Have I Remembered this Before? Remembering (and forgetting) of prior Remembering

by

Michelle Marie Arnold
B.A., University of Lethbridge, 1997

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

We accept this thesis as conforming to the required standards

[redacted]

Dr. D. Stephen Lindsay, Supervisor (Department of Psychology)

[redacted]

Dr. Michael E. J. Masson, Departmental Member (Department of Psychology)

[redacted]

Dr. Helena Kadlec, Departmental Member (Department of Psychology)

[redacted]

Dr. Geraldine Van Gyn, External Examiner (School of Physical Education)

Supervisor: Dr. D. Stephen Lindsay

Abstract

Schooler, Bendiksen, and Ambadar (1997) described cases of "recovered memories" of abuse in which individuals had forgotten episodes of talking about the abuse when they were supposedly amnestic for it. We developed a laboratory analogue of this "forgot-it-all-along" effect. 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. Experiments 3, 5, and 6 were similar to Experiment 1, but the cues always 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 forget prior remembering when they had been cued to think of the items differently on the two tests.

Examiners:

[redacted]

Dr. D. Stephen Lindsay, Supervisor (Department of Psychology)

[redacted]

Dr. Michael E. J. Masson, Departmental Member (Department of Psychology)

[redacted]

Dr. Helena Kadlec, Departmental Member (Department of Psychology)

[redacted]

Dr. Geraldine Van Gyn, External Examiner (School of Physical Education)

Table of Contents

List of Tables

List of Figures

Acknowledgments

I wish to thank my supervisor Steve Lindsay for introducing me to the topic of the forgot-it-all-along effect, and for all of his guidance and contributions in constructing, completing, and interpreting my thesis research. I give special thanks to Vincenza Gruppuso, Tanya Berry, Iris van Rooij, and Denise Broekman for helpful discussions of my work, and for general support and encouragement. Finally, I would like to thank my family because even though they admit they don't really understand what I do (and why I love it!), they support me nonetheless.

Have I Remembered this Before? Remembering

(and forgetting) of prior Remembering

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.

A number of survey studies have asked adults who reported histories of childhood sexual abuse if there had been a time when they did not remember that abuse (see Read & Lindsay, 2000, and Scheflin & Brown, 1996, for contrasting reviews). The rate of reports of prior periods of not remembering abuse vary dramatically across such studies (from a low of 16% in Williams, 1995, to a high of 77% in Roe & Schwartz, 1996), likely due to differences in samples and in how the question was phrased. Although sometimes cited as evidence of a special traumagenic amnesia mechanism, such studies pose numerous problems for interpretation (e.g., when respondents reply affirmatively, do they mean that they were amnestic or merely that they avoided thinking about the abuse? see Melchert & Parker, 1997, and Williams, 1995). Even if it was clear that participants were reporting prior periods of amnesia, important questions would remain. How do individuals go about judging whether or not they previously remembered something that they currently remember? How accurate are such judgments? As Schooler (in press) noted, judgments of prior remembering may be analogous to other sorts of metamemory judgments (e.g., feeling-of-knowing; Koriat & Levy Sadot, 1999; Nelson & Narens, 1994) and as such constitute an important topic for investigation (Nelson, 1996).

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 (i.e., emotionally wrenching experiences of recollecting long-forgotten memories of sexual abuse that contradict the person's beliefs about his/her past). 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. For example, at the age of 24 a woman referred to as TW (Schooler, in press) had a recovered-memory experience of molestation (fondled by a family friend) that had occurred when she was nine. TW told her mother about the molestation shortly after it occurred, but TW believes that between the time she told her mother about it to the time she remembered it at the age of 24 she had absolutely no memory for the molestation. TW was subsequently shocked to learn that, during the period of time she believed she had been amnesic for the event she had told her former husband about it on several different occasions (Schooler, Bendiksen, & Ambadar, 1997).

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 (FIA) effect. That is, recollecting an event in manner X may cause one to forget having previously recollected it in manner Y. In the case of TW, there is some evidence to suggest that at the time the molestation occurred

Have I Remembered this Before? Remembering (and forgetting) of prior Remembering

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.

A number of survey studies have asked adults who reported histories of childhood sexual abuse if there had been a time when they did not remember that abuse (see Read & Lindsay, 2000, and Scheflin & Brown, 1996, for contrasting reviews). The rate of reports of prior periods of not remembering abuse vary dramatically across such studies (from a low of 16% in Williams, 1995, to a high of 77% in Roe & Schwartz, 1996), likely due to differences in samples and in how the question was phrased. Although sometimes cited as evidence of a special traumagenic amnesia mechanism, such studies pose numerous problems for interpretation (e.g., when respondents reply affirmatively, do they mean that they were amnestic or merely that they avoided thinking about the abuse? see Melchert & Parker, 1997, and Williams, 1995). Even if it was clear that participants were reporting prior periods of amnesia, important questions would remain. How do individuals go about judging whether or not they previously remembered something that they currently remember? How accurate are such judgments? As Schooler (in press) noted, judgments of prior remembering may be analogous to other sorts of metamemory judgments (e.g., feeling-of-knowing; Koriat & Levy Sadot, 1999; Nelson & Narens, 1994) and as such constitute an important topic for investigation (Nelson, 1996).

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 (i.e., emotionally wrenching experiences of recollecting long-forgotten memories of sexual abuse that contradict the person's beliefs about his/her past). 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. For example, at the age of 24 a woman referred to as TW (Schooler, in press) had a recovered-memory experience of molestation (fondled by a family friend) that had occurred when she was nine. TW told her mother about the molestation shortly after it occurred, but TW believes that between the time she told her mother about it to the time she remembered it at the age of 24 she had absolutely no memory for the molestation. TW was subsequently shocked to learn that, during the period of time she believed she had been amnesic for the event she had told her former husband about it on several different occasions (Schooler, Bendiksen, & Ambadar, 1997).

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 (FIA) effect. That is, recollecting an event in manner X may cause one to forget having previously recollected it in manner Y. In the case of TW, there is some evidence to suggest that at the time the molestation occurred (age 9) TW viewed the event as "uncomfortable" or "awkward," but that when she experienced the recovered-memory experience (age 24) she labeled the event as sexual abuse/assault. Further, TW's former husband stated that her discussions about the molestation (during the period she believes she had no memory for the event) were very "matter of fact," with little emotion expressed. In contrast, TW's recovered-memory experience involved a vivid and powerfully emotional memory of the molestation–a phenomenological experience vastly different from her previous remembering of the event (Schooler, Bendiksen, & Ambadar, 1997).

Schooler named the forgot-it-all-along effect in reference to another well-known phenomenon, the knew-it-all-along effect. Several studies have shown that people often have difficulty retrospectively determining the level of knowledge they had prior to the occurrence of an event (e.g., Begg, Robertson, Gruppuso, Anas, & Needham, 1996; Fischhoff, 1975; Wood, 1978). Fischhoff (1977) coined the phrase the knew-it-all-along effect to describe this result, although it also commonly referred to as hindsight bias and as the illusory-knowledge effect. In the basic knew-it-all-along paradigm participants respond to a set of questions, after which they are given feedback (i.e., the correct answers) for a portion of the questions; participants are then given the questions and instructed to respond with the same answers they had given prior to being exposed to the feedback. The knew-it-all-along effect occurs when participants correctly answer significantly more feedback questions (in comparison to non-feedback questions), indicating an overestimation in the amount of knowledge they believe they previously possessed (Fischhoff, 1977).

Several studies have explored the effects of extended efforts to retrieve memories of childhood on adults' assessments of the completeness of their autobiographical memories. Research by Belli, Winkielman, Read, Schwarz, and Lynn (1998) demonstrated effects of experimental manipulations on individuals' assessments of the completeness of their memories of childhood in general: Individuals asked to recollect numerous childhood experiences during a brief experiment tended to rate their memories of childhood as being less complete than individuals asked to recollect only a few childhood experiences. Presumably the former participants were surprised at the difficulty of recollecting the required number of childhood experiences, which led them to revise downward their estimates of the completeness of their memories of childhood. Read (1997) and Read and Lindsay (2000) found that having participants engage in extended efforts to remember more about a particular past experience (e.g., their high-school graduation ceremony) increased the likelihood that they would report prior periods during which they remembered less about the experience than they currently do (presumably simply because they had, over the course of the experiment, come to recollect many long-forgotten details). Although fascinating, these studies do not get at the specific issue of how individuals judge whether or not they had previously recollected a particular past event that they currently recollect.

To the best of my 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 events. Parks found evidence that participants had often forgotten their Phase-1 reminiscences a few minutes later, when making the Phase-2 judgments. Relatedly, a study by Joslyn, Loftus, McNoughton, and Powers (in press) required participants to study a passage (Phase 1), answer questions about the initial session 24-36 hrs later (Phase 2), and answer those same questions six weeks later (Phase 3). During Phase 3, participants were also asked to indicate whether or not they had remembered particular answers during Phase 2. Participants sometimes forgot that they had previously recalled an answer, and often falsely claimed to have remembered answers that they had not recalled during Phase 2. Finally, in a 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 context 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. These pioneering studies demonstrate that, as one might expect, people do sometimes forget prior occasions of recollection. They also suggest various mechanisms that may influence such forgetting, but they do not directly explore Schooler's hypothetical account of the forgot-it-all-along effect.

We developed a laboratory analog of the process that Schooler (in press) proposed gives rise to the forgot-it-all-along effect (i.e., reinterpretation). It is possible that the underlying cognitive processes and mechanisms involved in recollecting childhood memories of abuse are not equivalent to the processes and mechanisms involved in recalling events created in a laboratory setting. However, if a change in the way an event is thought about on different occasions does lead to forgetting for previous episodes of recalling abuse, then a change in context should also result in the forgetting of previous recollection within the laboratory setting. 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 forgot-it-all-along effect by manipulating the meaning of studied homographs between the two recall attempts. The other three experiments (Experiment 3, 5, and 6) tested for the forgot-it-all-along 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 context change is a factor in 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 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.

