participants sometimes recalled a word from a list other than that on which they were being tested (across participants, an average of .92 items were recalled from a list other than the one being tested), and these items were scored as correct so long as they were not items that had been recalled on an earlier test. Overall, participants recalled an average of 36.75 target words in free recall (45.94%). On average, participants classified 71.54% of recalled words as "remember." On the final test, proportion correct cued recall was significantly higher for items in the studied-context condition ($M = .88$) than for items in the other-context condition ($M = .81$), $t(23) = 3.75$, $p = .001$.

Judgment of previous recollection. The proportions of items correctly judged following cued recall (Test 2) as having also been recalled during free recall (Test 1) are shown in Table 5. The analyses reported here were performed on the judgment data for target items correctly recalled in both free and cued recall (shown in bold in Table 5), although the same pattern of results was found when analyses were contingent only on correct recall on Test 1. No reliable effects of the study list or test list factors for the judgment task were found in an initial omnibus ANOVA (all Fs < 1.49, ps ≥ .24), and therefore the data were collapsed across these variables. A within-subjects ANOVA was performed on the proportion of correct "Yes" judgments, with context in cued recall (studied vs. other) as the within-subjects factor. Participants were significantly more likely to fail to remember that they had freely recalled an item on Test 1 if it had been cued with the other-context word on Test 2 than if it had been cued with the studied-context word, $F(1, 23) = 8.08$, $MSE = .01$, $p < .01$. Participants rarely erred on items that they had not freely recalled on Test 1.

We conducted a subanalysis restricted to items on which participants had made a "Remember" (as opposed to "Know") response during Test 1. The data from 7 participants were excluded because they claimed to "remember" fewer than eight items per condition. Among the remaining participants, the effect of context cues on judgments of prior recall was the same as in the overall analysis reported above: Participants more often forgot that they had recalled "remembered" items if the items were cued with the other-context word on Test 2 ($M = .81$) than if they were cued with the studied-context word ($M = .91$), $F(1, 16) = 7.66$, $MSE = .01$, $p = .01$.

General Discussion

As predicted, participants more often forgot that they had previously recalled a studied word if they were led to think about that word differently on the two recall episodes. Thus our results support the existence of a forgot-it-all-along effect.

Remembering prior remembering is likely similar to remembering other sorts of past experiences. That is, an episode of recollection may be remembered (or forgotten) just as other sorts of episodes are. Similarly, judgments of prior recollection are likely based on the same mechanisms as judgments of other sorts of prior occurrences. In general terms, when attempting to judge whether or not they have previously recollected a particular event, individuals cue memory and assess its output (a la Tulving's 1984 concept of synergistic ecphory; cf. Anderson & Bower, 1972; Johnson, Hashtroudi, & Lindsay, 1993; Whittlesea & Williams, 1998).

Memories of episodes of recollection likely have two special (although not unique) characteristics. One is that they share content with memories of the remembered event itself. The more vividly, completely, and veridically a past experience was recollected on a particular occasion, the more the memories of that episode of recollection will share content with memories of the initial experience itself. This may pose a variety of problems for the cognitive system. For example, instances of prior recollection may be difficult to recall as distinct episodes because cues for those prior recollections will also be cues for (and perhaps better cues for) memories of the initial event itself; this may limit revival of the memory information for the prior recollection via cue-overload (Watkins & Watkins, 1976), or produce blended ecphoric products in which the information from prior recollections is experienced as part of the recollection of the event itself (indeed, this may be an important part of the way rehearsal works). If cuing conditions selectively favor revival of one or more prior instances of recollection over revival of the event itself, that memory information may be mistaken as a memory of a perceptual experience (i.e., the individual thinks s/he is remembering an actual experience but is really reviving memories of prior recollections of that experience rather than memories of the event itself); in other cases in which cuing conditions selectively favor revival of memories of prior instances of recollection, the individual may mistakenly judge that s/he never experienced the event in question but rather had only thought about or imagined experiencing it (cf. Johnson et al., 1993).

The mechanisms governing forgetting prior episodes of word recall are not, of course, necessarily the same as those involved in forgetting prior episodes of recalling childhood abuse. For example, when participants in our experiments recalled a word on Test 2 and failed to remember that they had also previously recalled that word, they probably did not have emotionally charged "recovered-memory experiences" akin to those reported by Schooler (in press). That is, with our materials, participants would rarely spontaneously think something like "Wow, this is amazing–I had totally forgotten about the word 'palm' until now!" Principles of memory developed in the laboratory have, for the most part, fared well in terms of generalizing to more naturalistic settings (e.g., Banaji & Crowder, 1989), but the question of generalizability is an empirical one. At minimum, it is likely that forgetting of prior instances of recalling abuse is a more complex and multifaceted phenomenon than is forgetting of prior instances of recalling study-list words. Nonetheless, the fact that in all four experiments a substantial forgot-it-all-along effect was obtained is consistent with the hypothesis that changes in the way an event is thought about on different occasions can contribute to forgetting of prior episodes of recollection.

