Remembrance of remembrance past

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

Arnold and Lindsay (2002) found that individuals more often failed to remember they had previously recalled an item if that item had been cued in a qualitatively different way on two recall occasions: the ''forgot-it-all-along'' (FIA) effect. Experiment 1 was designed to determine if the FIA effect arises because participants incorrectly believe they have not been previously tested for an item, or because they incorrectly believe they have failed to recall the item when previously tested. Experiment 2 measured participants' confidence in their incorrect prior-recall judgements, and Experiment 3 tested participants' ability to ''recover'' their previous recollection when the prior-recall context was restored. Results indicated that participants usually believed they had not previously been cued for the items they failed to remember previously recalling; they were often confident in their incorrect judgements of prior non-remembering; and re-introducing the context of prior remembering sometimes enabled them to recapture their memories of previous recall.

In what may be the literary scene most often cited by memory researchers, the protagonist of *Remembrance of Things Past* experiences a flood of childhood memories after eating a childhood treat (a madeleine cookie soaked in tea). The scene is usually cited as an illustration of the power of distinctive retrieval cues, but what interests us here is that Proust's protagonist believes that his childhood in Combray had long lain dormant and unremembered: ''Many years had elapsed during which nothing of Combray, save what was comprised in the theatre and the drama of my going to bed there, had any existence for me'' (Proust, 1923, p. 58). He notes that the mere sight of the madeleines had not prompted such reminiscences ''... perhaps because [he] had so often seen such things in the interval, without tasting them, on the trays in pastry-cooks' windows, that their image had dissociated itself from those of Combray days to take its place among others more recent'' (Proust, 1923, p. 61).

Reports of past events being forgotten for years before being recollected in later life are not peculiar to fiction. In her work on involuntary memories, Salaman (1970/1982) noted that ''we are all exiles from our past'' (p. 55). In addition to describing many of her own late-life recollections, Salaman cited several examples of autobiographical writers reporting the recovery of long-forgotten memories. Furthermore, claims of prior periods of amnesia are a key element of the phenomenon of recovered-memory experiences, in which adults report newly discovered memories of childhood sexual abuse (e.g., Courtois, 1999; Lindsay & Read, 1994; Porter, Campbell, Birt, & Woodworth, 2003). Conversely, people sometimes make the opposite sort of claim regarding prior reminiscences, reporting that they recollect a particular past event each and every time a special stimulus is encountered. To cite a prosaic example, Conway (2003, p. 218) asserted that every time he goes to the grocer store he recalls the time he failed to bring a coin for a shopping trolley. To the best of our knowledge, this phenomenon has not been studied systematically, but popular songs and literature furnish many examples of claims

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

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through a post-graduate scholarship to Michelle M. Arnold and by NSERC Grant OGPO121316-95 to D. Stephen Lindsay. Portions of this research were reported at the annual meeting of the Northwest Cognition and Memory Society, Vancouver, British Columbia, May 2001.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210444000214

along the lines of ''every time I see/hear/smell X, I remember Y.''

What led Proust's protagonist to believe that he had not previously reminisced about any of those hundreds of pages of childhood events? Do you really remember X every time you encounter Y? More generally, how do people make judgements about their history of recollecting past events?

Memory researchers have long noted that free recall requires an output-monitoring process that enables participants to avoid repeating items they have already reported (Koriat, Ben-Zur, & Sheffer, 1988). For example, Koriat et al. contended that older adults exhibit higher repetition in recall than younger adults, because older adults are less able to monitor their previous recall performance. In related work, Gardiner and co-authors (Gardiner & Klee, 1976; Gardiner, Passmore, Herriot, & Klee, 1977; Klee & Gardiner, 1976, 1980) investigated ''memory for remembered events'' (MRE) by having participants study multiple lists, each followed by a recall or recognition test, and then presenting a final recognition test in which participants were to indicate which items they had previously recalled or recognised on the initial tests.[1] Performance on the final test was generally quite accurate, but was influenced by manipulations that would be expected to modulate the distinctiveness and strength of memory records of the initial memory-test response (e.g., MRE was more accurate when memory had initially been assessed with recall as opposed to recognition tests, or if participants had both spoken and written their initial memory responses rather than only speaking them).

Several theorists have pointed to the role of an output-monitoring process in prospective memory tasks (e.g., Dobbs & Reeves, 1996; Leynes & Bink, 2002; Marsh, Hicks, Hancock, & Munsayac, 2002); participants need to remember that they have already remembered to perform an intended action to avoid needlessly repeating it. Dobbs and Reeves (1996) noted that very little prospective memory research has focused on the role of output monitoring, but that it seems likely that repetitions sometimes occur due to failures to monitor performance. Marsh et al. (2002) reported research demonstrating that repetition errors in prospective memory can occur from a failure in output monitoring, and that such errors can be reduced by elaboratively responding to an intention (e.g., verbalising and pressing a key to a target when it appears, as compared to merely pressing a key). In summary, although research on output monitoring has not yet resolved the issue of how individuals judge whether or not they had previously recollected a particular past event, output monitoring has been implicated by numerous researchers as an important process for accurate memory performance in both prospective and retrospective memory domains.

In the course of studying reports of recovered-memory experiences, Schooler, Ambadar, and Bendiksen (1997a) discovered dramatic evidence that judgements of prior autobiographical remembering can be in error (see also Schooler, 2001; Schooler, Bendiksen, & Ambadar, 1997b). In three of the cases they investigated, a woman who reported a full-blown recovered-memory experience had apparently told another person about the abuse during the period of supposed amnesia. Schooler and co-authors proposed that this reflected the operation of a ''forgot-it-all-along'' (FIA) mechanism that can lead to forgetting of prior instances of recollecting a past event. Their idea was that during the memory-recovery experience the traumatic event was recalled in a qualitatively different way from past occasions of remembering it (e.g., more completely, more episodically, or as abuse *per se* rather than as some more innocuous category of childhood event), giving rise to a strong emotional response which in turn contributed to a spontaneous and unconscious inference that the event must not previously have been recalled (e.g., ''Given the tremendous emotion I am experiencing now, I must not previously have recollected this event''; see also Conway & Pleydell-Pearce's, 2000, related discussion of a self-memory system).