A more detailed comparison of the forgot-it-all-along effect with related memory phenomena–such as recognition failure and the knew-it-all-along effect–has been relegated to the discussion section so as to supply a groundwork for comparing and contrasting the various memory effects. Additionally, the theoretical interpretations presented in the discussion section for the following six experiments will emphasize an attribution approach to the forgot-it-all-along effect–that is, an approach that incorporates a variety of factors that may potentially influence the reconstruction of prior episodes (e.g., encoding specificity, fluency, etc.).

Experiment 1

Method

Overview. 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.

Subjects. Twenty six University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course. Two of these 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 (v. 2.1) 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 in place of 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 105 target 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 know 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

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

Recall performance. The proportions correctly recalled on Test 1 were initially analyzed in a 2 (Test 1: studied-context vs. other-context) x 2 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. Test 1 recall performance was not reliably affected by the counterbalancing factors of study list or test list (all Fs < 1.11, ps $\geq .35$), so the data were collapsed across these manipulations and a t-test was conducted. Proportion correctly recalled on Test 1 was significantly higher for items cued with studied-context words ($M = .91$, $SD = .07$) than for items cued with the other-context words ($M = .77$, $SD = .11$), $t(23) = 9.62$, $p < .0001$.

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 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. Test 2 recall performance was not reliably affected by the counterbalancing factors of study list or test list (all Fs < 2.97, ps $\geq .10$) and the data were collapsed across these manipulations. Performance on Test 2 was near ceiling (as intended), and there was no significant difference in proportion of items recalled for the studied-context (M = .94, SD = .06), other-context (M = .92, SD = .07), and not-tested (M = .92, SD = .08) 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 effects of the counterbalancing factors of study list or test list were found (all $Fs < 1$), 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.

Experiment 2

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.[1] 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 was designed to assess these two accounts of the results of Experiment 1 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.

Method

Overview. 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), (c) 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).

Subjects. Twenty seven University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course or a $10 payment. Three of these 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

Recall performance. The proportion correctly recalled for each condition of Test 1 and Test 2 is shown in Table 2. The proportion of items correctly recalled on Test 1 were analyzed in a 2 (Test 1: studied-context and other-context) x 2 (Study list: study list 1 vs. study list 2) x 6 (Test list: test lists 1-6) mixed-models ANOVA. There was no effect of the test list counterbalancing factor on Test 1 recall, (all Fs < 1.90, ps $\geq .17$), and the data were subsequently collapsed across this variable. There was a marginal main effect of study list, $F(1, 22) = 4.05$, $MSE = .01$, $p = .06$. Collapsing across study list, proportion correctly recalled on Test 1 was higher for items cued with the studied-context words ($M = .88$, $SD = .07$) than with other-context words ($M = .77$, $SD = .09$), $F(1, 23) = 57.76$, $MSE = .003$ $p < .0001$. This same pattern is found for participants who studied list 1, $F(1, 11) = 22.68$, $MSE = .002$, $p = .001$, and for participants who studied list 2, $F(1, 11) = 38.48$, $MSE = .003$, $p < .0001$, but the participants who studied list 1 were significantly more likely to recall target words cued with different context words ($M = .81$, $SD = .09$) than the participants who studied list 2 ($M = .73$, $SD = .08$), $t(22) = 2.34$, $p = .03$. Thus Test 1 recall for both of the study lists adhered to the principle of encoding specificity, but due to a fluke of random assignment the items in the other-context condition were recalled at a higher rate in the first study list than in the second study list.

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) x 2 (Study list: study list 1 vs. study list 2) x 6 (Test list: test lists 1-6) mixed-models ANOVA. Overall, correct recall on Test 2 was higher for items cued with the studied-context words (M = .90, SD = .06) than with the other-context words (M = .80, SD = .12), $F(1, 12) = 43.57$, $MSE = .01$, $p < .0001$. This effect was moderated by a reliable Test 2 context x Study List interaction, $F(1, 12) = 5.15$, $MSE = .01$, $p = .04$, and a reliable Test 2 context x Test List interaction $F(5, 12) = 3.40$, $MSE = .01$, $p = .04$. However, when the two conditions not-tested on Test 1 are dropped from the analysis no effects of the study list or test list factors are found (all Fs < 2.73, ps ≥ .12). Therefore, it appears as though some of the counterbalancing lists were more difficult to recall on Test 2 for the items not-tested on Test 1 (e.g., there was a large fluctuation of recall performance for the not-studied/other condition across the six test lists, with recall ranging from .60 to .88), but this pattern is not discussed further because the focus of the experiment is the judgment of prior recall. Excluding the two conditions not-tested on Test 1 from the analysis, and collapsing across the study list and test list factors, correct recall on Test 2 was significantly higher for items cued on Test 2 with studied context words (M = .91, SD = .05) than with different context words (M = .82, SD = .11), $F(1, 23) = 24.93$, $MSE = .01$, $p < .0001$. With all of the conditions included in the analysis, there was a trend for correct recall on Test 2 to be influenced by context on Test 1, $F(2,24) = 3.08$, $MSE = .01$, $p = .06$, however this pattern disappeared when the two conditions not-tested on Test 1 were dropped from the analysis, $F < 1$. That is, results indicated that items not-tested on Test 1 were recalled at a lower rate on Test 2 than the items (regardless of studied- or other-context) that had been tested for on Test 1. 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. In an initial omnibus ANOVA no effect of the test list counterbalancing factor for the judgment task was found (all $Fs < 1.73$, $ps \geq .20$), 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 similar: 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), $\underline{t}(23) = 4.65$, $\underline{p} < .0001$.[2] Additionally, there was a significant main effect of Test 1 context, $\underline{F}(1,22) = 11.43$, $\underline{MSE} = .010$, $\underline{p} = .003$, with correct memory for prior recall being greater for items that had been cued with studied-context words on Test 1 ($\underline{M} = .76$, $\underline{SD} = .11$) than for items that had been cued with other-context words on Test 1 ($\underline{M} = .70$, $\underline{SD} = .12$). 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, $\underline{F} < 1$. 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 (e.g., "palm" as a kind of tree or as a part of the hand). To the extent that the manipulation worked, what participants recalled on Test 2 was quite different from what they recalled on Test 1; that is, a subject 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). 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, such that there is a sense in which the person is correct in denying past recollections of what they now remember. Nonetheless, it may also be the case that more subtle shifts in context can lead to a forgot-it-all-along effect. Experiment 3 was designed to explore this possibility.

Method

Overview. 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.

Subjects. 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 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 (c) changing the symbols of the to-be-recalled target word on Test 1 (*) and Test 2 (?).

Results

Recall performance. The proportions correctly recalled on Test 1 were initially analyzed in a 2 (Test 1: studied-context vs. other-context) x 2 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. Test 1 recall performance was not reliably affected by the counterbalancing factors of study list or test list (all Fs < 4.28, ps $\geq .07$), so the data were collapsed across these manipulations and a t-test was conducted. Proportion correctly recalled on Test 1 was significantly higher for items in the studied-context condition ($M = .94$, $SD = .06$) than items in the other-context-condition ($M = .81$, $SD = .08$), $t(11) = 5.02$, $p < .001$.

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 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. There was no significant difference in proportion of items recalled on Test 2 for the studied-context ($M = .96$, $SD = .05$), other-context ($M = .94$, $SD = .07$), and not-tested ($M = .93$, $SD = .08$) conditions, $F < 1$. There were reliable interactions of Test 1 Context x Study List x Test List, $F(4, 12) = 3.95$, $MSE = .002$, $p = .03$, and Test 1 Context x Test List, $F(4,12) = 5.58$, $MSE = .002$, $p = .01$, as well as a marginal main effect of study list, $F(1,6) = 5.74$, $MSE = .003$, $p = .054$. However, when the condition not-tested on Test 1 was excluded from the analysis there were no significant effects of the counterbalancing factors (all Fs < 4.10, ps $\geq .08$), and still no reliable difference in proportion of items recalled for the studied and other conditions, $\underline{F} < 1$. It appears that, through a fluke of random assignment certain combinations of the study and test lists for the not-studied condition were easier to recall on Test 2 than others, but even when these counterbalancing effects were accounted for there was no significant difference in recall between the conditions on Test 2.

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. In an initial omnibus ANOVA, no effect of the study list or test list factors for the judgment task were found (all $\underline{F}$s < 2.17, $\underline{p}$s $\geq$ .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, $\underline{F}(1, 11) = 48.54$, $\underline{MSE} = .01$, $\underline{p} < .0001$. 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 forgot-it-all-along 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.

Method

Overview. 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). 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 (<u>other-context</u> condition). Additionally, the second test required participants to judge whether they remembered recalling the word on the free-recall test.

<u>Subjects.</u> Twenty seven University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course. Three of these 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.

<u>Materials.</u> 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 (<u>study list</u> 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 (<u>test list</u> 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, in contrast, 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 encountered 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. Before starting the first list, participants were given the practice list to ensure that they understood the instructions and were comfortable with the tasks. After completing the first phase of the experiment, 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 with 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

Recall Performance. During free recall 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 in an earlier list. The proportions correctly recalled on Test 1 were initially analyzed in a 2 (Study list: study list 1 vs. study list 2) x 2 (Test list: test list 1 vs. test list 2) between-subjects ANOVA. The proportion of items correctly recalled in Phase 1 was reliably affected by the counterbalancing factor of study list, but not test list ($F < 1$): Collapsing across test list, participants who studied list 2 ($M = .51$, $SD = .13$) were significantly more likely to recall the target words than participants who studied the target items with list 1 ($M = .41$, $SD = .09$), $t(22) = 2.27$, $p = .03$. Therefore, even though the studied meaning of the target items had been randomly assigned to the study list factor, participants free-recalled more of the targets in study list 2 than study list 1. However, the finding that one group of target words was more difficult to recall on Test 1 when studied with one of two meanings is not important for the present paper, especially when further analysis revealed that such between-conditions results are not found for the judgment of prior remembering. Overall, participants recalled an average of 36.75 target words (45.94%) and 2.04 incorrect intrusions in free recall. Finally, participants classified an average of 71.54% of the recalled target events as "Remember."