Although the idea for the present experiments grew out of the recovered memory arena, our findings also have implications for other domains within cognitive psychology. Research on a variety of memory phenomena (e.g., flashbulb memories, permastore) often uses retrospective self-report measures of the number of times a person previously remembered or thought about an event (e.g., to assess the effects of rehearsal). The results presented here suggest that such measures should not be taken at face value and that, like memory for an event itself, memory for previous recollection is subject to systematic biases.

References

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

Azuma, T. (1996). Familiarity and relatedness of word meanings: Ratings for 110 homographs. *Behavior Research Methods, Instruments and Computers, 28*, 109-124.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist, 44*, 1185-1193.

Bouton, M. E., Nelson, J. B., & Rosas, J. M. (1999). Stimulus generation, context change, and forgetting. *Psychological Bulletin, 125*, 171-186.

Eich, E. (1995). Searching for mood dependent memory. *Psychological Science, 6*, 67-75.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 349-358.

Franks, J. L., Bilbrey, C. W., Lien, K. G., McNamara, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory & Cognition, 28*, 1140-1151.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23-30.

Gawlick-Grendell, L. A., & Woltz, D. J. (1994). Meaning dominance norms for 120 homographs. *Behavior Research Methods, Instruments and Computers, 26*, 5-25.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Joslyn, S., Loftus, E. F., McNoughton, A., & Powers, J. (in press). Memory for memory. *Memory & Cognition*.

Light, L. L., & Carter-Soboll, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 1-11.

Lindsay, D. S., & Read, J. D. (in press). The recovered memories controversy: Where do we go from here? In G. M. Davies & T. Dalgleish (Eds.), *Recovered memories: Seeking the middle ground*. Chichester, UK: John Wiley & Sons.

Martin, E. (1975). Generation-recognition theory and the encoding specificity principle. *Psychological Review, 82*, 150-153.

Padilla-Walker, L. M., & Poole, D. A. (in press). Memory for previous recall: A comparison of free and cued recall. *Applied Cognitive Psychology*.

Parks, T. E. (1999). On one aspect of the evidence for recovered memories. *American Journal of Psychology, 112*, 365-370.

Schneider, W. (1988). Micro-Experimental Laboratory: An integrated system for IBM PC Compatibles. *Behavior Research Methods, Instruments, and Computers, 20*, 206-217.

Schooler, J. W. (1999). Seeking the core: The issues and evidence surrounding recovered accounts of sexual trauma. In L. M. Williams (Ed.), *Trauma and memory* (pp. 203-216). Thousand Oaks, CA, USA: Sage Publications.

Schooler, J. W. (in press). Discovering memories in light of meta-consciousness. *The Journal of Aggression, Maltreatment and Trauma*.

Schooler, J. W., Ambadar, Z., & Bendiksen, M. (1997). A cognitive corroborative case study approach for investigating discovered memories of sexual abuse. In J. D. Read & D. S. Lindsay (Eds.), *Recollections of trauma: Scientific evidence and clinical practice* (pp. 379-387). New York: Plenum Press.

Schooler, J. W., Bendiksen, M., & Ambadar, Z. (1997). Taking the middle line: Can we accommodate both fabricated and recovered memories of sexual abuse? In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 251-292). New York: Oxford University Press.

Tulving, E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences, 7*, 223-268.

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

Tulving, E., & Watkins, O. C. (1977). Recognition failure of words with a single meaning. *Memory & Cognition, 5*, 513-522.

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition, 22*, 111-126.

Watkins, M. J., & Watkins, O. C. (1976). Cue-overload theory and the method of interpolated attributes. *Bulletin of the Psychonomics Society, 7*, 289-291.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98*, 141-165.

Table 1

Mean Number of Items and Mean Proportion of Items Judged as "Recalled" as a Function of Recall Status on Test 1 and Test 2 for Experiment 1

Test 1 Cue	Test 1/Test 2 Recall Status	Number of Items	Proportion Judged as "Recalled" on Test 1
Studied-context			
	Not Recalled/Not Recalled	1.79	.09 (.06)
	Not Recalled/Recalled	1.37	.04 (.02)
	Recalled/Recalled	**31.42**	**.81 (.02)**
	Recalled/Not Recalled	.42	.57 (.17)
Other-context			
	Not Recalled/Not Recalled	1.62	0 (--)
	Not Recalled/Recalled	6.38	.11 (.03)
	Recalled/Recalled	**25.83**	**.54 (.04)**
	Recalled/Not Recalled	1.17	.10 (.05)
Not-tested			
	NA/Not Recalled	2.79	.14 (.06)
	NA/Recalled	32.21	.10 (.03)

Note. Lines in bold are those for which statistical analyses are reported in the manuscript. There were 35 items per condition. Standard error of the means are in parentheses.