Arnold and Lindsay (2002) developed a laboratory analogue that captures some aspects of Schooler et al.'s FIA mechanism (see also Joslyn, Loftus, McNoughton, & Powers, 2001; Padilla-Walker & Poole, 2002, for related experimental work on the FIA effect). Our idea was that if remembering sexual abuse in qualitatively different ways on two occasions can lead a person to fail to remember the prior instance of remembering, then an analogous effect may occur if individuals are led to recall innocuous laboratory materials in qualitatively different ways on two occasions. In Arnold and Lindsay's Experiment 1, participants studied a list of homophones, each accompanied by a biasing context word (e.g., hand–*palm*, plaster–*cast*). They were then tested on a subset of the study list, with some target items being cued with

[1] We thank Richard L. Marsh for bringing this research to our attention.

the studied-context word (e.g., hand: p***m) and others cued with an other-context word (e.g., theatre: c**t). Thereafter, participants were tested on all of the studied items, this time with the studied-context cues given as recall prompts, and after each recalled word they judged whether or not they had recalled that word on the first test. As predicted, participants more often forgot their prior recall of words cued with other-context words on Test 1 than of words cued with studied-context words on Test 1. We took this result as evidence that remembering a past event in a different way can lead one to fail to remember a prior instance of recalling that event, and proposed that this effect can be understood in terms of the notion of transfer-appropriate processing (TAP) (e.g., Morris, Bransford, & Franks, 1977). Briefly, the idea is that current thoughts about an item will be a poor cue for a prior instance of recalling that item if the current way of thinking about the item differs from how it was thought of during the prior recall.

In three follow-up studies, Arnold and Lindsay (2002) eliminated various alternative explanations for their findings. For example, the results of Experiment 1 might be attributed to weaker Test 1 remembering of items in the other-context condition, rather than to the qualitative mismatch between Test 1 and Test 2 recall of those items. In Experiment 2, studied-context versus other-context cues were manipulated in both recall tests. The important finding was that an FIA effect was obtained for items cued on Test 1 with studied-context cues and on Test 2 with other-context cues (as well as for items cued on Test 1 with other-context cues and on Test 2 with studied-context cues). Experiment 3 showed that the FIA effect is not restricted to manipulations that vary the dictionary definition of the to-be-recalled words (e.g., from hand-*palm* to tree-*palm*) but can also be obtained with more subtle shifts from Test 1 to 2. For example, participants who on Test 1 recalled ''palm'' in response to the cue ''The fortune teller traced the lifeline on the p*** of his hand'', and recalled the same studied item on Test 2 in response to the cue ''He used his p*** to swat the fly'', were less likely to remember their Test 1 recall of that item than were participants who were given the same cues on both tests. Experiment 4 showed that the effect was not merely due to participants judging whether or not they had seen the Test 2 retrieval cues on Test 1 (as opposed to judging whether or not they had recalled the target). To this end, we used multiple short study lists and free recall procedures for Test 1, followed by cued recall in Test 2. Even though there were no retrieval cues in Test 1, participants more often remembered their Test 1 recollection of target words if they were cued in Test 2 with studied-context cues than if they were cued in Test 2 with other-context cues. Collectively, the experiments reported by Arnold and Lindsay provide compelling evidence that people are more likely to forget an instance of past remembering if it differs from their current recollections.

In the experiments reported here, we explored three issues left unresolved in the Arnold and Lindsay (2002) studies. First, when participants fail to remember their Test 1 recollection of an item cued differently on Tests 1 and 2, do they think that they were cued for that item on Test 1 but failed to recall it, or do they think they were not even cued for the item on the first test? Second, when participants falsely judge that they did not recall an item on Test 1, are they confident that they have *not* recalled that item, or are they merely unsure that they recalled it? Recovered-memory experiences are defined, in part, by the fact that those who report them confidently deny prior recollections of the trauma. At least one of the cases described by Schooler and co-authors (Schooler et al., 1997a, 1997b) steadfastly denied prior memories of the abuse, even when confronted with evidence that she had told another person about the abuse during the period of supposed amnesia. Without claiming that our simple laboratory task is a model of recovered memories of childhood trauma, we nonetheless sought to determine whether false reports of ''No, I didn't recall that one on Test 1'' in this paradigm are made with great or little confidence. Finally, given that a person cued differently on Tests 1 and 2 subsequently fails to remember the Test 1 recall episode, might s/he ''recover'' memories of recalling that item if prompted with the Test 1 recall cue? Each of these issues was explored in turn in the following three experiments. The theoretical implications of the findings, and their relationship to other recent memory research, are elaborated in the General Discussion.

EXPERIMENT 1

Arnold and Lindsay (2002) asked participants, for each Test 2 item, whether they had or had not recalled that item during Test 1. When participants responded ''No'' to this question, they may have meant ''No, I don't remember anything about that item on Test 1'' or they may have

meant ''No, I remember that a cue was presented for that item but I was unable to retrieve the item''. Distinguishing between these alternatives is important because they are qualitatively different reasons for claiming no prior remembering of an event. An individual who claims ''I never even previously considered the idea that I might have been abused'' is reporting a qualitatively different sort of non-remembering compared to someone who claims ''I had previously been asked about abuse, but despite that prompt I didn't remember being abused''. Experiment 1 was designed to examine more specifically what participants mean when they incorrectly give a ''No'' response in the prior-memory judgement task, and the parallels between the pattern of ''No'' responses reported here and real-world cases of recovered memories are examined in the discussion of this experiment.

Method

Overview. The prior-recall judgement task used in the Arnold and Lindsay (2002) experiments was modified to give participants three response options: (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).

Participants. A total of 12 University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course.

Materials. The set of 116 homographic target words with two dominant meanings (e.g., *palm* in the part-of-hand sense and the type-of-tree sense) from Experiment 3 of Arnold and Lindsay (2002) was used for Experiment 1. These stimuli were designed to test for the FIA effect by manipulating the context of the target items, but not their sense or meaning (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. 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. Participants were tested individually on an IBM-compatible personal computer using Schneider's Micro Experimental Laboratory Professional 2.0 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 the experimenter would read aloud a target word, followed by a sentence containing that target word. Participants were instructed to repeat the sentence aloud and then write down the target word on a sheet provided by the experimenter.

The first cued recall test immediately followed the study phase, with the items presented in random order. Participants were informed that they would be tested on half of the target words (they were actually tested on two-thirds of the items, but it was easier to explain the task in terms of half of the items) and that this would be done by presenting a context sentence along with the first letter of a target word (e.g., *He swatted the fly with the *** of his hand–p*). Further, participants were instructed that for half of the trials the context sentences would correspond to the context sentences presented with the targets during the study phase, and that for the remaining trials the context sentences would not be the same as at study but that nonetheless the context sentences 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* hearing during the study phase: Participants were instructed to say ''Pass'' if they did not remember the answer or if the answer they came up with was a guess (e.g., filling in the blanks instead of remembering the word from the study phase). Participants were given a hand-held microphone and told that their responses would be recorded on audio-tape. The tape recorder was set so that participants could hear their voices come through the speakers. This arrangement of the tape recorder was used to help participants later understand the difference between only remembering being tested on a target word in Test 1 (i.e., being cued for a target item, but not necessarily recalling that target word) versus remembering recalling the target word. The experimenter keyed in the spoken responses

(coding them as correct, incorrect, or no response) and 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-minute break during which they conversed with the experimenter before moving on to the second cued recall test.