The proportions of items correctly recalled on Test 2 were analyzed in a 2 (Test 2: studied-context vs. other-context) x 2 (Study list: study list 1 vs. study list 2) x 2 (Test list: test list 1 vs. test list 2) mixed-models ANOVA. Recall performance in Phase 2 was not reliably affected by either the study list or test list factors (all $Fs < 1.80$, $ps \geq .20$), and therefore the data were collapsed across these manipulations. On the final test, proportion correct cued recall was significantly higher for items in the studied-context condition (M = .88, SD = .06) than items in the other-context condition (M = .81, SD = .09), $t(23) = 3.75$, $p = .001$.

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. In an initial omnibus ANOVA, no effect of the study list or test list factors for the judgment task was found (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, but recalled during Test 2, by saying that they had recalled those items 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, SD = .13) than if they were cued with the studied-context word (M = .91, SD = .08), $F(1, 16) = 7.66$, MSE = .01, $p = .01$.

Experiment 5

Because the judgment task used in the previous four experiments was a 2-alternative forced choice decision (Yes/No), it is unclear whether the forgot-it-all-along effect arose from participants' incorrect belief that they had not been previously tested for an item (in Experiments 1-3), or from an incorrect belief that they had been previously tested for an item but had failed to recall that item on Test 1. That is, when participants incorrectly judged on Test 2 that they had not previously recalled an other-context item on Test 1, were they saying "No" because they believed the item had not been tested for on Test 1, or were they saying "No" because they believed they had been cued for the item on Test 1 but had not correctly recalled the target word (gave an incorrect answer or said "Pass")? Experiment 5 was designed to examine more specifically what participants mean when they incorrectly give a "No" response in the judgment task.

Method

Overview. Experiment 5 was designed to examine participants' "No" responses on the judgment task. The judgment task was modified to give participants three possible responses: (a) the item had been tested for on Test 1, and had been correctly recalled (cued and recalled), (b) the item had been tested for on Test 1, but had not been correctly recalled (cued but not recalled), and (c) the item had not been tested for on Test 1 (not cued).

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

Materials. This experiment used the same target words and context sentences as Experiment 3. Four of the items were used as primacy buffers and four as recency buffers. The target words were randomly 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. Except for modifications to the judgment instructions, the procedure from Experiment 3 was replicated. For the judgment task, participants were instructed to say "cued and recalled" if they believed that they had been tested for an item on Test 1 and they could remember recalling that item. They were told to say "cued but not recalled" if they believed that they had been tested for an item on Test 1 but had not recalled the item when tested for it (i.e., they remembered saying "Pass" or giving an incorrect answer). Finally, participants were instructed to say "not cued" if they believed the item had not been tested for on Test 1. The three judgment options were displayed on the screen for each trial.

Results

Recall performance. The proportions correctly recalled on Test 1 were initially analyzed in a 2 (Test 1: studied-context vs. other-context) x 2 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. Test 1 recall performance was not reliably affected by the counterbalancing factors of study list or test list (all Fs < 3.40, ps $\geq .10$), so the data were collapsed across these manipulations and a t-test was conducted. Proportion correctly recalled on Test 1 was significantly higher for items in the studied-context condition (M = .93, SD = .07) than items in the other-context condition (M = .81, SD = .05), $t(11) = 5.94$, $p < .0001$.

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 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. There was no effect of the study list counterbalancing factor on Test 2 recall, (all Fs < 1.60, ps ≥ .28), and the data were subsequently collapsed across this variable. There was no significant difference in proportion of items recalled for the studied-context (M = .95, SD = .05), other-context (M = .95, SD = .04), and not-tested (M = .93, SD = .03) conditions, $F(2,18) = 2.50$, MSE = .002, $p > .11$. However, there was a reliable interaction of Test 1 Context x Test List, $F(4, 12) = 4.92$, MSE = .004, $p = .01$. Although there was no difference in the proportion of items recalled between the conditions on Test 2, it appears that some of the combinations of Test 1 context and test lists were better recalled on Test 2 than others.

Judgment of previous recollection. The proportions of items on Test 2 correctly judged as recalled on Test 1 are shown in Table 6. 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 6), 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 effect of the study list or test list factors for the judgment task were found (all Fs < 2.85, ps ≥ .14), 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, $F(1, 11) = 52.01$, $MSE = .01$, $p < .0001$. Replicating the previous experiments, participants rarely erred on not-tested items by saying that they had remembered those items on Test 1.

The average number of items that participants correctly recalled on both tests but failed to remember previously recalling (i.e., collapsing across the "Not Cued" and "Cued but Not Recalled" judgment options) was 2.58 for the studied-context condition and 10.67 for the other-context condition. The majority of these items were classified by participants as "Not Cued" on Test 1, rather than "Cued but Not Recalled," and this was true for both the studied- ($M = .95$, $SD = .11$) and other-context ($M = .91$, $SD = .10$) conditions. This pattern of responses for the judgment task does not appear to be the result of a difficulty of participants in understanding the distinction between the three judgment choices (e.g., confusion concerning the difference between "not cued" and "cued but not recalled") because participants almost always correctly judged the items that they recalled on Test 2 in the not-tested condition ($M = 32.92$) as "Not Cued" on Test 1 ($M = .96$, $SD = .05$). Instead, these results clearly indicate that when participants forgot they had previously recalled an item it was usually because they mistakenly believed they had not previously been tested for that item.

Experiment 6

The results of the previous five experiments support the hypothesis that a forgot-it-all-along effect can arise from changes in context between different episodes of recollection. It could be argued, however, that the forced-choice format of the judgment task in the previous five experiments treated all responses as equal. More specifically, it is possible that participants were not very confident about their incorrect "No" responses in the other-context condition, and the restricted range of response for the judgment task did not register this low confidence. In the two examples of the forgot-it-all-along effect presented by Schooler (in press), the women involved were quite adamant that they had not been able to recall their abuse prior to their recovered memory experiences–that is, both women were highly confident that they possessed no recollection of the abuse during the period of time they believed they were amnesic for it. Thus it is important to examine how participants in the forgot-it-all-along paradigm experience their failure to remember prior remembering. If our experimental paradigm parallels the forgot-it-all-along effect described by Schooler, then participants should, at least some of the time, be confident in their incorrect belief that they had not previously recalled a context-change item. To explore the issue of confidence we implemented a confidence rating for the judgment of prior recall task.

Method

Overview. Experiment 6 was designed to examine participants' responses on the judgment task by using a confidence rating task. Participants were required to provide a confidence rating on a 6-point scale (from 1 = very low confidence, to 6 = very high confidence) after each judgment decision.

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

Materials. Experiment 6 used the same list of target words and context sentences as Experiment 3. Four of the items were used as primacy buffers and four as recency buffers. The target words were randomly 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. Experiment 6 was a replication of Experiment 3, with the only modification being the addition of a confidence rating task. After making each "Yes/No" judgment of prior remembering decision, participants were shown a screen with a 6-point confidence scale; (a) very low, (b) quite low, (c) somewhat low, (d) somewhat high, (e) quite high, and (f) very high. Participants were instructed to choose the option that best described their confidence for the judgment decision they had just completed. The experimenter stressed to the participants that there was no right or wrong answer to the confidence rating task, and that they should use the full range of the scale to select the option that best reflected their level of confidence.

Results

Recall performance. The proportion of items correctly recalled on Test 1 were analyzed in a 2 (Test 1: studied-context vs. other-context) x 2 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. There was no effect of the study list counterbalancing factor on Test 1 recall, (all Fs < 1.92, ps $\geq .22$), and the data were subsequently collapsed across this variable. There was a reliable interaction of Test 1 Context x Test List, $F(1, 9) = 7.20$, $MSE = .01$, $p = .01$. That is, it appears that some combinations of the test lists with Test 1 context were recalled better than others. Because recall is not the focus of the experiment, this result is not discussed further. Collapsing across test list, proportion correctly recalled on Test 1 was significantly higher for items in the studied-context condition (M = .93, SD = .06) than items in the other-context condition (M = .80, SD = .09), $t(11) = 5.29$, $p < .0001$.

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 (Study list: study list 1 vs. study list 2) x 3 (Test list: test lists 1-3) mixed-models ANOVA. Test 2 recall performance was not reliably affected by the counterbalancing factors of study list or test list (all Fs < 1) and the data were collapsed across these manipulations. 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 = .94, SD = .06), other-context (M = .93, SD = .05), and not-tested (M = .92, SD = .07) 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 7. 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 7), 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 effect of the study list or test list factors for the judgment task were found (all Fs < 1.76, ps ≥ .25), 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, $F(1, 11) = 30.75$, MSE = .01, $p < .0001$. Participants rarely erred on not-tested items by saying that they had remembered those items on Test 1.

We conducted a subanalysis restricted to items for which participants claimed to have "somewhat high," "quite high," or "very high" confidence in their judgments regarding previous recollection. The effect of context cues on judgments of prior remembering was the same as the overall analysis reported above: Even when confidence fell in the "high" range of the confidence scale, participants less often remembered that they had recalled an item if the item was cued with the other-context word on Test 1 (M = .70, SD = .20) than if they had been cued with the studied-context word on Test 1 (M = .95, SD = .08), $F(1, 11) = 23.48$, MSE = .02, p = .001. A further analysis was restricted to the two highest confidence ratings ("quite high" and "very high" confidence). The data from 2 participants were excluded from this analysis because less than 10 items were recalled in either the studied- or other-context conditions for the two highest confidence ratings. For the remaining participants, the same pattern of results previously reported for judgments of prior remembering were found: Participants more often failed to remember that they had recalled an item if the item was cued with the other-context word on Test 1 (M = .80, SD = .17) than if they were cued with the studied-context word on Test 1 (M = .98, SD = .05), $F(1, 9) = 8.80$, MSE = .02, $p < .02$.

