Memory for memory

SUSAN JOSLYN, ELIZABETH LOFTUS, AMANDA McNOUGHTON, and JAYME POWERS
University of Washington, Seattle, Washington

Participants read short passages and 1 day later they answered questions via telephone about the passages (text facts) and about the experimental session (event facts). They were telephoned again 6 weeks later and answered the same questions about text and event facts. They also answered new questions about whether they remembered the answers they had given in the initial telephone interview (recall for prior memory performance). Although participants accurately remembered the majority of past memory successes, they were poor at remembering past memory failures. After being provided with the correct answer and tested again, the participants' performance improved somewhat, especially for memory failures. This suggests that some errors in recalling past forgetting might have been due to correctly remembering the answer previously given, but failing to realize that it had been wrong. These findings have implications for a variety of situations in which people are queried about past memory performance.

How well do people remember whether they remembered something in the past? If someone claims to have forgotten something until recently, such as childhood sexual abuse (CSA), how accurate is that claim? It is important to note that this is a question not about the accuracy of the claim of abuse but rather about the accuracy of the claim of *forgetting*. In other words, how accurate is one's memory for memory itself?

Memory for memory is a form of metamemory, the ability to evaluate one's own memory capacities. Memory for memory is similar, in this respect, to the well-studied *judgments of learning* (JOL) in which people attempt to predict *future* memory performance (Dunlosky & Nelson, 1997). Curiously, however, there is little comparable literature on metamemory judgments extending backward in time—that is, retrospective metamemory judgments.

The ability to remember past *forgetting*, an occasion in the past when someone was *unable* to remember a particular event, can be critical. At the very least, chronic forgetting suggests that one's current memory strategies may be inadequate. For instance, we know a woman, Mary, who frequently forgets where she puts her keys, yet she fails to remember these occasions. Unable to remember her pattern of forgetting, Mary sees no need to implement a better retrieval strategy (e.g., depositing her keys in the same place each time).

Ironically, diagnosis of more serious memory problems also relies in large part on the patients' self-reported memory of past memory failures. People are asked such questions such as *How would you rate your ability to remember the name of someone just introduced to you?*

Self-evaluation of general memory abilities is presumably based on memory for past memory performance. Some evidence suggests that self-reported memory abilities are not related to scores on subsequent laboratory administered memory tests. This is especially true of people with memory impairments (e.g., Alzheimer's), but has also been demonstrated with nonimpaired individuals (Feher, Larrabee, Sudilovsky, & Crook, 1994; Zelinski, Gilewski, & Thompson, 1980). Self-reported memory abilities are more highly correlated with memory tests given before memory assessment than with memory tests given after memory assessment (Gervasio & Blusewicz, 1988; Herrmann, Grubs, Sigmundi, & Grueneich, 1986), suggesting that the recent remembering experience provides useful insight.

Retrospective metamemory judgments are crucial to the study of autobiographical memory, especially the study of *repression*, which involves the claim of an extended period of past forgetting. Most of the research investigating repression is based on reports by individuals of unverifiable periods of past forgetting. How do we know the "repressed" events were ever really forgotten?

This brings us back to the original question, *How well do people remember previous remembering and forgetting?* Recent work by Parks (1999) suggests that people sometimes forget recalling a childhood event that they thought about only minutes earlier. In this unique study, adult participants were asked to recall certain aspects of their childhood and adolescence. Later, when they were asked how recently they had thought about some of the same events, the majority of participants failed to report remembering at least one event that had been remembered earlier in the same experimental session.

Some evidence suggests that a recent remembering experience can bias judgments of past memory. In one study, the more childhood events that participants attempted to

Correspondence concerning this article should be addressed to S. Joslyn, Box 351525, Department of Psychology, University of Washington, Seattle, WA 98195 (e-mail: susanj@u.washington.edu).

remember, the more likely they were to report amnesia for childhood events, even when current retrieval was successful (Belli, Winkelman, Read, Schwarz, & Lynn, 1998). Perhaps the increased effort required in recalling additional events led participants to overestimate the difficulty of remembering.

In another recent study (Padilla & Poole, 1999) it was found that people were less accurate when they remembered previous recollections if they had participated in a recognition test of the same material just prior to being asked. People made the mistake of thinking they had forgotten some sentences that they had actually heard and had accurately recalled approximately half an hour earlier. This suggests that people are more likely to think that something was previously forgotten if their most recent memory experience was facilitated by closely matched cues. This finding is especially relevant in the context of clinical practice in which clients are asked directly about CSA. Conceivably, questions including specific cues (e.g., *Were you ever sexually abused by a family member?*) could increase the impression of previous forgetting.