All 108 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 sentence with the first letter of a target word (e.g., *He swatted the fly with the p?? of his hand*) and asked to recall the target word from the study phase. Participants were correctly informed that each of the context sentences on Test 2 corresponded to a context sentence presented with a target word during the Study phase. As in Test 1, participants were told to respond with an answer only if they remembered hearing the target word during the Study phase. After each recall attempt the three judgement options were displayed on the screen, 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. For the judgement 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. Finally, participants were instructed to say ''not cued'' if they believed the item had not been tested for on Test 1. Participants were explicitly instructed that their judgements should not be based on whether or not they had seen the Test 2 context sentence 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 therefore could not have been recalled) on Test 1. The experimenter emphasised that the task was not to judge if the context sentence 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 'cued but not recalled'.''). The experimenter also told participants that another way to think of the judgement task involved the tape-recording from Test 1–''If the tape recording was played back, would you hear your voice saying the target word?'' Participants were instructed to say ''cued and recalled'' only if they could remember recalling and saying the target word on Test 1. During the test, participants were stopped two or three times and reminded of the instructions for the judgement task.

Results and discussion

Although our interest focused on participants' judgements about prior remembering, in this and the subsequent Results and discussion sections we first report analyses of recall performance on Test 1 and Test 2 (collapsed across the counterbalancing factors of study list and test list)[2] before presenting the analyses of the data for the prior-remembering judgement.

Recall performance. Proportion correctly recalled on Test 1 was significantly higher for items in the studied-context condition ($M = .93$) than items in the other-context condition ($M = .81$), $F(1, 11) = 35.31$, $MSE = .003$, partial $\eta^2 = .76$, $p < .0001$. A within-subjects analysis of variance (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 reliable difference in proportion of items recalled for the studied-context ($M = .95$), other-context ($M = .95$), and not-tested ($M = .93$) conditions, $F(2, 22) = 1.44$, $MSE = .002$, $p = .26$.

Judgement of previous recollection and cueing. The proportions of items on Test 2 judged as recalled on Test 1 are shown in Table 1.

[2] The analyses of recall performance for all experiments were initially performed using omnibus ANOVAs that included the study list and test list counterbalancing factors. There were sometimes significant effects of the study list and/or test list factors (although the pattern was not consistent across experiments). It is likely that by chance of random assignment some of the conditions contained more memorable senses of the target words (study list), and sometimes led one version of the test list to be easier than the other (although some or all of these inconsistent counterbalancing effects may be Type I errors). Regardless of the reasons, differences in recall performance due to the counterbalancing of the study/test lists are not pertinent to our present interests and we do not report these analyses here; they may be requested from either author.

TABLE 1
Experiment 1

Test 1 cue	*Test1/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	8 (–)	1.00 (–)
	Not recalled/Recalled	1.08	.04 (.04)	.85 (.10)
	Recalled/Recalled	**33.25**	**.92 (.02)**	.01 (.004)
	Recalled/Not recalled	.17	.50 (.50)	.50 (.50)
Other-context	Not recalled/Not recalled	.92	0 (–)	.38 (.16)
	Not recalled/Recalled	6.08	.08 (.04)	.17 (.04)
	Recalled/Recalled	**28.00**	**.62 (.04)**	.03 (.01)
	Recalled/Not recalled	1.00	.26 (.15)	.21 (.15)
Not-tested	NA/Not recalled	2.50	0 (–)	.13 (.06)
	NA/Recalled	33.50	.02 (.01)	.04 (.01)

Mean number of items and mean 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 1.

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

The analyses reported here were performed on the judgement data for target 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 study list or test list factors for the judgement task were found (all Fs < 2.85, p $\geq$ .14), and therefore the data were collapsed across these variables in the subsequent analyses.

A within-subjects ANOVA was performed on the proportion of "cued and recalled" judgements, 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, $F(1, 11) = 52.01$, $MSE = .01$, partial $\eta^2 = .83$, $p < .0001$. This finding replicates our FIA effect (Arnold & Lindsay, 2002).

The average number of items that participants correctly recalled on both tests but failed to remember previously recalling was 2.58 for the studied-context condition and 10.67 for the other-context condition. The vast 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-context (M = .95) and other-context (M = .91) conditions. Further, there was no significant difference between the proportion of "not cued on Test 1" judgements for the studied-context and other context conditions, $F < 1$. Finally, participants almost always correctly responded "not cued" to the items in the not-tested condition (M = .96).

These results indicate that when participants forgot they had previously recalled an item, they usually mistakenly believed they had not even been tested for that item. This outcome makes sense from a TAP point of view: At Test 2, participants fail to remember the Test 1 cue for the same reason that they fail to remember their Test 1 recall of that item, namely because the way they are currently thinking about the item doesn't map on to the way they thought about it on Test 1.

Although this TAP account is plausible, it is also possible that the low rate of "cued but not recalled" responses to other-context items merely reflects probability matching.[3] Specifically, participants may have seldom responded with "cued but not recalled" simply because they knew that they had rarely failed to recall items for which they were cued on Test 1. That is, participants might have thought something like "I have no idea whether I was cued to recall this item on Test 1, but I recalled almost all of the items for which I was cued on Test 1, so given that I don't remember recalling this item on Test 1 I probably wasn't cued for it."

Participants' beliefs regarding their Test 1 performance probably did influence their prior-cueing judgements in ways consistent with this probability-matching account. More specifically,

[3] We thank an anonymous reviewer for suggesting the idea of probability matching as another explanation for the judgement data.

knowing that they had recalled most of the items for which they were cued on Test 1 very likely discouraged participants from making a large number of ''cued but not recalled'' responses. However, probability matching does not provide a complete account of the low rate of ''cued but not recalled'' responses on other-context items that participants forgot they had recalled. Participants almost always correctly responded ''cued but not recalled'' when asked about the few studied-context items that they had in fact failed to recall on test 1 (as shown in Table 1). This result shows that prior-cueing judgements were not based solely on probability matching. Rather, it appears that those judgements were based largely on whether or not participants could bring to mind evidence of having been cued for the item on Test 1.

How do our results compare to real-world cases of the FIA effect like those described by Schooler and colleagues (e.g., Schooler, 2001; Schooler et al., 1997b)? Our intuition is that individuals who believe they were amnesic for abuse for a period during which they in fact told another person about the abuse would usually fail to recall the entire conversational context in which they had previously reported the abuse, not just their report of the abuse itself. That is, just as our participants denied memories of even being asked about the target word ''palm'' if the way they were currently thinking about ''palm'' differed from the way they had previously thought about it, so real-world cases of the FIA effect would likely fail to recall the specific conversations and other cues that led them to report the abuse during the period of time they now believe they were amnesic for it. Consistent with this intuition, one of the FIA case studies presented by Schooler (2001; Schooler et al., 1997b) reportedly denied ever discussing the topic of abuse with the person to whom she had disclosed the abuse. We will return to the issue of the relationship between our experiments and real-world cases of recovered memories in the General Discussion.