Although the forgot-it-all-along effect was found for items in the high confidence range, those analyses do not get at the question of interest–How confident are participants when they fail to recall their previous recollection of an item in the other-context condition? The confidence ratings for the total number of items for the studied- and other-context conditions (collapsing across participants) that were incorrectly judged as not recalled on Test 1 are shown in Figure 1. As Figure 1 demonstrates, participants were often confident about their incorrect "No" judgments in the other-context condition. That is, more than half of the time (61.67%) in the other-context condition participants rated their incorrect "No" judgments in the high-range of the confidence scale. These confidence data demonstrate that participants are not simply choosing the "No" option for the judgment of prior recall because of low confidence (e.g., "I guess I'll say 'No', but only because I have to give either a 'Yes' or 'No' answer"). Rather, these results clearly indicate that–similar to the two forgot-it-all-along case studies described by Schooler (in press)–participants in our experiments can be highly confident in their erroneous belief that they had not previously recalled an item.

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. Additionally, participants were often in the high range of confidence regarding their incorrect judgments for context-change items. Thus our results support the existence of a forgot-it-all-along effect.

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). Indeed, if we hadn't asked participants to make judgments about prior recollection they probably would not have done so spontaneously. 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!" Nothing about recalling a word on Test 2 is likely to evoke such an experience. In contrast, for a person who has not self-identified as a survivor of abuse, remembering a past experience as an instance of abuse would produce a great deal of surprising discrepancy–that is, the memory would be deeply discrepant with beliefs about the self. As Whittlesea and colleagues have argued (e.g., Whittlesea & Williams, 1998), detecting such discrepancies creates a problem for the cognitive system, and such problems must be solved. In this case, one way to solve the problem is to infer that the event in question had long been forgotten. Of course, this inference can be made only if the individual fails to recollect prior occasions on which s/he had remembered the event in question. Our results suggest that people may well fail to remember a prior instance of recollecting an event if they currently recollect that event in a qualitatively different way than they previously had.

Remembering prior remembering is likely similar to remembering other sorts of past experiences. That is, an episode of recollection can 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).

It may seem obvious that remembering prior remembering is analogous to remembering other sorts of prior events, but it often appears that these two types of episodes are handled differently by researchers. That is, remembering prior remembering is sometimes treated as being somewhat different from remembering for other prior events (e.g., claims of prior remembering unquestionably taken as "true"). For example, in Bahrick's (1984) investigation of permastore (i.e., long-term retention of semantic memory) for highschool Spanish, participants rarely claimed to have rehearsed the Spanish material during the retention interval under investigation. Further, Bahrick found that rehearsal was not correlated with the retention of Spanish learned in highschool, and he concluded that "the data therefore reflect no important influence of the rehearsal variables" (p. 12). Although permastore is most likely not solely the result of past rehearsal for the semantic information, it may be that rehearsal does play a part in the effect: Participants may have failed to recall previous instances of rehearsal because their present situation (i.e., the experimental setting) did not match with their previous episodes of rehearsal for the material being tested. The important issue is that self-reports concerning prior remembering/rehearsal cannot be treated as strictly veridical in nature, and it is important to account for the potential factors that could influence these reports.

The remainder of this discussion examines the forgot-it-all-along effect in respect to related memory effects and other metamemory judgments. The following sections explore relevant commonalities and distinctions between the forgot-it-all-along effect and the related areas, including a focus on an attributional approach interpretation of the phenomena discussed. As its name suggests, the attributional approach is simply that–an approach (as compared to a detailed theory). However, an attributional approach is advocated here because of its emphasis on the reconstructive nature of memory and the multiple factors that can influence the reconstruction of past events. Although several potential factors are discussed below (e.g., context, fluency, etc.), these sections are by no means an exhaustive examination of the influences and conditions that impact memory performance.

The Knew-it-all-along Effect

As mentioned in the introduction, the forgot-it-all-along effect was named in reference to the knew-it-all-along effect (overestimation of prior knowledge). It would be premature to assume that these two effects are merely opposite ends of the same continuum, but it is reasonable to compare the two phenomena. For example, both the knew-it-all-along and forgot-it-all-along effects centre around retrospective knowledge: Participants are instructed to focus on their past in order to perform a given task in the present. More specifically, both effects arise when participants are unable to assess their prior performance (i.e., level of pre-feedback knowledge for the knew-it-all-along effect, and prior recall for the forgot-it-all-along effect) to the extent that is necessary to accurately complete their present task. Further, it is the addition of or change in information (correct answers for knew-it-all-along, and context-change for forgot-it-all-along) that hinders performance; participants are fully aware that they must ignore the additional/changed information but it appears that they are unable to do so.

Fischhoff (1975, 1977; Fischhoff & Beyth, 1975) hypothesized that the knew-it-all-along effect may result from an automatic assimilation of the correct feedback with pre-existing knowledge. A feeling of "knew it all along" occurs because the assimilation of new information with prior knowledge effectively eradicates the original knowledge state, making it impossible for an individual to "recapture" their previous level of knowledge. Because this process is automatic and immediate it is difficult for people to comprehend the impact post-event information has on their perception of past knowledge, even when they are warned about the phenomenon. The automatic assimilation hypothesis does not explicitly make a distinction between semantic and episodic memory systems, such that the assimilation of feedback occurs within a semantic memory system. However, the hypothesis implicitly denotes a separation between semantic and episodic memory: Feedback alters semantic knowledge, but it does not overwrite the prior episodic event itself. For example, giving participants the correct answers to general-knowledge questions in a knew-it-all-along experiment hinders their ability to accurately reconstruct the answers they gave prior to the feedback, but it does not leave them unable to remember the fact that they had previously answered the same questions. Further, this hypothesis implies that the updating effect of feedback on existing knowledge also appears to leave no trace of its occurrence in the semantic system; rather, the relevant knowledge and beliefs are simply altered in accord with the new information. Fischhoff (1977) pointed to the failure of debiasing instructions (i.e., informing participants of the effect and cautioning them to avoid overestimating their previous knowledge) to reduce the hindsight bias as supporting the notion that participants lack awareness that the post-event information influenced their hindsight judgments: Even when warned of the bias, participants are unable to adjust adequately for the effects of being exposed to feedback information.

The automatic assimilation hypothesis described above does not appear to transfer readily to the forgot-it-all-along effect. The automatic assimilation hypothesis would argue that a change in context alters the pre-existing context of a target item (i.e., the context in which an item had previously been studied or recalled), however it does not seem clear as to why this alteration of prior context would result in a forgot-it-all-along effect. Rather, a logical prediction that would follow from an automatic assimilation perspective is that participants would remember on Test 2 that they had previously recalled a context-change item, but they would be unable to correctly remember the context within which they had previously remembered that item (e.g., they believe the Test 1 context was the same as the Test 2 context). The fundamental problem for the automatic assimilation explanation of the forgot-it-all-along effect is that it does not seem sensible to say that there is a semantic entry corresponding to the study episode (because it is an episode), nor to say that this bit of semantic memory would be updated by a Test 1 and/or Test 2 encounter in a different context. Further, the semantic representation could not be used to make judgments of prior remembering because by definition it does not record details of the Test 1 event.

In terms of the knew-it-all-along effect, the automatic assimilation hypothesis has proven difficult to invalidate. Although there is some evidence that the knew-it-all-along effect can be moderated with experimental manipulations (e.g., Hasher, Attig, & Alba, 1981; Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988), there has been no solid evidence that the integration of feedback information does not result in the elimination of original states of knowledge. Part of the problem may be that, in the way most researchers have chosen to define it, the automatic assimilation hypothesis is unfalsifiable: Even if strong evidence for mitigation of the effect by controlled processes is found, one could still argue that the use of strategies other than the recall of prior beliefs are being used to make the judgments. For example, Slovic and Fischhoff (1977) reported that forcing participants to consider alternate solutions while judging their prior knowledge reduced the hindsight bias (Experiment 2). Slovic and Fischhoff argued that participants may use an availability heuristic to assess prior knowledge (rather than attempting to access memory), and that reducing the availability of feedback by introducing alternative solutions therefore leads to a smaller knew-it-all-along effect. However, the incorporation of strategies as a basis for reconstructing prior knowledge into the automatic assimilation approach lessens the need for the main tenant of the hypothesis. That is, if the automatic assimilation hypothesis claims that participants may use a strategy to reconstruct previous knowledge (regardless of whether the strategy leads to accurate reconstruction) then there is no need to postulate that receiving feedback actually eliminates prior knowledge. Therefore, as in the case of the forgot-it-all-along effect, the automatic assimilation hypothesis does not appear to be the most suitable interpretation of the knew-it-all-along effect.

Jacoby and Kelley (1987) proposed an attributional approach to the knew-it-all-along effect, in which they argued that giving participants feedback "spoils" their subjective experience, thereby contaminating the chief basis upon which they would judge an answer. For example, it is likely that feedback information is more accessible at test (e.g., due to recency) than prior knowledge, and this accessibility leads to more fluent processing of the feedback information. High accessibility of the feedback information would not, in itself, lead to a knew-it-all-along effect, but rather individuals must erroneously attribute the fluently generated feedback information to prior knowledge. It is important to note that there is a critical distinction between Jacoby and Kelley's (1987) attributional approach and the automatic assimilation hypothesis. For the attributional approach, no claims of a destruction of original memory are made and the operation of unconscious processes is not banished to the time of feedback, in that the unaware influences of memory could be occurring at the time of retrieval/reconstruction of the hindsight judgments. Unlike the automatic assimilation theory, an attributional approach to the knew-it-all-along and forgot-it-all-along effects does not distinguish between remembering an original knowledge state and reconstructing it.