Table 2

Mean Proportion of Items Correctly Recalled in Experiment 2

Test1/Test 2 Cues	Test 1	Test 2
studied/studied	.89 (.02)	.91 (.02)
other/studied	.77 (.02)	.91 (.02)
not-tested/studied		.89 (.02)
studied/other	.88 (.02)	.82 (.03)
other/other	.77 (.03)	.82 (.02)
not-tested/other		.77 (.03)

Note. Standard error of the means are in parentheses.

Table 3

Mean Number of Items and Mean Proportion of Items Judged as "Recalled" as a Function of Recall Status on Test 1 and Test 2 for Experiment 2

Test 1/Test 2 Cues	Test 1/Test 2 Recall Status	Number of Items	Proportion Judged as "Recalled" on Test 1
Studied/Studied			
	Not Recalled/Not Recalled	1.33	.07 (.06)
	Not Recalled/Recalled	.71	.25 (.13)
	Recalled/Recalled	**15.75**	**.83 (.03)**
	Recalled/Not Recalled	.21	.38 (.24)
Other/Studied			
	Not Recalled/Not Recalled	.71	.03 (.03)
	Not Recalled/Recalled	3.42	.03 (.01)
	Recalled/Recalled	**12.92**	**.60 (.03)**
	Recalled/Not Recalled	.96	.27 (.11)
Not-tested/Studied			
	NA/Not Recalled	2.04	.02 (.02)
	NA/Recalled	15.96	.08 (.02)
Studied/Other			
	Not Recalled/Not Recalled	.83	.21 (.11)
	Not Recalled/Recalled	1.42	.03 (.03)
	Recalled/Recalled	**13.46**	**.70 (.03)**
	Recalled/Not Recalled	2.29	.57 (.10)
Other/Other			
	Not Recalled/Not Recalled	2.67	.02 (.01)
	Not Recalled/Recalled	1.54	.07 (.04)
	Recalled/Recalled	**13.25**	**.79 (.03)**
	Recalled/Not Recalled	.54	.45 (.16)
Not-tested/Other			
	NA/Not Recalled	4.17	.05 (.03)
	NA/Recalled	13.83	.08 (.02)

Note. Lines in bold are those for which statistical analyses are reported in the manuscript. There were 18 items per condition. Standard error of the means are in parentheses.

Table 4

Mean Number of Items and Mean Proportion of Items Judged as "Recalled" as a Function of Recall Status on Test 1 and Test 2 for Experiment 3

Test 1 Cue	Test 1/Test 2 Recall Status	Number of Items	Proportion Judged as "Recalled" on Test 1
Studied-context			
	Not Recalled/Not Recalled	1.42	0 (--)
	Not Recalled/Recalled	.83	.05 (.05)
	Recalled/Recalled	**33.58**	**.93 (.01)**
	Recalled/Not Recalled	.17	.50 (--)
Other-context			
	Not Recalled/Not Recalled	.75	0 (--)
	Not Recalled/Recalled	6.25	.06 (.03)
	Recalled/Recalled	**27.67**	**.63 (.05)**
	Recalled/Not Recalled	1.33	.18 (.12)
Not-tested			
	NA/Not Recalled	2.58	0 (--)
	NA/Recalled	33.42	.03 (.01)

Note. Lines in bold are those for which statistical analyses are reported in the manuscript. There were 36 items per condition. Standard error of the means are in parentheses.

Table 5

Mean Number of Items and Mean Proportion of Items Judged as "Recalled" as a Function of Recall Status in Test 1 and Test 2 for Experiment 4

Test 2 Cue	Test 1/Test 2 Recall Status	Number of Items	Proportion Judged as "Recalled" on Test 1
Studied-context			
	Not Recalled/Not Recalled	3.33	.08 (.05)
	Not Recalled/Recalled	17.87	.07 (.01)
	Recalled/Recalled	**17.29**	**.86 (.03)**
	Recalled/Not Recalled	1.50	.63 (.10)
Other-context			
	Not Recalled/Not Recalled	5.25	.03 (.02)
	Not Recalled/Recalled	16.79	.06 (.01)
	Recalled/Recalled	**15.42**	**.77 (.03)**
	Recalled/Not Recalled	2.54	.56 (.07)

Note. Lines in bold are those for which statistical analyses are reported in the manuscript. There were 40 items per condition in Test 2. Standard error of the means are in parentheses.