EXPERIMENT 2

Experiment 2 explored the subjective phenomenology of failures to remember prior remembering. In our earlier studies, participants were required to make a yes/no judgement as to whether or not they had previously recalled each item. What did participants mean when they responded ''No''? Were they merely expressing a low level of confidence (e.g., ''I don't have a clear memory of recalling this item, so I guess I probably didn't'') or were they fairly sure that they had not recalled the item? If our experimental paradigm parallels the FIA effect described by Schooler (2001), then participants should, at least some of the time, be quite confident in their incorrect belief that they had not previously recalled a context-change item. In contrast, if participants are merely guessing that they had not previously recalled the context-change items, then the laboratory FIA effect becomes a less interesting phenomenon. Further, the issue of confidence is also important because it potentially speaks to the robustness of the FIA effect. More specifically, if the effect is mainly the result of guessing (i.e., participants act conservatively by choosing the ''not recalled'' option when they are unsure of an other-context item), then reducing or eliminating the FIA effect should be easier than if participants are highly confident in their incorrect ''No'' responses to the other-context items. To explore this issue we added a confidence rating to the prior-recall judgement task.

Method

Participants. A total of 12 University of Victoria undergraduates participated in exchange for optional extra credit in an introductory psychology course.

Materials and procedure. Experiment 2 used the same list of target words and context sentences as Experiment 1. Further, Experiment 2 followed the same basic procedure of Experiment 1, with two alterations. The first adjustment involved changing the judgement task from the three-alternative forced-choice task used in Experiment 1 (cued and recalled, cued but not recalled, and not cued) to the two-alternative choice task used in previous FIA experiments. That is, participants were instructed to say ''Yes'' to the judgement task if they remembered recalling and saying the target word on Test 1, and ''No'' if they believed they had not recalled the target word or had not been tested for the target word on Test 1. The second modification involved the addition of a confidence rating task. After making each ''Yes/No'' judgement of prior remembering decision, participants were shown a screen with a 6-point confidence scale; (1) very low, (2) quite low, (3) somewhat low, (4) somewhat high, (5) quite high,

and (6) very high. Participants were instructed to choose the option that best described their confidence for the judgement 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 and discussion

Recall performance. Proportion correctly recalled on Test 1 was significantly higher for items in the studied-context condition (M = .93) than items in the other-context condition (M = .80), $F(1, 11) = 28.00$, $MSE = .004$, partial $\eta^2 = .72$, $p < .0001$. A within-subjects ANOVA was performed on the proportion correctly recalled on Test 2, with context on Test 1 (studied, other, or not-tested) as the within-subjects factor. There was no significant difference in proportion of items recalled for the studied-context (M = .94), other-context (M = .93), and not-tested (M = .92) conditions, $F < 1$.

Judgement of previous recollection. The proportions of items on Test 2 judged as recalled on Test 1 are shown in Table 2. The analyses reported here were performed on the judgement data for target items correctly recalled on both Test 1 and Test 2 (shown in bold in Table 2), 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 study list or test list factors for the judgement 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'' judgements as a function of context on Test 1 (studied vs other). 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, $F(1, 11) = 30.75$, $MSE = .01$, partial $\eta^2 = .74$, $p < .0001$. Participants rarely incorrectly claimed to have recalled not-tested items on Test 1.

A frequency distribution of the confidence ratings for items in the studied- and other-context conditions (collapsing across participants) that were incorrectly judged as not recalled on Test 1 is shown in Figure 1. As Figure 1 demonstrates, participants were often confident about their incorrect ''No'' judgements 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'' judgements in the high range of the confidence scale (compared to 41.1% in the studied-context condition). These confidence data demonstrate that participants were not simply choosing the ''No'' option for the judgement 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 suggest that par-

TABLE 2
Experiment 2

Test 1 cue	*Test1/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 (.10)
	Not recalled/Recalled	.92	.19 (.13)
	Recalled/Recalled	**32.92**	**.90 (.03)**
	Recalled/Not recalled	.50	.50 (.29)
Other-context	Not recalled/Not recalled	1.25	.05 (.05)
	Not recalled/Recalled	6.00	.04 (.02)
	Recalled/Recalled	**27.50**	**.63 (.05)**
	Recalled/Not recalled	1.25	.13 (.08)
Not-tested	NA/Not recalled	2.92	.08 (.06)
	NA/Recalled	33.08	.06 (.02)

Mean number of items and mean proportion of items judged as ''recalled'' and as a function of recall status on Test 1 and Test 2 for Experiment 2.

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

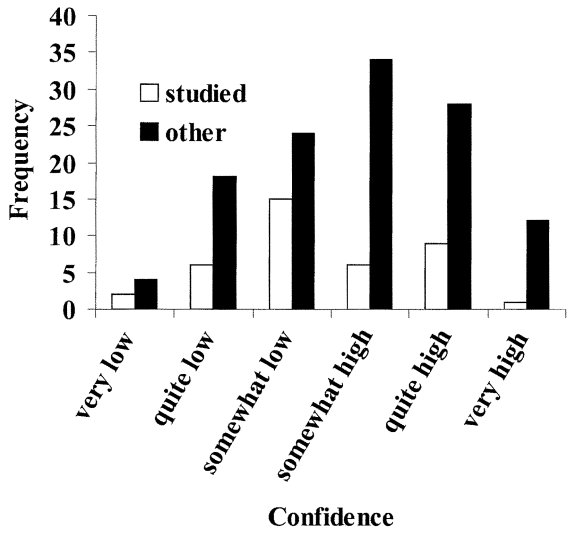

Figure 1. Confidence ratings for the total number of items incorrectly judged as not recalled on Test 1 (collapsed across participants) in the studied- and other-context conditions.

ticipants in our experiments are often quite confident in their erroneous belief that they had not previously recalled an item.

We conducted a sub-analysis restricted to items for which participants claimed to have "somewhat high", "quite high", or "very high" confidence in their judgements regarding previous recollection. The effect of context cues on judgements 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) than if they had been cued with the studied-context word on Test 1 (M = .95), $F(1, 11) = 23.48$, $MSE = .02$, partial $\eta^2 = .68$, $p = .001$. A further analysis was restricted to the two highest confidence ratings ("quite high" and "very high" confidence). The data from two participants were excluded from this analysis because fewer 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 judgements 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) than if they were cued with the studied-context word on Test 1 (M = .98), $F(1, 9) = 8.80$, $MSE = .02$, partial $\eta^2 = .49$, $p < .02$.