An attributional approach can be seen as focussing on how available information is used to make subjective judgments, depending on the task at hand. For instance, Begg et al. (1996) proposed that the knew-it-all-along effect may be due, at least in part, to basing judgments on familiarity: Facts learned during an experiment feel familiar, and this familiarity is mis-attributed to prior knowledge. Begg et al. used three different tests following a study phase to examine what they called the illusory-knowledge effect. Participants in the total-recall test were simply asked to answer as many question as they could (number of questions answered with and without study), whereas in the knew-it test participants were instructed to answer only the questions they believed they could have answered prior to the study phase. In the learned-it test participants were told to respond to a question only if they had learned the answer during the study phase (and had not known the answer prior to the experiment). A measure of the knew-it-all-along effect was obtained by comparing the learned-it and knew-it tests to the total-recall test. In general, a hindsight bias was found for subjects in the knew-it test, but not the learned-it test. Begg et al. (1996) claimed that the lack of bias in the learned-it test, along with the knew-it-all-along effect in the knew-it test, indicates that the effects of feedback depend on where attention is focussed. That is, participants are capable of correctly identifying items they have just learned when their task is to deal with this new information in the present. It may be that the familiarity produced from feedback results in a hindsight bias when attention is focussed on the past, but that this same familiarity leads to correctly identifying new (learned) items when attention is focussed on the present: What you attribute the familiarity to depends on the goal you are trying to accomplish (Jacoby & Kelley, 1987).

The attributional approach to the knew-it-all-along effect discussed above can also be applied to the forgot-it-all-along effect. Because judgments of prior remembering occur after an item is recalled in a different context (other-context condition), the change in context alters the information that a participant has available to make the judgment. For example, the inconsistency of context between recall tests most likely results in a weaker feeling of familiarity for target items in the other-context condition, in comparison to the level of familiarity produced by the repetition of context in the studied-context condition. Thus the forgot-it-all-along effect may arise because participants falsely attribute the lower familiarity produced by target items in the other-context condition to not having recalled these items on the first test. As mentioned previously, though, it is not the difference in the level of familiarity itself that would lead to a forgot-it-all-along effect, but rather that participants mis-attribute this difference to some aspect of the past (i.e., whether or not an item was previously remembered). Creating a task that required participants to deal directly with context (i.e., focusing attention on the present) likely would not result in the performance deficit found in the forgot-it-all-along paradigm. For instance, say the Test 2 judgment of prior recall was changed to a task in which participants simply judged whether the target items had the same or different context as Test 1. In this case, participants may attribute the lower level of familiarity for target items in the other-context condition to Test 1 context, thereby correctly judging that these items had a different context on the first test. The advantage of the attributional approach to the forgot-it-all-along effect is its emphasis on the idea that the use of the same source of information across tasks will not necessarily result in the same observed outcome for each task (i.e., relying on familiarity can lead to both accurate and inaccurate performance, depending on the nature of the task).

Recognition Failure

The experimental model used to investigate the forgot-it-all-along effect in the experiments presented in this paper is essentially an adapted version of a recognition failure paradigm. Recognition failure has been investigated using numerous methods, but the paradigm that is of most interest for the forgot-it-all-along effect involves the manipulation of cues. Tulving and Thompson (1973, Experiment 2) first had participants study a list of target words paired with weak-associate cues, and then each participant completed the following five tasks; (1) generated six free-association responses to strong-associate (not-studied) cue words for half of the studied targets, (2) examined free-association responses and indicated which responses were studied target items, (3) presented with free-association responses generated by another participant to the strong-associate (not-studied) cue words for the remaining half of the studied targets and asked to indicate which responses were studied target items, (4) wrote down beside each recognized target word from Task 3 the corresponding studied (weak-associate) cue, and (5) presented with all of the studied weak-associate cues and asked to recall as many target items as could be remembered. Interestingly, participants could generate 80% of the target words in the free-association test (task 1), but they were only able to recognize 2.8% of these words as studied target items (task 2). Further, participants' recognition of target words was no better when they were given another participant's free-association responses (task 3), although for the items they did recognize as targets they were able to report 71% of the weak-associate cues they had studied with those targets. Finally, when given the studied weak-associate cues, participants were able to recall 59% of the studied target items. In sum, the results indicated that participants were much more likely to remember that a word was a studied target item if they completed a task in the context of the weakly-associated studied cue word than if they completed a task in the context of the strongly-associated unstudied cue word.

The main distinction between the recognition failure paradigm and the forgot-it-all-along paradigm is researchers' focus on the type of event that participants fail to remember. In the forgot-it-all-along paradigm the effect of interest is the failure to remember prior successful recall of an item, whereas in the realm of recognition failure the effect of interest is the failure to recognize the item itself. However, both effects could conceivably be produced within the same paradigm. If participants in Tulving and Thompson's (1973) recognition failure paradigm had been asked on the final task (recalling target words to weak-associate studied cues) whether they had recognized each item as a target word in the second or third task (free-association with strong-associate unstudied cues), a forgot-it-all-along effect would most likely have been found.

As mentioned in the introduction, the dominant explanation for recognition failure is based on the encoding specificity principle. Tulving (1984) argued that the strength of a memory trace should not be discussed in isolation because "any given trace has many different 'strengths', depending upon retrieval conditions" (p. 233). Related to the encoding specificity principle is the transfer-appropriate processing hypothesis. The transfer-appropriate processing hypothesis states that the closer the match between the processes used at encoding and retrieval, the better performance will be at retrieval (see

Franks et al., 2000; Morris, Bransford, & Franks, 1977; Rajaram, Srinivas, & Roediger, 1998). For example, Morris et al. (1977) demonstrated that an encoding task that involves a deep level of processing does not necessarily lead to better performance on a subsequent task than an encoding task that involves more superficial processing. In the Morris et al. experiments, participants engaged in both semantic (deep) and rhyme (superficial) acquisition of target items, followed by either a standard recognition test or a rhyme recognition test. Recognition on the standard recognition test was higher for items that had been studied with semantic acquisition than for items that had been studied with rhyme acquisition. Conversely, however, recognition on the rhyme recognition test was higher for items that had been studied with rhyme acquisition than for items that had been studied with semantic acquisition. Morris et al. argued that the transfer-appropriate processing hypothesis could account for these results because one of the major strengths of this hypothesis is that "[it] emphasizes that the value of particular acquisition activities must be defined relative to particular goals and purposes" (p. 528).

The encoding specificity principle and the transfer-appropriate processing hypothesis compliment and closely mirror one another, with the main distinction being that the former tends to focus more on features while the latter focuses more on processes. Henceforth, the focus will be on the transfer-appropriate processing hypothesis, although discussions concerning this hypothesis are largely relevant to the encoding specificity principle as well.

A major advantage of the transfer-appropriate processing hypothesis is that the various factors proposed to have an impact on memory performance are not restricted to either the encoding stage or the retrieval stage. Instead, transfer-appropriate processing emphasizes the importance of the interaction between encoding and retrieval conditions (Franks et al., 2000; Tulving & Thompson 1973). Additionally, the transfer-appropriate processing hypothesis highlights Tulving's (1984) concept of synergistic ecphory. The term ecphory refers to the establishment of a memory trace during encoding, and synergistic ecphory involves the interaction of ecphoric information and retrieval cues in producing the recollection of an event. Tulving's concept of synergistic ecphory emphasizes that it is the combination of the encoding and retrieval conditions that gives rise to a recollective experience: The recollection of an event is not dependant on ecphoric information in isolation from retrieval cues (and vice versa).

On the surface, results from the recognition failure experiments seem counterintuitive–strong-associate cues should lead to higher recall than weak-associate cues and recognition performance should be better than recall performance. The transfer-appropriate processing hypothesis can easily account for these curious results–that is, it is not simply the conditions within which a target word is studied that drives the effect (e.g., weak vs. strong associative cue words), but rather it is the interplay of the conditions at study and the conditions at the time of recall that are responsible for the production of the effect. Thus, for example, participants failed to recognize a target item in the context of a strong-associate non-studied cue because that cue led them to think of the target in a way that did not support the experience of having studied that item as a target.

The transfer-appropriate processing interpretation of recognition failure readily transmits to the forgot-it-all-along effect. When a context-change target item is thought about one way during Test 1 and in another way during Test 2, then the way of thinking about that item during Test 2 is a relatively poor cue for evoking memories of the Test 1 episode of recollection. That is, the processes employed at Test 2 are not "transfer appropriate" for eliciting memories of Test 1 recollection for the context-change target items. Further, the results from Experiment 2 demonstrate that a change in context is only detrimental for the judgment of prior recollection if the context change occurs on only one of the two recall tests: A change in context that remains constant for both recall tests (other/other condition) does not produce the forgot-it-all-along effect. Granted, the condition for which a change in context remains constant across the two recall tests is not technically a context-change situation, because context on Test 1 matches context on Test 2 (i.e., in this situation context-change refers to a difference in context from the Study Phase, not a difference in context between the two recall tests). However, the important point is that there can be many combinations of contexts across episodes, and a change in context between episodes does not necessarily translate in to significantly poorer performance, depending on the task at hand.

The importance of the type of task being performed is clearly shown in the example of the consistent context-change given in the above paragraph. That is, a change in context that remained consistent (other/other condition) was not detrimental to performance on the judgment task, but it was detrimental to performance on both recall tasks. This example also serves to illustrate that even though recognition failure and the forgot-it-all-along effect can be produced using the same manipulations (i.e., context), the outcome observed for each phenomenon will not always be equivalent because they are two different tasks that do not necessarily tap the same processes or sources of information.

Other Metamemory Judgments

Although there are a variety of metamemory judgments (e.g., memorability ratings, tip-of-the-tongue, etc; see Guttentag & Carroll, 1998; Metcalfe, 2000) this section focuses on three areas that are particularly relevant to the forgot-it-all-along effect; feelings-of-knowing, judgments-of-learning, and false memories for false memories.

Feeling-of-knowing. Hart (1965) is credited with fueling interest in metamemory judgments, particularly in the area of the feeling-of-knowing. A feeling-of-knowing is said to occur when a person is unable to recall some piece of information, but feels s/he could recognize (or recall) the information at a later time (Koriat & Levy-Sadot, 1999). To examine the feeling-of-knowing, Hart employed a recall-judgment-recognition (RJR) paradigm; participants attempted to recall answers to general-information questions, and for the questions that they were unable to answer they judged whether they believed they could recognize the correct answers among a set of distractors. Finally, participants were given a multiple-choice test for all of the questions. Hart found that the feeling-of-knowing judgment was quite accurate in predicting subsequent performance on the recognition task. That is, participants were able to predict both whether they would be successful or not successful in recognizing the answers to the questions for which they had not previously recalled answers.