Cues are also a problem in remembering past recall if they are different on the two remembering occasions (Arnold & Lindsay, 2000). In a laboratory list learning study, people were much less likely to remember that they had recalled homographic words (e.g., palm) in an earlier test when the cue (e.g., hand) was different from the one provided (e.g., tree) in the more recent test.

These pioneering studies imply that memory for memory can be fallible even at short intervals. Experiments such as these are important to our understanding of memory and are fundamental to the interpretation of naturalistic and autobiographical studies based on self-reported past forgetting. The work presented here was designed to contribute to that understanding.

We begin by taking a closer look at the original question: How well do people remember their own past memory performance? In other words, when people have remembered correctly in the past, do they now remember having been correct? Likewise, when memory failed them in the past, do they remember that? Take for example a student's attempt to recall whether an answer was remembered accurately on a test the previous week. The student is asked, *Did you remember the capital of Finland on the test last week?* How does the student decide whether the answer was remembered accurately? It is possible to decide without appealing to an episodic memory. The judgment can be based instead on inference. For example, if the question is very difficult or if the student neglected to study for the test, the student may infer that the question was not answered correctly on the test. Likewise, if the question was very easy or if the information was common knowledge, the student may conclude that it had been answered correctly.

We assume, however, that episodes of remembering are represented in memory like other kinds of episodes. As such, judgments of past remembering can be based

on an explicit memory, which, like other episodic memories, must first be retrieved and then evaluated (Anderson & Bower, 1972; Bahrack, 1970; Kintsch, 1970). Thus, the student may approach the problem by generating cues and searching memory for a record of reading and answering the question. One potential cue is the student's current ability to answer the question. We will focus on this approach because it is promoted by our experimental paradigm (described below). Taking this approach, the student begins by attempting to answer the test question itself (i.e., by asking, *What is the capital of Finland?*). The outcome of the retrieval attempt is then used to cue retrospective metamemory. For instance, if no answer comes to mind, the student asks whether retrieval was similarly unsuccessful during the test the previous week. This may lead to an explicit memory of forgetting on the test. Similarly, if a particular answer is now recalled, the student then asks whether it is the same answer that had been retrieved during the test. We refer to this as the *matching* judgment. The matching judgment involves comparing the present remembering experience, whether successful or not, to one's memory for remembering in the past. Although we could speculate about the cognitive steps involved in the matching judgment, such as evaluation of familiarity (Mandler, 1980) or source monitoring (Johnson, Hashtroudi, & Lindsay, 1993), they are not relevant to the present work. We assume that a matching judgment is made; however, our hypotheses do not concern the exact nature of the sub-components.

If the present answer matches the one on the test, the student must then decide whether the answer was correct. We refer to this decision as the *accuracy* judgment. The accuracy judgment is necessary because remembering is a mental state that was more or less accurate at the time, as compared with the actual event. Again, we assume that such a judgment is necessary, without hypothesizing about the underlying processes that lead to the judgment.

The bottom line is that, when retrospective metamemory is based on an explicit memory of the remembering experience, at least two judgments must be made, first a matching judgment and then an accuracy judgment. In other words, the student can only claim to have remembered the answer on the test when he/she judges both that the answer currently remembered is the same as the one on the test and that it is the correct answer. As a result, there are two kinds of errors that can be made. The rememberer can misjudge whether the current retrieval attempt duplicates the previous one (matching judgment error), or the rememberer can misjudge the accuracy of the answer (accuracy judgment error). In either case, the claim about past memory will be in error. This two-judgment model provides the theoretical framework for the study reported here.

Participants were asked to remember facts from an experimentally controlled situation similar to a test in school. We hypothesized that several variables might af-

fect retrospective metamemory of this sort. The accuracy of the target memory experience, whether or not remembering was successful on the test, could have an impact. That is, successful memory performance may be remembered differently than unsuccessful memory performance. There are several reasons why this might be the case. Successful remembering might be more memorable simply because it constitutes an additional rehearsal of the original material, long thought to strengthen memory (Atkinson & Shiffrin, 1968). Successful remembering might be more memorable because it is a positive experience. Research suggests that, in most cases, successes and positive events are recalled more readily than are failures and negative events (Bahrick, Hall, & Berger, 1996; Linton, 1986; Wagenaar, 1986). Yet, people might claim successful remembering whether or not they actually recall remembering. There is a well-documented overconfidence in prospective metamemory judgments (see Metcalfe, 1999). Perhaps people are overconfident in retrospective metamemory as well.