To reiterate, Experiment 2 demonstrated that the FIA effect is not simply due to guessing: When participants incorrectly claimed that they had not previously recalled an other-context item, they were often quite confident that they had not previously remembered that item. This result would be less interesting if participants were just generally prone to use the high confidence options, but this behaviour was not the case. For example, in over half of the trials on which studied-context items were incorrectly judged as not previously recalled, participants were on the low side of confidence (showing that participants were willing to use the full range of confidence ratings).

The high confidence with which participants often claimed not to have previously recalled other-context items that they had in fact recalled suggests that the FIA effect may be difficult to eliminate. That is, given that participants are often confident when they incorrectly judge that an other-context item had not been recalled on Test 1, it may be difficult to dissuade them. Yet, as explained in the next paragraph, the TAP account suggests that it may be possible to lead participants to recapture recollections of prior remembering. We began to explore this issue in the final experiment.

EXPERIMENT 3

The third experiment was constructed to investigate whether participants could recapture their Test 1 recollections for items that they had failed to remember previously recalling. More specifically, the question of interest centred around reinstating prior context: After failing to recall their previous recollection of other-context items, is it possible for participants to "recover" their Test 1 recall if they are prompted with the context in which they had been tested in Test 1? This idea of reinstating Test 1 context is related to research on recognition failure of recallable words. In the classic example of this phenomenon, Tulving and Thomson (1973, Experiment 2) had participants study target words paired with weakly associated context words (e.g., glue – TABLE). When the target words were subsequently paired with strong associates (e.g., chair – TABLE) and presented on a yes/no recognition test, participants often failed to recognise them, even though they were often later able to recall the target words when the studied weak associates (e.g., glue) were presented as retrieval cues. Thus an item participants failed to recognise when the study and test contexts differed could later be recalled when the study and

test contexts were reconstituted. Likewise, we expected that when participants forgot during Test 2 that they had recalled an other-context item on Test 1, they would sometimes recover memories of that prior instance of recollection when provided with the studied-context cue.

As in Arnold and Lindsay's (2002) Experiment 4, during Test 1 participants indicated, for each word freely recalled, whether they recollected episodic details of studying that word (''remember'') or merely ''knew'' that the word had been studied (as in Gardiner & Java, 1990). Forgetting of prior instances of remembering is arguably more interesting when those prior recalls were characterised as ''remember'' rather than ''know'' experiences. Arnold and Lindsay found that the FIA effect was maintained in a sub-analysis restricted to items participants had previously classified as ''remembered'', and we sought here to replicate that finding. Relatedly, we expected to find a smaller FIA effect in Experiment 3 than in the previous two experiments because the free recall task is rather memorable. That is, the items correctly recalled are those brought to mind by the participants themselves, and participants had made a Remember/Know judgement for each recalled item (thereby increasing its memorability). Additionally, because the first test was free recall, there was no opportunity for the familiarity of cues to add to the FIA effect.

Method

Overview. This experiment used a free-recall procedure for Test 1 (as in Arnold & Lindsay, 2002, Experiment 4). Participants studied six lists of 12 homographic target words presented in disambiguating context sentences. After each study list was presented, participants attempted free recall for the target words (Test 1). For each word recalled on Test 1, participants indicated whether they ''remembered'' studying that word or merely ''knew'' that the word had been in the study list. Later, participants were given a cue to recall each of the target words (Test 2); half of the items were cued with context words relating to the studied sense of the targets (*studied-context* condition) and half were cued with context words that did not relate in sense to the sentences presented with the target items at study (*other-context* condition). Additionally, the second test required participants to complete two judgements regarding whether they remembered recalling the word on the free-recall test. Both judgements were exactly the same in nature, but the second judgement was done in the presence of the context that had not been used to prompt Test 2 recall (i.e., other-context targets tested with context words relating to the studied context, and studied-context targets paired with context words not related to the studied sense).

Participants. A total of 28 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 condition on the cued recall test that they had also previously recalled in free recall, and one participant failed to follow instructions for both the recall and judgement tasks—their data were dropped from the analyses.

Materials. A list of 82 target words was constructed from the set of homographic target words used in Experiment 4 of Arnold and Lindsay (2002). These stimuli were constructed so that the senses of the study sentences were not changed from those used in the previous FIA experiments, but the context words themselves were dropped from the sentences. For example, in the prior two experiments the target word ''palm'' was sometimes studied with the sentence ''He used the palm of his hand to swat the fly'', and the recall of ''palm'' was sometimes cued with the context sentence relating to part-of-hand (i.e., the same sentence that participants studied with the target word). In this experiment, in contrast, some participants studied ''palm'' with the context sentence ''He used his palm to swat the fly'', and in Test 2 ''palm'' was cued with either ''hand – p_ _ m'' or ''tree – p_ _ m''. Consequently, in both of the context conditions, participants were never exposed to the context words (e.g., ''hand'') until the final cued-recall test. Of the 82 items, 10 were used as a practice list. Two study lists were constructed to counterbalance the meanings of the target words between subjects (*study list* factor). For each participant, the computer randomly divided the 72 target words into six lists of 12 items. The target words were randomly partitioned into two lists of 36 words (*test list* factor), with each list appearing equally often across participants in the studied-context and other-context within-subject conditions of Test 2.

Procedure. The study phase was modified slightly from that used in Experiment 1 and 2.

Participants were told that on each trial the experimenter would read a sentence aloud, and that one of the words would be verbally emphasised as a target word. Their task was to repeat the sentence aloud, making sure that they also verbally emphasised the target word. Participants were instructed to pay close attention to each sentence as it was being read so that it would not need to be repeated, but if a participant was unable to repeat a sentence or failed to emphasise the target word the experimenter repeated the sentence (participants rarely needed a sentence to be repeated). The experimenter informed participants that they would have to recall the target words after each list of 12 sentences was presented. Additionally, the experimenter cautioned participants to write down a word only if they were confident that they had heard that item as a target word in the study list. Beside each target word that they recalled, participants were required to make a Remember-Know judgement (as in Gardiner & Java, 1990). Participants were asked to write an ''R'' beside the recalled word if they could remember something about the experience of having studied that word (e.g., remember the sound of the target word, images it evoked, etc.), and to write a ''K'' beside the recalled word if they knew the item was in the study list but could not recollect anything specific about studying that word. For each list, participants were given 1.5 minutes to recall as many target words as possible and complete the Remember-Know judgements. The practice list was administered before starting the first critical list to make sure that participants understood the tasks. After completing recall of the final study list, participants were given a 10-minute break during which they conversed with the experimenter before moving on to the cued-recall test.