Hart (1965) proposed a direct-access explanation for the feeling-of-knowing. According to the direct-access approach, the memory system can monitor what it has stored, and the feeling-of-knowing functions as a helpful indicator of what knowledge is present in memory. More specifically, a strong feeling-of-knowing indicates to the system that, even though an item cannot presently be recalled the item is in memory and therefore efforts to retrieve that item should continue. Conversely, the absence of a feeling-of-knowing for an item that currently cannot be recalled indicates to the system that it is unlikely that the item is in storage, and retrieval efforts should be terminated.

In contrast to this direct-access approach, Koriat (1993; Koriat & Levy-Sadot, 1999) suggested that the feeling-of-knowing is produced by an accessibility heuristic. If an item cannot presently be recalled, partial information relating to that item may nonetheless come to mind, and the more information that comes to mind, the stronger the feeling-of-knowing. Koriat (1993) found that when participants were unable to recall a studied nonsense string of letters, the number of correct and incorrect letters from the string that they were able to produce correlated with their feeling-of-knowing judgments: The more correct/incorrect letters produced for an unrecalled item, the higher the feeling-of-knowing judgment. Further, Koriat stated that the feeling-of-knowing is usually an accurate judgment of future recognition because the partial information that comes to mind tends to be accurate (e.g., in the nonsense letter string experiment, 90% of the letters reported for the unrecalled items were correct). These results do not seem to be compatible with the direct-access approach, in that there does not appear to be any parsimonious way of accounting for such findings within the direct-access approach (Koriat & Levy-Sadot, 1999).

The accessibility heuristic approach to the feeling-of-knowing is very similar to the attributional approach that has been advocated in this paper. The advantage of an attributional approach, however, is that it takes into account a wider variety of influences that may affect the feeling-of-knowing judgment. For example, Maki (1999) investigated the feeling-of-knowing judgment with an interference paradigm in which participants studied and were tested on stimulus-response pairs for two separate lists. Interference was created by manipulating the similarity of the stimuli and responses between the two tests (e.g., stimuli repeated for both lists, but paired with different responses), and Maki found that feeling-of-knowing judgments were higher when repeated stimuli were paired across the two lists with responses that were associated (same-associate condition), compared to when repeated stimuli were paired with different responses across the two lists (same-different condition). Maki claimed that these results were incompatible with an accessibility heuristic explanation because the accessibility heuristic approach would have predicted higher judgments for the same-different condition: Two different responses for the same stimuli across lists should result in more partial information coming to mind for an unrecalled item, compared to when the two responses are associates.

Although Maki's (1999) results appear problematic for the accessibility heuristic approach, they can be explained using the attributional approach. The attributional approach would posit that individuals may use the amount of information that comes to mind about an unrecalled item as one of the basis for the feeling-of-knowing judgment, but that individuals are not limited to only this source of information in making the judgment. In Maki's experiments, attempting to recall the responses in the same-associate condition may have been more fluent than in the same-different condition because the responses in the former condition were related (e.g., partial information may come to mind faster or more vividly when the responses are associates because of the similarities between the items). Further, participants may (mis)attribute this fluency to the item they are currently attempting (but unable) to recall, and consequently they assign a higher feeling-of-knowing judgment to that item because they feel they are very close to remembering it.

The above hypothesis generated from the attributional approach also can be applied to the forgot-it-all-along effect. Overall, recalling a target item cued with a studied-context word on both recall tests is likely more fluent than recalling a target item to an other-context word on the first test and a studied-context word on the second test (or vice versa). Part of the forgot-it-all-along effect may be due to this difference in fluency; changing context between tests leads to less fluent recall, which may alter the information available for judging prior remembering (a change in context reduces the fluency produced from encountering and recalling the item on Test 1). In sum, whether or not fluency plays a role in either the forgot-it-all-along effect or the feeling-of-knowing judgment, the important point is that the attributional approach is capable of integrating a variety of sources of information that may potentially influence these judgments about memory.

Judgments-of-learning. Judgments-of-learning require a person to rate the likelihood (typically during encoding) that s/he will remember a recently studied item in the future (Koriat & Levy-Sadot, 1999; Schwartz, 1994). Researchers have typically been interested in judgments-of-learning in relation to whether people can accurately determine the amount of time and resources they must invest in learning material to the extent that they could remember that material in the future (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989).

Results pertaining to the accuracy of judgments-of-learning for predicting future recollection have been mixed. Judgments-of-learning made immediately after studying an item tend to correlate only modestly with future performance, whereas judgments-of-learning that are made after a delay (e.g., 5-min delay between studying an item and completing the judgment-of-learning) tend to correlate almost perfectly with subsequent performance (Nelson & Dunlosky, 1991). Nelson and Dunlosky (1991) claimed that immediate judgments-of-learning are less accurate because these judgments are based on short-term memory, whereas success in a future task depends on long-term memory. They further stated that delayed judgments-of-learning are good predictors of subsequent performance because people are more likely to rely on long-term memory to make these judgments; short-term memory is no longer available and therefore not an influence in delayed judgments-of-learning.

Spellman and Bjork (1992) argued that delayed judgments-of-learning correlate highly with future performance because participants treat the delayed judgment as a recall task. More specifically, when participants perform the delayed judgment-of-learning task they cue themselves for the studied item–if the item comes to mind it is given a high judgment-of-learning rating (and vice versa, if the studied item cannot be brought to mind it is given a low judgment-of-learning score). Post-test reports from participants support this hypothesis. Of the 20 participants who made delayed judgments in Nelson and Dunlosky's (1991) experiment, 19 stated that they had tried to recall the studied item before making the judgment, whereas only half of the participants who made immediate judgments claimed to have made attempts at recalling the items (Spellman & Bjork, 1992).

Setting aside the immediate versus delayed debate, other researchers have discovered factors that influence the accuracy of judgments-of-learning for predicting future performance. For instance, Begg et al. (1989) demonstrated that high-frequency words were given higher judgment-of-learning ratings than low-frequency words, yet the low-frequency words were remembered at a higher rate on a subsequent cued recall test. In another experiment, Begg et al. found that after studying word pairs (e.g, "A-B") participants were more accurate in predicting their future recall performance if they were shown the first item and asked to judge how likely they thought it was they would recall the second word ("A → B") than if they were shown the second item and asked to predict their future recall of the first word ("B → A"). Using a transfer-appropriate processing approach, Begg et al. asserted that judgments-of-learning (as well as other metamemory judgments) are accurate to the extent that the processes underlying the judgment task and the memory test are the same; "if [a] test requires producing an item, predictions should be based on production rather than comprehension..." (p. 631).

Using this transfer-appropriate processing approach, some interesting predictions regarding the correlation of judgments-of-learning with the forgot-it-all-along effect can be made. Granted, judgments-of-learning are prospective judgments whereas judgments of prior recall are retrospective. However, because a transfer-appropriate processing explanation has been proposed for both judgments some predictions concerning the relationship between the two judgments can be made. For example, let's say in a forgot-it-all-along paradigm where context was manipulated on Test 1 that participants made judgments-of-learning on Test 1 regarding how likely they thought they would be able to recall each item on Test 2. If participants use the fluency of processing as a basis for the judgments-of-learning, then items in the other-context condition should receive lower judgments than items in the studied-context condition. Lower judgments given to other-context items on Test 1 would lead to a low correlation between judgments-of-learning and subsequent recall performance because there is no difference in recall on Test 2 between the studied- and other-context conditions (as in Experiments 1, 3, 5, and 6).

In the scenario described above, judgments-of-learning would have a low overall correlation with Test 2 recall, but a high correlation with the judgment of previous recall (i.e., they both would show less accurate judgments for the other-condition). But scenarios in which the judgment-of-learning and the judgment of previous recall do not correlate can also be proposed. For instance, let's consider the free recall forgot-it-all-along paradigm (Experiment 4); suppose during the free recall test participants were required to give judgment-of-learning ratings for each word they recall. Unlike the previous example, here context is not manipulated until after the judgments-of-learning are completed. Because context does not change until Test 2, there is no reason to predict that the judgments-of-learning will correlate with the judgment of prior recall, and most likely there would be little or no correlation between the two judgments. These examples illustrate that–like the memory performance they are trying to capture–metamemory judgments (such as judgments-of-learning and judgments of previous recall) will only correlate to the extent that they rely on similar processes and sources of information.

False memory for false memory. Using an associated word list paradigm, Marsh and Hicks (2001) recently demonstrated that individuals can produce false memories about recalling items that they incorrectly believe they had previously encountered. In a typical associated word list paradigm (a.k.a., Deese-Roediger-McDermott [DRM] paradigm) participants study a list of words that are related to a non-presented critical lure (e.g., "candy", "sugar", and "sour" for the critical non-presented lure "sweet"), followed by a free-recall test for the word list (see Deese, 1959; Read, 1996; Roediger & McDermott, 1995). Results from standard associated word list experiments not only demonstrate that participants frequently recall the non-presented critical lures, but that they often report they actually remember studying these lures (Roediger & McDermott, 1995).

Marsh and Hicks (2001) modified the associated word list paradigm to include a second memory test and judgment task. Participants studied six associated word lists, and after each list they engaged in free recall. After studying/recalling all six lists, participants were given a recognition test for each individual list. On this second test participants were shown a list of words that included the studied list items, the critical non-presented lure, and new (semantically dissimilar from list/critical lure) items. Participants were instructed to label each item as studied and not recalled, studied and recalled, or new. Results from the free recall task showed the typical pattern of false memories that have been found in previous associated word list experiments–that is, participants recalled about 50% of the non-presented critical lures. The results of interest, though, concern how the unrecalled critical lures were judged on the recognition test. Almost 90% of the non-presented critical lures that had not been recalled on the first test were judged on the recognition test as having been studied. More interestingly, almost half of the time that participants claimed to have studied these non-presented critical lures, they also believed that they had previously recalled them.[3] This outcome will be referred to subsequently as the false-memory-for-false-memory effect.