Another factor that may affect retrospective metamemory is present memory performance. When using present remembering as a cue for past remembering, what happens when the two retrieval attempts do *not* match? Conceivably, someone remembering correctly now may be more inclined to think that past memory was correct regardless of whether that was actually the case. Likewise, someone currently unable to remember may be more inclined to think that past memory failed. There is evidence that people are similarly biased when making judgments about other kinds of past mental states. A good example is the *knew it all along* effect or hindsight bias. This phenomenon concerns memory for another kind of past mental state, a prediction (i.e., what one thinks will happen in the future). People tend to think that current knowledge would have been more predictable in the past than it actually had been (Fischhoff, 1975, 1977). The errors people made in remembering their predictions of the outcomes of the episodes involving Clarence Thomas (Dietrich & Olson, 1993) and Rodney King (Gilbertson, Deitrich, Olson, & Guenther, 1994) were predominantly consistent with the actual outcomes. Participants were more likely to believe that they had predicted the actual outcomes—Clarence Thomas was confirmed and Rodney King's assailants were initially acquitted—even when they did not predict them. In a similar manner, retrospective metamemory judgments may be biased toward the outcome of the current retrieval attempt.

Retrospective metamemory may be affected by the nature of the material to be remembered. For instance, people may find it easier to remember past memory for autobiographical events than to remember past memory for textual information (i.e., facts learned indirectly, such as by reading). Autobiographical memories might be richer than memories for impersonal facts. Memories for actual events are often encoded from a personal perspective and in conjunction with information from several sensory modalities, making them easier to remember in

the first place (Wheeler, Stuss, & Tulving, 1997). However, in educational settings, textual information is often the focus of deliberate memory strategies that may enhance their memorability. Usually, adults are more accurate in remembering intentionally learned facts as compared with items that they do not deliberately attempt to remember (Hyde & Jenkins, 1973). In addition, memory is usually more accurate for facts that are considered central to the situation (Christianson & Loftus, 1987). Factors such as these might well affect retrospective metamemory as well.

In order to explore these issues, we brought participants into a classroom and asked them to read several passages (e.g., one on the career of Diane Arbus). The next day, we tested memory for both the facts in the passages (e.g., *What did Diane Arbus do for a living?* [text questions]) and the experimental setting in which they were learned (e.g., *Did the experimenter wear glasses?* [event questions]). Six weeks later, we asked the same questions again as well as questions about the participants' ability to recall the previous answers. We assume that this procedure (i.e., asking participants to first answer the question and then asking them to assess their previous memory for the answer) encouraged them to make matching and accuracy judgments. Then, we told the participants the correct answer and again asked them whether they had recalled it correctly 6 weeks earlier. We regard the second question as comparable to a yes/no recognition test, in that we eliminated the need for the accuracy judgment. The participants were simply asked to judge whether the correct answer was the one they had provided 6 weeks earlier. We will refer to this as *answer recognition*. In addition to allowing us to compare accuracy of the two types of tests, answer recognition allowed us to isolate the source of errors in the earlier recall test. If participants reported having remembered correctly in the recall test but reversed their opinions after the correct answer was provided (answer recognition), we inferred that the recall errors were in judging the accuracy of past performance rather than in recalling it accurately.

We explored two main issues related to the relative accuracy of retrospective metamemory judgments. We hypothesized that memory successes would be remembered more accurately than memory failures. We also hypothesized that the participants would be biased toward present memory performance. In other words, if the participants were able to remember the answer in the present, they would be more likely to say that they had remembered it in the past, and vice versa. These issues were examined for both text and event material.

METHOD

Participants

Fifty-four University of Washington students participated in return for extra credit points in psychology courses. Twelve participants failed to complete all three phases of the experiment. The 42 who completed all three phases ranged in age from 17 to 30 years with a mean age of 20. Thirty were women and 12 were men.

Procedure

In the first phase of the experiment, the participants were seated in a classroom, and they read a 10-page packet described below. In the second and third phases, an experimenter conducted telephone interviews of each participant at prearranged times.

Phase 1. In Phase 1, an experimenter brought groups of 5–17 participants into a classroom. In all, there were five groups meeting at different times of the day and on different days of the week over a period of about 1 week. Each of two experimenters conducted approximately half of the Phase 1 sessions. Two different classrooms were used.