All 72 of the critical target words were tested for in random order with cued recall on Test 2. Each trial began with a context word and the first and last letters of the target word presented on the computer screen; participants were instructed to recall the corresponding studied target word, and to respond with an answer only if they were confident that they had heard that word as a target item during study. Participants were told that the context words for half of the trials 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 help participants understand these instructions they were given an example of a studied-context and other-context word that could be used to probe for a target word (e.g., ''You studied the target word 'palm' with 'He used his palm to swat the fly'. We could test you for this item 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'.''). After each item was recalled (or, if recall was unsuccessful, the experimenter told the participant the item) participants were asked to judge whether they had free-recalled that target word during Test 1; judgement instructions were the same as those used in Experiment 2. After the first judgement task was completed, participants were shown the context word that had not been used to cue recall on Test 2 and were asked to perform the prior-recall judgement again. The experimenter explained to participants that the context word used to cue recall might not be enough to induce remembering for previous recall, and that they should think of the second context word as additional information that might help them with the judgement task.

Results

Recall performance. Participants occasionally free-recalled a target word from a list other than that on which they were being tested (an average of .42 items across participants), and these items were scored as correct as long as they were not items that had been recalled in an earlier list. On average, participants produced 41.25 target words (57.29%) and 1.42 incorrect intrusions in free recall. Finally, participants classified an average of 65.14% of the recalled target events as ''Remember''. On the final test, proportion correct cued recall was reliably higher for items in the studied-context condition ($M = .88$) than items in the other-context condition ($M = .84$), $F(1, 23) = 9.16$, $MSE = .003$, partial $\eta^2 = .23$, $p = .01$.

Judgement of previous recollection. The proportions of items judged following cued recall (Test 2) as having also been recalled during free recall (Test 1) are shown in Table 3. The analyses reported here were performed on the judgement data for target items correctly recalled in both free and cued recall (shown in bold in Table 3), although the same pattern of results was found

TABLE 3
Experiment 3

Test 2 cue	*Test1/Test 2 recall status*	*Mean number of items*	*Mean proportion judged as "recalled" on Test 1 for Judgement 1*	*Mean proportion judged as "Recalled" on Test 1 for Judgement 2*
Studied-context	Not recalled/Not recalled	2.17	.09 (.04)	.11 (.04)
	Not recalled/Recalled	13.13	.15 (.03)	.16 (.03)
	Recalled/Recalled	**18.71**	**.88 (.02)**	**.89 (.02)**
	Recalled/Not recalled	2.00	.84 (.06)	.87 (.05)
Other-context	Not recalled/Not recalled	3.17	.16 (.06)	.18 (.06)
	Not recalled/Recalled	12.29	.15 (.03)	.16 (.03)
	Recalled/Recalled	**17.79**	**.80 (.03)**	**.83 (.03)**
	Recalled/Not recalled	2.75	.69 (.07)	.77 (.07)

Mean number of items and mean proportion of items judged as "recalled" and as a function of recall status in Test 1 and Test 2 for Experiment 3.

Lines in bold are those for which statistical analyses are reported in the manuscript. There were 36 items per condition in Test 2. Standard error of the means are in parentheses.

when analyses were contingent only on correct free recall on Test 1. An initial omnibus ANOVA showed no effects of the study list or test list factors for the judgement task (all Fs < 3.90, ps $\geq .06$), and therefore the data were collapsed across these variables. The proportion of correct "Yes" judgements was analysed in a 2 (Prior-remembering judgement: judgement 1 vs judgement 2) $\times$ 2 (Test 2 context: studied vs other) within-subjects ANOVA. Overall, participants were significantly more likely to fail to remember that they had free-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, $F(1, 23) = 6.81$, $MSE = .02$, partial $\eta^2 = .23$, $p < .02$. Planned comparisons showed that this effect of initial context on Test 2 (i.e., other-context cues) was reliable for both the first judgement, $F(1, 23) = 8.42$, $MSE = .01$, partial $\eta^2 = .27$, $p < .01$, and the second judgement, $F(1, 23) = 4.89$, $MSE = .04$, partial $\eta^2 = .18$, $p = .04$. There was also a main effect of judgement, in that participants were reliably more likely to claim prior remembering of the target items on the second judgement task ($M = .86$) than on the first judgement task ($M = .84$), $F(1, 23) = 11.25$, $MSE = .001$, partial $\eta^2 = .33$, $p < .01$.

As anticipated, however, there was a significant interaction between the judgement tasks and context, $F(1,23) = 5.25$, $MSE = .001$, partial $\eta^2 = .19$, $p = .03$: Participants were significantly more likely to fail to recall their Test 1 recollections in the other-context condition on the first judgement task than after context was reinstated on the second judgement task, $F(1, 23) = 9.22$, $MSE = .001$, partial $\eta^2 = .29$, $p < .01$. There was no reliable difference between judgement tasks 1 and 2 for the studied-context condition, $F(1, 23) = 3.22$, $p = .09$, thereby showing that the additional context presented for the second judgement task was effective in increasing correct performance on the second previous recall judgement only for the other-context items. Finally, 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 during Test 1.

We conducted a sub-analysis for the first prior-recall judgement task restricted to items that participants had claimed to "remember" rather than "know" during the Test 1 free recall. We excluded the data from three participants because they assigned a "remember" response to fewer than eight items per condition. For the remaining participants, the effect of Test 2 context on judgements of prior recall was the same as in the analyses reported above. That is, participants more often forgot that they had recalled "remembered" items if the items were cued with the other-context word on Test 2 ($M = .70$) than if they were cued with the studied-context word ($M = .82$), $F(1, 20) = 9.85$, $MSE = .02$, partial $\eta^2 = .33$, $p < .01$. This pattern of results for "remembered" items also was found for the studied-context ($M = .83$) and other-context conditions ($M = .72$) of the second prior-recall judgement task, , $F(1, 20) = 6.73$, $MSE = .02$, partial $\eta^2 = .25$, $p < .02$.

The central finding of Experiment 3 was that, as predicted, reinstating Test 1 context helped participants to recapture their prior recall of the other-context items. This effect is particularly

interesting given the evidence from Experiment 2 that participants are often quite confident in their erroneous FIA reports regarding other-context items. Interestingly, however, reinstating Test 1 context did not completely eliminate the FIA effect. That is, when participants recalled an item in response to an other-context cue on Test 2 and failed to remember that they had previously recalled that item on Test 1, the subsequent presentation of the studied context word sometimes enabled them to remember recalling the item on the prior test, but it did not always do so. Perhaps the studied-context cue, presented immediately after the other-context cue, did not lead participants to think about the target word solely in its studied sense, but rather led them to think of the word in both senses. From a TAP perspective, this polysemous way of thinking about the word would be a better cue for their prior-recall than would thinking about it solely in the other-context sense (leading to the significant reduction in the FIA effect observed from judgement 1 to judgement 2), but it might be a less effective cue than thinking about the studied-context sense alone (in which case it would not entirely eliminate the FIA effect). Future research may develop more powerfully effective cues to help reinstate prior context, perhaps by more widely separating the presentation of other-context and studied-context cues or by using more detailed and elaborate cues (e.g., the context sentences used in Experiments 1 and 2, as opposed to the single-word context items used in this experiment).