The false-memory-for-false-memory effect shares many similarities to the forgot-it-all-along effect. Both effects arise because, under certain context conditions, participants are unable to accurately assess their previous recall, resulting in an over/under estimation of their previous recall output. Additionally, the paradigms used to investigate the two effects are very similar: Participants are required to make judgments on a second memory test regarding their recall of the same items on a previous memory test. However, a major distinction between the two paradigms centers around the stimuli used to test for the effects. For the forgot-it-all-along paradigm, context is manipulated for each individual item, whereas in the false-memory-for-false-memory paradigm context is manipulated for a group of items (i.e., across separate lists). In the false-memory-for-false-memory experiments, there are three separate occasions (study/recall/recognition) where there is the potential for every item of an associated word list to serve as a repetition for a global context (the non-presented critical lure) that connects the items and theme of the list together. Conversely, in the forgot-it-all-along experiments there is no possibility for the strengthening of a global context across target items because every target has its own unique contexts (i.e., the two contexts for each target were designed so as not to overlap with any other target). This difference in how context is used between the two paradigms is most likely an important factor that leads to what looks like opposite effects. It is possible that the strengthening of a global context across episodes results in a false belief that a related item (critical lure) was previously recalled, whereas a change in context between episodes (e.g., a "weakening" of contextual information) results in a false belief that an item was not previously recalled. More to the point, different processes do not have to be postulated for each effect, but rather (as stated in previous sections) it is more likely that it is the interplay of context during and between each episode that results in the different outcomes for each phenomenon.

Another important difference between the forgot-it-all-along and false-memory-for-false-memory paradigms concerns the items included in the second memory test. In the forgot-it-all-along paradigms the second memory test only included items that participants had previously studied, whereas the second test in the false-memory-for-false-memory experiments also contained new semantically-dissimilar items. If participants use a variety of informational sources available to them at retrieval to ascertain their previous recall, then it is possible that the inclusion/exclusion of new items will impact their judgments. For example, it is possible that encountering new items on a second memory test alters how the fluency of an item is used in deciding whether it previously had been recalled.

If factors such as the impact of fluency on judgments for prior recall are changed by the addition of non-studied items, then it is possible that dropping the new items from the second memory test would reduce or eliminate the false-memory-for-false-memory effect. That is, participants in the false-memory-for-false-memory paradigm most likely process the non-presented critical lure more fluently than the new semantically-dissimilar items, and they may mis-attribute this higher processing fluency to having previously recalled the critical lure. If this hypothesis is correct, then dropping the new items from the second test should eliminate the advantage of processing fluency that the critical lures have over the new items (and it would change the information participants have available for judging previous recollection). In the same vein, adding non-studied items to the second test in the forgot-it-all-along paradigm could lead participants to experience the target items in the other-context condition as being more fluent/familiar than the new items, perhaps leading participants to attribute this difference in fluency to having previously recalled the other-context items (resulting in a reduction or elimination of the forgot-it-all-along effect).

Investigation of both the forgot-it-all-along effect and the false-memory-for-false-memory effect has only just begun, and further research is warranted before strong conclusions regarding the link between the two effects can be substantiated. However, it would be a mistake to ignore the similarities between the two effects because research conducted in one arena could lend understanding and potential avenues of exploration to the other.

Summary and Conclusion

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 episodes of prior recollection via cue-overload (Watkins & Watkins, 1976). It may also produce blended ecphoric products in which 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. 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 six experiments a substantial forgot-it-all-along effect was obtained, minutes after the initial recall episode, 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. As previously mentioned, research on a variety of memory phenomena (e.g., flashbulb memories, permastore) 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.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. Journal of Experimental Psychology: General, 113, 1-29.

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

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. Journal of Memory and Language, 28, 610-632.

Begg, I. M., Robertson, R. K., Gruppuso, V., Anas, A., & Needham, D. R. (1996). The illusory-knowledge effect. Journal of Memory and Language, 35, 410-433.

Belli, R. F., Winkielman, P., Read, J. D., Schwarz, N., & Lynn, S. J. (1998). Recalling more childhood events leads to judgments of poorer memory: Implications for the recovered/false memory debate. Psychonomic Bulletin & Review, 5, 318-323.

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

Deese, J. (1959). On the prediction of occurrence of a particular verbal intrusions in immediate recall. Journal of Experimental Psychology, 58, 17-22.

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

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1, 288-299.

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

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. Organizational Behavior and Human Performance, 13, 1-16.

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.

Guttentag, R., & Carroll, D. (1998). Memorability judgments for high- and low-frequency words. Memory & Cognition, 26, 951-958.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. Journal of Educational Psychology, 56, 208-216.

Hasher, L., Attig, M. S., & Alba, J. W. (1981). I knew it all along: Or, did I? Journal of Verbal Learning and Verbal Behavior, 20, 86-96.

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? Memory & Cognition, 16, 533-538.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. Personality and Social Psychology Bulletin, 13, 314-336.

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.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. Psychological Review, 100, 609-639.

Koriat, A., & Levy Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), Dual process theories in social psychology (pp. 483-502). New York: The Guilford Press.

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.

Maki, R. H. (1999). The roles of competition, target accessibility, and cue familiarity in metamemory for word pairs. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 1011-1023.

Marsh, R. L., & Hicks, J. L. (2001). Output monitoring tests reveal false memories of memories that never existed. Memory, 9, 39-51.

Melchert, T. P., & Parker, L. (1997). Different forms of childhood abuse and memory. Child Abuse & Neglect, 21, 125-135.

Metcalfe, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), The Oxford handbook of memory (pp. 197-211). New York: Oxford University Press.

Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519-533.

Nelson, T. O. (1996). Consciousness and metacognition. American Psychologist, 51, 102-116.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". Psychological Science, 2, 267-270.

Nelson, T. O. & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 1-25). Cambridge, MA: MIT Press.

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.

Rajaram, S., Srinivas, K., & Roediger, H. L. III (1998). A transfer-appropriate processing account of context effects in word-fragment completion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 993-1004.

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. Psychonomic Bulletin and Review, 3, 105-111.

Read, J. D. (1997). Memory issues in the diagnosis of unreported trauma. In J. D. Read & D. S. Lindsay (Eds.), Recollections of trauma: Scientific evidence and clinical practice (pp. 79-108). New York: Plenum Press.

Read, J. D., & Lindsay, D. S. (2000). "Amnesia" for summer camps and high school graduation: Memory work increases reports of prior periods of remembering less. Journal of Traumatic Stress, 13, 129-147.

Roe, C. M., & Schwartz, M.. F. (1996). Characteristics of previously forgotten memories of sexual abuse: A descriptive study. Journal of Psychiatry and Law, 24, 189-206.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 803-814.

Scheflin, A. W., & Brown, D. (1996). Repressed memory or dissociative amnesia: What the science says. Journal of Psychiatry and Law, 24, 145-188.

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.

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. Psychonomic Bulletin and Review, 1, 357-375.

Slovic, P. & Fischhoff, B. (1977). On the psychology of experimental surprise. Journal of Experimental Psychology: Human Perception and Performance, 3, 544-551.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. Psychological Science, 3, 315-316.

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.

Williams, L. M. (1995). Recovered memories of abuse in women with documented child sexual victimization histories. Journal of Traumatic Stress, 8, 649-673.

Wood, G. (1978). The knew-it-all-along effect. Journal of Experimental Psychology: Human Perception and Performance, 4, 345-353.

Footnotes

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

[2]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; $M = .69$, $SD = .16$) than when they were cued with studied-context words on Test 2 (studied/studied; $M = .78$, $SD = .14$), $t(11) = 1.92$, $p < .05$. 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; $M = .63$, $SD = .13$) than when they were cued with other-context words on Test 2 (other/other; $M = .72$, $SD = .15$), $t(11) = 2.03$, $p < .05$. 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; $M = .70$, $SD = .13$) than if they were cued with studied-context words on Test 2 (studied/studied; $M = .89$, $SD = .09$), $t(11) = 4.55$, $p = .001$. 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; $M = .58$, $SD = .16$) than when they were cued with other-context words on Test 2 (other/other; $M = .85$, $SD = .15$), $t(11) = 5.04$, $p < .0001$.

[3]Marsh and Hicks (2001) recognized that the unbalanced judgment scale for the recognition task (two options for studied, one option for new) could be a potential confound. In a subsequent experiment, participants were given a 4-point judgment scale; (1) studied and not recalled, (2) studied and recalled, (3) not studied, but erroneously recalled on Test 1, and (4) not studied and not recalled on Test 1. Additionally, the free recall test was manipulated so that half of the participants could see their recall output throughout the recall period (present condition), whereas for the other half of the participants the recalled words disappeared immediately after each word was produced (absent condition). Results once again showed a false-memory-for-false-memory effect, but participants in the present condition were much less likely to judge in the recognition test that an unrecalled critical lure had been studied and recalled, in comparison to the absent condition. It appears that not giving participants the opportunity to view their output during recall inflates their false beliefs that they had previously recalled a non-presented critical lure.

Table 1

Mean Number of Items and 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	Mean Number of Items	Mean Proportion Judged as "Recalled" on Test 1
Studied-context			
	Not Recalled/Not Recalled	1.79	.09 (.26)
	Not Recalled/Recalled	1.37	.04 (.09)
	Recalled/Recalled	**31.42**	**.81 (.10)**
	Recalled/Not Recalled	.42	.57 (.45)
Other-context			
	Not Recalled/Not Recalled	1.62	0 (--)
	Not Recalled/Recalled	6.38	.11 (.15)
	Recalled/Recalled	**25.83**	**.54 (.19)**
	Recalled/Not Recalled	1.17	.10 (.19)
Not-tested			
	NA/Not Recalled	2.79	.14 (.26)
	NA/Recalled	32.21	.10 (.12)

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

Table 2

Mean Proportion of Items Correctly Recalled in Experiment 2

Testl/Test 2 Cues	Test 1	Test 2
studied/studied	.89 (.09)	.91 (.08)
other/studied	.77 (.09)	.91 (.08)
not-tested/studied		.89 (.10)
studied/other	.88 (.08)	.82 (.14)
other/other	.77 (.13)	.82 (.11)
not-tested/other		.77 (.16)

Note. Standard deviations are in parentheses.