At the outset of the session, the experimenter introduced herself, asked the participants to sign consent forms, explained how extra credit would be allotted, and arranged for subsequent telephone interviews. The participants were told that the experiment concerned memory and that they would be tested for recall and comprehension. The experimenter passed out a reading packet and asked the participants to read each passage carefully once. The participants were instructed not to reread any passage to ensure equal exposure. To encourage the participants to read for content, the experimenter asked them to write the answers to two simple comprehension questions after they finished each passage (e.g., *What time did Dennis Grossini get up on game day?*). To determine the possible role of previous exposure, the participants were also asked to indicate their prior familiarity with each topic on a 7-point rating scale, ranging from 1, *not at all familiar*, to 7, *very familiar*. The comprehension questions and the rating scale were printed on a separate sheet. Before the participants began reading the packet, the experimenter asked them whether they had any questions about the procedure, cautioning that she would not be able to answer questions after the session had begun. All participants finished within the allotted time, and most completed Phase 1 within 30 min.

Phase 2. On the evening following the Phase 1 session, a different experimenter telephoned each participant. Everyone was called between 24 and 36 h of the initial session. During the telephone interview, which lasted 12–15 min, the experimenter asked the participants 30 new factual questions (*not* the comprehension questions of Phase 1), 16 about the original passages (2 questions for each passage), and 14 about the testing session (see the Appendix). The order in which the experimenter asked the questions in the telephone interview (see the Appendix) did not follow the order in which the passages were read in Phase 1, although order was held constant across participants. The participants were instructed to make their best guess if they were unsure about the answer. A small percentage of participants refused to guess on some questions (see the Results section). The experimenter recorded the participants' responses but gave no feedback in regard to the accuracy of the responses.

The participants were not given explicit instructions about note taking. However, no participant was observed taking notes during Phase 1 of the experiment, nor did interviewers have the impression that the participants were referring to notes during either this or the subsequent phone interview. Phase 2 accuracy (see below) supports this impression.

Phase 3. Six weeks after the Phase 2 interview, the experimenter telephoned the participants a second time. During the Phase 3 interview, which lasted 25–35 min, the experimenter (the same one who had conducted the previous telephone interview) asked the participants the same factual questions in the same order as were asked in Phase 2.¹ The participants were again encouraged to make their best guess if they were unsure of the answer. In Phase 3, however, three more parts followed each question. (1) The experimenter asked, "Did you remember the answer to this question during your first telephone interview?", and the participant answered. Then, (2) the experimenter told the participant the correct answer, and (3) again asked, "Did you remember the answer to this question during your first telephone interview?" (answer recognition). The participant answered.

Materials

The 10-page reading packet presented in Phase 1 comprised eight short passages on diverse topics. Each passage was fairly simple and accessible to nonexperts. However, each was chosen because the information was judged as not being common knowledge. Passages were taken from *National Geographic* and similar sources. The passages had an average Flesch-Kincaid reading level of grade 9. The topics were Diane Arbus, superstitions of baseball players, flavors of beer, the history of coffee, red crabs, dinosaurs, kohlrabi, and horseback riding. We selected relatively obscure topics in an attempt to control the amount of exposure the participants had to the material and to minimize the possible effects of prior knowledge. Each passage ranged in length from one half to one page and was followed by two comprehension questions. All participants read the passages in the same order.

In Phases 2 and 3, the experimenters asked the participants 30 questions about the passages and the testing session. These questions are shown in the Appendix in the order in which they were asked. Six questions (4, 11, 12, 24, 29, and 30) were not included in the analyses reported below because of potential ambiguities in the answers would make it difficult to interpret the participants' memory for accuracy. These 6 questions had several (unanticipated) correct answers. For instance, 1 question asked the name of the experimenter and some participants gave the principal investigator's name, presumably because it was also printed on the consent form. In fact, the question referred to the name of the experimenter who had administered the original session (not to the principal investigator). Numbers 11 and 30 were the only questions referring to a single story (i.e., kohlrabi), leaving no remaining questions on that particular reading.