GENERAL DISCUSSION

As expected, an FIA effect was found in all three of the present experiments: Participants were more likely to forget that they had previously remembered an item if they were cued in a different way in the two recollective episodes. Further, the results indicated that: (a) when participants forgot that they had previously recollected an item, they almost always incorrectly believed that they had not previously been tested for that item (as opposed to believing that they had been cued for the target item but that they had been unable to recall that item), (b) participants were often quite confident in their mistaken belief that they had not previously recollected an item, and (c) reinstating the context of a prior recall episode sometimes enabled participants to recapture their memory of previously recalling that episode.

FIA in the lab versus the real world

As we have noted previously (Arnold & Lindsay, 2002, p. 528), participants in the FIA paradigm are unlikely to experience a spontaneous and powerful feeling during Test 2 that they had not recalled an other-context item on Test 1. That is, they are unlikely to think ''Wow, 'palm!' I didn't remember 'palm' before!''. In contrast, real-world recovered-memory experiences are defined, in part, by the rememberer's spontaneous and powerful conviction that the events had not previously been recollected. The case studies reported by Schooler and colleagues (1997a, 1997b) indicate that this conviction is sometimes misplaced, but the point remains that our laboratory procedure does not capture this aspect of recovered-memory experiences.

A related point is that participants in our paradigm probably do not believe they were ''amnesic'' for words that they do not remember recalling on Test 1. On the contrary, the results of Experiment 1 indicate that they believe they had not been cued for most of those items (and, presumably, that they probably would have recalled them had they been cued). Of course, those beliefs were erroneous (i.e., they HAD recalled the words in question), but the point we want to emphasise here is that there is little reason to believe our participants thought they had been ''amnesic'' for words they didn't remember recalling. Here again, this appears to contrast with real-world recovered-memory experiences, in which individuals may believe that memories of trauma had long been inaccessible (perhaps due to the operation of a special mechanism such as dissociation or repression).[4]

[4] In fact, although a number of studies have sought to estimate the prevalence of forgetting and/or recovering memories of childhood trauma (for contrasting reviews, see Brown, Scheflin, & Hammond, 1998; Read & Lindsay, 2000), it is not clear what individuals mean when they report a prior period of non-remembering. In a study by Read (1997) of 83 individuals who reported an experience that they had ''forgotten about'' for an extended period of time, only one person selected a response option indicating ''amnesia'' for that event, and more than two thirds indicated that they would have remembered the ''forgotten'' event if directly asked about it during the period of non-remembering. Similarly, one of the FIA case studies presented by Schooler (2001; Schooler et al., 1997b) indicated that she believed it possible that she would have remembered the abuse if she had been asked about it during the period of supposed amnesia. Nonetheless, it is clear that individuals who report recovered-memory experiences often believe that their memories of trauma had long been inaccessible to them (e.g., R. E. Cheit, personal communication, October 10, 1997).

These differences in subjective experience between real-world and laboratory FIA effects do not necessarily reflect the operation of fundamentally different mechanisms in the two phenomena. Rather, the differences may arise from differences in how the cognitive system interprets evidence of prior non-remembering of real-life trauma versus of words. That is, the differences may have to do with content rather than with mechanism.

Many current models posit that autobiographical remembering involves dynamic reconstructive processes in which the content of information generated in response to cues interacts with the individual's beliefs and expectations. For example, Conway and Pleydell-Pearce (2000) proposed a self-memory system in which autobiographical memories are constructed via interactions between currently active goals (the ''working self''), expectations about what is to be retrieved, and a distributed autobiographical knowledge base. From this perspective, the experience of recovering memories of childhood sexual abuse would have profound implications for the self-memory system, whereas recovering memories of study-list words would not, even if the same basic mechanism was involved in giving rise to the belief that memories had been recovered.

Similarly, models that emphasise the role of inferential processes in the phenomenology of remembering (e.g., Jacoby, Kelley, & Dywan, 1987; Johnson, Hashtroudi, & Lindsay, 1993; Ross & Wilson, 2003; Whittlesea & Williams, 1998) would predict qualitative differences in the attributions (including affective responses) made to memorial evidence of prior non-remembering of abuse versus of words. Indeed, Schooler and colleagues speculated that the emotional intensity of memories of abuse may contribute to the belief that the abuse had not previously been remembered: Rememberers make an inference along the lines of ''If I had remembered the abuse before, then it wouldn't be such a powerfully emotional and overwhelming memory now'' (Schooler et al., 1997a, 1997b). Unlike memories of abuse, the stimuli used in the FIA laboratory paradigm do not lend themselves to this sort of spontaneous, after-the-fact attributional process. Thus, although memories of abuse differ in many ways from memories of words, it may be that the same basic mechanisms underlie the real-world and laboratory FIA effects. Different combinations of information used by the same mechanisms can produce similar results (e.g., the FIA effect under different circumstances) with qualitatively different experiences.

Related research and theory

The experimental procedure we have used to investigate the FIA effect shares many characteristics with the recognition failure paradigm (e.g., Tulving & Thomson, 1973). The main distinction between the recognition failure paradigm and the FIA paradigm is researchers' focus on the type of event that participants fail to remember. In the FIA 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 recognise the item itself (although both effects could conceivably be produced within the same paradigm). The dominant explanation for recognition failure is based on the encoding specificity (ES) 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). Relatedly, the TAP hypothesis states that the closer the match between the processes used at encoding and retrieval, the better performance will be at retrieval (see Franks, Bilbrey, Lien, & McNamara, 2000; Morris et al., 1977; Rajaram, Srinivas, & Roediger, 1998, for a more detailed discussion of the TAP hypothesis).

According to ES/TAP the various factors proposed to have an impact on memory performance are not restricted to either the encoding stage or the retrieval stage. Instead, ES and TAP emphasise the importance of the interaction between encoding and retrieval conditions (Franks et al., 2000; Tulving & Thomson 1973). Thus, for example, participants in the recognition failure paradigm failed to recognise 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 cueing of the experience of studying that item as a target. As briefly mentioned in the introduction, this interpretation of recognition failure readily applies to the FIA effect. When a 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, and hence participants judge that they

had not previously recalled the word (much as participants in a recognition experiment would reject as new a test item that does not evoke sufficient evidence of oldness; cf. Ghetti, 2003; Rotello, Macmillan, & Van Tassel, 2000).