Table 3

Mean Number of Items and 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	Mean Number of Items	Mean Proportion Judged as "Recalled" on Test 1
Studied/Studied			
	Not Recalled/Not Recalled	1.33	.07 (.24)
	Not Recalled/Recalled	.71	.25 (.45)
	Recalled/Recalled	**15.75**	**.83 (.13)**
	Recalled/Not Recalled	.21	.38 (.48)
Other/Studied			
	Not Recalled/Not Recalled	.71	.03 (.09)
	Not Recalled/Recalled	3.42	.03 (.07)
	Recalled/Recalled	**12.92**	**.60 (.14)**
	Recalled/Not Recalled	.96	.27 (.39)
Not-tested/Studied			
	NA/Not Recalled	2.04	.02 (.08)
	NA/Recalled	15.96	.08 (.10)

[continued]

Table 3

Mean Number of Items and Proportion of Items Judged as "Recalled" as a function of Recall Status on Test 1 and Test 2 for Experiment 2 (continued)

Test 1/Test 2 Cue	Test 1/Test 2 Recall Status	Mean Number of Items	Mean Proportion Judged as "Recalled" on Test 1
Studied/Other			
	Not Recalled/Not Recalled	.83	.21 (.38)
	Not Recalled/Recalled	1.42	.03 (.12)
	Recalled/Recalled	**13.46**	**.70 (.14)**
	Recalled/Not Recalled	2.29	.57 (.45)
Other/Other			
	Not Recalled/Not Recalled	2.67	.02 (.06)
	Not Recalled/Recalled	1.54	.07 (.17)
	Recalled/Recalled	**13.25**	**.79 (.16)**
	Recalled/Not Recalled	.54	.45 (.50)
Not-tested/Other			
	NA/Not Recalled	4.17	.05 (.13)
	NA/Recalled	13.83	.08 (.09)

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

Table 4

Mean Number of Items and 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	Mean Number of Items	Mean Proportion Judged as "Recalled" on Test 1
Studied-context			
	Not Recalled/Not Recalled	1.42	0 (--)
	Not Recalled/Recalled	.83	.05 (.13)
	Recalled/Recalled	**33.58**	**.93 (.04)**
	Recalled/Not Recalled	.17	.50 (--)
Other-context			
	Not Recalled/Not Recalled	.75	0 (--)
	Not Recalled/Recalled	6.25	.06 (.10)
	Recalled/Recalled	**27.67**	**.63 (.16)**
	Recalled/Not Recalled	1.33	.18 (.23)
Not-tested			
	NA/Not Recalled	2.58	0 (--)
	NA/Recalled	33.42	.03 (.03)

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

Table 5

Mean Number of Items and 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	Mean Number of Items	Mean Proportion Judged as "Recalled" on Test 1
Studied-context			
	Not Recalled/Not Recalled	3.33	.08 (.24)
	Not Recalled/Recalled	17.87	.07 (.05)
	Recalled/Recalled	**17.29**	**.86 (.13)**
	Recalled/Not Recalled	1.50	.63 (.43)
Other-context			
	Not Recalled/Not Recalled	5.25	.03 (.09)
	Not Recalled/Recalled	16.79	.06 (.06)
	Recalled/Recalled	**15.42**	**.77 (.15)**
	Recalled/Not Recalled	2.54	.56 (.33)

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 deviations are in parentheses.

Table 6

Mean Number of Items and Proportion of Items Judged as "Recalled" and "Cued but Not Recalled" as a Function of Recall Status on Test 1 and Test 2 for Experiment 5

Test 1 Cue	Test 1/Test 2 Recall Status	Mean Number of Items	Mean Proportion Judged as "Recalled" on Test 1	Mean Proportion Judged as "Cued but Not Recalled" on Test 1
Studied-context				
	Not Recalled/Not Recalled	1.50	0 (--)	1.00 (--)
	Not Recalled/Recalled	1.08	.04 (.12)	.85 (.27)
	Recalled/Recalled	**33.25**	**.92 (.07)**	.01 (.01)
	Recalled/Not Recalled	.17	.50 (.71)	.50 (.71)
Other-context				
	Not Recalled/Not Recalled	.92	0 (--)	.38 (.44)
	Not Recalled/Recalled	6.08	.08 (.15)	.17 (.15)
	Recalled/Recalled	**28.00**	**.62 (.15)**	.03 (.04)
	Recalled/Not Recalled	1.00	.26 (.38)	.21 (.39)

[continued]

Table 6

Mean Number of Items and Proportion of Items Judged as "Recalled" and "Cued but Not Recalled" as a Function of Recall Status on Test 1 and Test 2 for Experiment 5 (continued)

Test 1 Cue	Test 1/Test 2 Recall Status	Number of Items	Mean Proportion Judged as "Recalled" on Test 1	Mean Proportion Judged as "Cued but Not Recalled" on Test 1
Not-tested				
	NA/Not Recalled	2.50	0 (--)	.13 (.20)
	NA/Recalled	33.50	.02 (.03)	.04 (.05)

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

Table 7

Mean Number of Items and Proportion of Items Judged as "Recalled" as a Function of Recall Status on Test 1 and Test 2 for Experiment 6

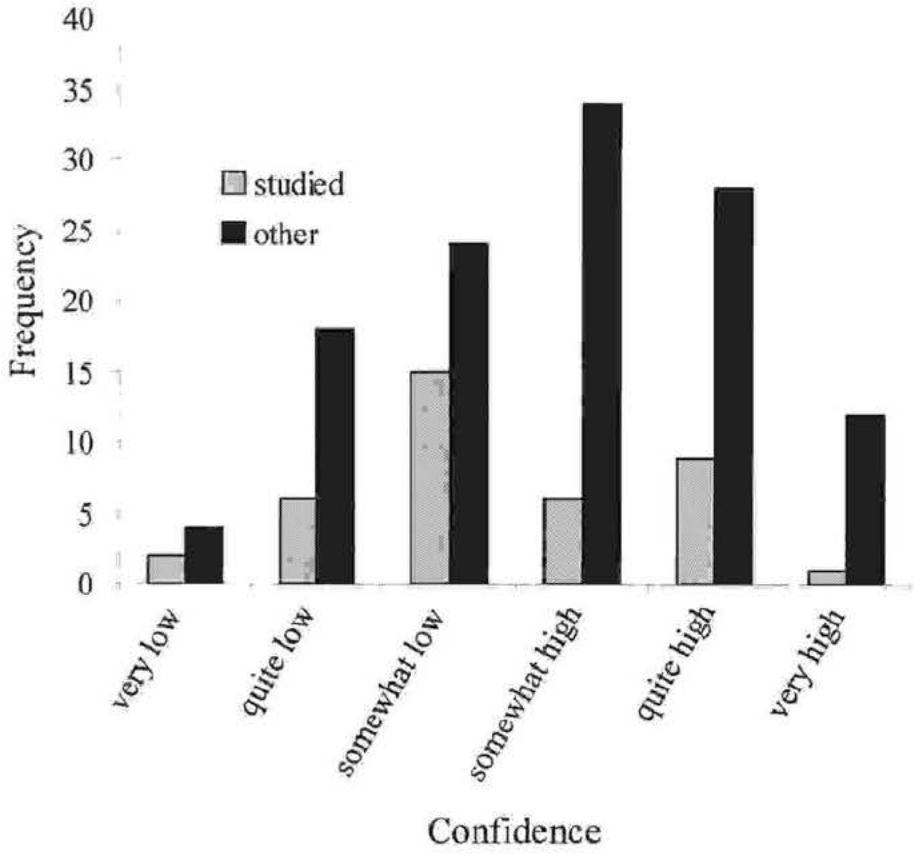
Test 1 Cue	Test 1/Test 2 Recall Status	Mean Number of Items	Mean Proportion Judged as "Recalled" on Test 1
Studied-context			
	Not Recalled/Not Recalled	1.67	.10 (.32)
	Not Recalled/Recalled	.92	.19 (.37)
	Recalled/Recalled	**32.92**	**.90 (.11)**
	Recalled/Not Recalled	.50	.50 (.56)
Other-context			
	Not Recalled/Not Recalled	1.25	.05 (.13)
	Not Recalled/Recalled	6.00	.04 (.08)
	Recalled/Recalled	**27.50**	**.63 (.17)**
	Recalled/Not Recalled	1.25	.13 (.23)
Not-tested			
	NA/Not Recalled	2.92	.08 (.18)
	NA/Recalled	33.08	.06 (.06)

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

Figure Caption

Figure 1. Confidence ratings for the total number of items (collapsed across participants) correctly recalled on Test 1 and Test 2, and incorrectly judged as "No" in the studied- (n = 39) and other-context (n = 120) conditions for Experiment 6.

Figure 1


40
35
30
25
20
15
10
5
0
Frequency
studied
other
very low
quite low
somewhat low
somewhat high
quite high
very high
Confidence

VITA

Surname: Arnold Given Names: Michelle Marie

Place of Birth: Eckville, Alberta, Canada

Educational Institutions Attended:

University of Victoria	1999 to 2001
University of Lethbridge	1994 to 1997
Red Deer College	1992 to 1994

Degrees Awarded:

B.A. (with Distinction)	University of Lethbridge	1997

Honours and Awards:

Natural Science and Engineering Research Council of Canada Postgraduate Scholarship A	1999 to 2001
President's Research Award	1999 to 2001
University of Lethbridge Scholarship	1996
Louise McKinney Scholarship	1995
Alexander Rutherford Scholarship	1992

UNIVERSITY OF VICTORIA PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain by the University of Victoria shall not be allowed without my written permission.

Title of Thesis:

Have I Remembered this Before? Remembering (and forgetting) of prior Remembering

Author [redacted]

Michelle Marie Arnold

August 20, 2001