Using an associated word list paradigm, Marsh and Hicks (2001) demonstrated that individuals can produce false memories of recalling items that they incorrectly believe they had previously encountered. In a typical associated word list paradigm (aka 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); participants frequently recall the non-presented critical lures, and they often report that they actually remember studying those lures (Roediger & McDermott, 1995). Marsh and Hicks (2001) modified the associated word list paradigm to include a second memory test and judgement task. After studying and free-recalling each of six associated word lists, participants were given a recognition test for each individual list 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. The results of interest 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. This outcome will be referred to subsequently as the *false-memory-for-false-memory* effect.[5]

[5] Marsh and Hicks (2001) recognised that the unbalanced judgement scale for the recognition task (two options for studied, one option for new) could be a confound. In a subsequent experiment, they used a 4-point judgement 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 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.

The false-memory-for-false-memory effect shares many similarities with the FIA 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 similar: Participants are required to make judgements 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 centres around the stimuli used to test for the effects. For the FIA 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 contribute to the evocation of a global context (corresponding to the non-presented critical lure) that connects the items and theme of the list together. Conversely, in the FIA 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 likely an important factor that leads to what looks like opposite effects (i.e., false-memory-for-false-memory vs forgetting of prior remembering). The strengthening of a global context across episodes may result in a false belief that a related item (critical lure) was previously recalled, whereas a change in context between episodes of remembering may result in a false belief that an item was not previously recalled.

At the most general level of analysis, our findings suggest that remembering a prior episode of recollection is equivalent to remembering other sorts of prior episodes. Hence retrieval of evidence of prior remembering follows the same principles that govern retrieval of evidence of other sorts of episodic memories (e.g., ES or TAP). We also expect that, as indicated by the March and Hicks (2001) findings, the experience of recollecting a prior instance of remembering is influenced by the same sorts of reconstructive and attributional processes that contribute to recollections of other sorts of events (e.g., source monitoring).

It may seem obvious that remembering prior remembering is analogous to remembering other

sorts of prior events. Yet researchers sometimes treat memory for prior remembering as though it were peculiarly reliable, by reporting self-report measures of rehearsal as though they can be taken at face value. In a study of ''permastore'' (forgetting curves that asymptote well above zero for very well-learned material), Bahrick (1984) found that participants' reports of how often they had rehearsed their high-school Spanish vocabulary over the intervening years were not correlated with their retention of that vocabulary, and concluded that ''the data therefore reflect no important influence of the rehearsal variables'' (p. 12). Some researchers studying flashbulb memories have made similar claims (e.g., Brown & Kulik, 1977). Our results indicate that self-reports concerning prior remembering/rehearsal cannot be assumed to be accurate, even if participants appear to be highly confident in them: Differences between the current testing situation and prior episodes of rehearsal may lead individuals to forget the latter. More generally, our results suggest that Proust's character may have been mistaken in his belief that he had long forgotten most of his childhood, and that (notwithstanding their beliefs to the contrary) even the most ardent romantics may sometimes encounter X without thinking of Y.

Manuscript received 5 August 2003
Manuscript accepted 16 April 2004

REFERENCES

Arnold, M. M., & Lindsay, D. S. (2002). Remembering remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 521–529.

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.

Brown, D., Scheflin, A. W., & Hammond, D. C. (1998). *Memory, trauma treatment, and the law*. New York: W. W. Norton & Co, Inc.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*, 73–99.

Conway, M. A. (2003). Cognitive-affective mechanisms and processes in autobiographical memory. *Memory*, *11*, 217–224.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*, 261–288.

Courtois, C. A. (1999). *Recollections of sexual abuse: Treatment principles and guidelines*. New York: W. W. Norton.

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

Dobbs, A. R., & Reeves, M. B. (1986). Prospective memory: More than memory. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 199–225). Mahwah, NJ: Lawrence Erlbaum Associates Inc. Q1

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.

Gardiner, J. M., & Klee, H. (1976). Memory for remembered events: An assessment of output monitoring in free recall. *Journal of Verbal Learning and Verbal Behavior*, *15*, 227–233.

Gardiner, J. M., Passmore, C., Herriot, P., & Klee, H. (1977). Memory for remembered events: Effects of response mode and response-produced feedback. *Journal of Verbal Learning and Verbal Behavior*, *16*, 45–54.

Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language*, *48*, 722–739.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Lawrence Erlbaum Associates Inc. Q2

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. (2001). Memory for memory. *Memory & Cognition*, *29*, 789–797.

Klee, H., & Gardiner, J. M. (1976). Memory for remembered events: Contrasting recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *15*, 471–478.

Klee, H., & Gardiner, J. M. (1980). Remembering the recall of cued and uncued words: Effects of initial accessibility. *Canadian Journal of Psychology*, *34*, 220–226.

Koriat, A., Ben-Zur, H., & Sheffer, D. (1988). Telling the same story twice: Output monitoring and age. *Journal of Memory and Language*, *27*, 23–39.

Leynes, P. A., & Bink, M. L. (2002). Did I do that? An ERP study of memory for performed and planned actions. *International Journal of Psychophysiology*, *45*, 197–210.

Lindsay, D. S., & Read, J. D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology*, *8*, 281–338.

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

Marsh, R. L., Hicks, J. L., Hancock, T. W., & Munsayac, K. (2002). Investigating the output monitoring component of event-based prospective memory performance. *Memory & Cognition*, *30*, 302–311.

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

Padilla-Walker, L. M., & Poole, D. A. (2002). Memory for previous recall: A comparison of free and cued recall. *Applied Cognitive Psychology*, *16*, 515–524.

Porter, S., Campbell, M. A., Birt, A. R., & Woodworth, M. T. (2003). ''He said, she said'': A psychological perspective on historical memory evidence in the courtroom. *Canadian Psychology*, *44*, 190–206.

Proust, M. (1923). *Swann's Way* (Vol. 1). (C. K. Scott Moncrieff, Trans.). New York: Henry Holt & Co.

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.

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.

Ross, M., & Wilson, A. E. (2003). Autobiographical memory and conceptions of self: Getting better all the time. *Current Directions in Psychological Science*, *12*, 66–69.

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language*, *43*, 67–88.

Salaman, E. (1982). A collection of moments. In U. Neisser (Ed.), *Memory observed* (pp. 49–63). San Francisco: W. H. Freeman & Co. [Original work published 1970.]

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. (2001). Discovering memories in light of meta-awareness. *The Journal of Aggression, Maltreatment and Trauma*, *4*, 105–136.

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

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

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

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.

QUERIES MEM 03.49 ARNOLD & LINDSAY

Page numbers refer to proofs

Q1 2, 16 Dobbs & Reeves, 1996, listed as 1986 – which is correct?

Q2 14, 16 Jacoby et al., 1987, listed as 1989 – which is correct?