RESULTS

Phase 2: Event and Text Question Accuracy

First, we examined the relative accuracy of the target memory experience (i.e., the Phase 2 answers to the 13 unambiguous *event* and 11 unambiguous *text* questions). To evaluate whether the answers that participants gave were correct, three judges reviewed all of the answers. Questions with more than one possible answer were omitted. In a few cases in which the wording was only slightly different, the judgment about accuracy was arrived at by consensus. Accuracy ranged from 93% to 10%. Only 14% of the errors occurred because participants said they did not know the answer. The majority of these responses were made to text questions (86%). Both kinds of errors, wrong answers and *don't knows*, were regarded as forgetting. As Table 1 indicates, in general, the participants were more accurate on *event* (9.1/13, 70%) than on *text* questions (4.7/11, 43%). Statistical support for this claim was obtained by calculating average percent accuracy for each participant and conducting a paired *t* test [$t(41) = 13.93, p < .0001$].²

Preliminary results suggested a situation, in which making assumptions about past memory performance without accessing explicit memories was discouraged. The average Phase 1 familiarity rating for reading packet topics was low, 1.5 on a scale ranging from 1 (not at all familiar) to 7 (very familiar). On average, the participants were 90% correct on the Phase 1 comprehension questions. The absence of ceiling effects in Phase 2 accuracy and the low familiarity ratings suggested that the ques-

Table 1
Mean Number of Correct, Wrong, and Don't Know Answers to Event and Text Questions in Phase 2 and Phase 3

	Event (<i>n</i> = 13)		Text (<i>n</i> = 11)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phase 2				
Correct	9.1	1.7	4.7	2.1
Wrong	3.3	1.5	3.5	1.8
Don't know	0.6	0.8	2.8	2.4
Phase 3				
Correct	8.1	1.7	3.7	2.5
Wrong	4.1	1.5	4.2	2.2
Don't know	0.8	1.2	3.1	2.9

tions were not extremely easy. Nor was the material well known. Likewise, neither the absence of floor effects in Phase 2 accuracy nor the high Phase 1 comprehension accuracy suggested that questions were extremely difficult or that the passages were not carefully read.

Phase 3 Interview: Memory for Memory

Our main interest was in how well participants remembered (during Phase 3) their previous Phase 2 performance (i.e., retrospective metamemory). Recall that 6 weeks after the first telephone interview, each participant was asked the same 30 questions again (e.g., *What time was your original testing session?*). Immediately after giving the answer (e.g., 10 A.M.), each participant was asked *Did you remember the answer to this question during your first phone interview?* and he or she responded *yes* or *no*. We examined whether memory for memory was influenced by content (*text* and *event*) and by Phase 2 accuracy (success and failure).

We calculated average percent memory for memory accuracy for each participant and found that the participants were often accurate when recalling successful Phase 2 remembering, for both *event* ($M = 87\%$, $SD = 12\%$) and *text* answers ($M = 79\%$, $SD = 26\%$). The participants were less accurate in their assessment of past memory failure than that of past memory success. They responded correctly that they had *not* remembered on average only 49% ($SD = 28\%$) of the event answers and 55% ($SD = 28\%$) of the text answers. Almost half of the time, they claimed to have remembered when they had not. This is what one would expect, given no prior information and no biases, from chance (50%) performance.

In order to explore the impact of content and Phase 2 accuracy on memory for memory, a repeated measures analysis of variance (ANOVA) was conducted. Although there was no reliable main effect for content [$F(1,41) = .37$, $MS_e = 0.02$, $p > .05$], there was a reliable main effect for Phase 2 accuracy [$F(1,41) = 37.58$, $MS_e = 3.47$, $p < .0001$]. Memory for successful remembering was reliably better than memory for memory failure. It is important to note, however, that even memory for success was far from perfect. The participants misremembered 13% of the correct *event* responses and 21% of the cor-

rect *text* responses, claiming that they had not remembered the answer when they actually had.

In addition, there was a reliable content $\times$ Phase 2 accuracy interaction [$F(1,41) = 4.56$, $MS_e = 0.30$, $p < .05$]. There was a much greater difference between successful ($M = 87\%$, $SD = 12\%$) and failed *event* memory ($M = 50\%$, $SD = 28\%$) than there was between successful ($M = 77\%$, $SD = 29\%$) and failed *text* memory ($M = 57\%$, $SD = 29\%$). Although memory for successful remembering was better overall for text answers, the large discrepancy in metamemory accuracy between memory for remembering and for forgetting was attenuated.

Bias

We hypothesized two sources of bias in retrospective metamemory judgments. First, people may be overconfident about past remembering (i.e., they may be more inclined to think that they remembered regardless of whether or not they did). Another source of potential bias is current memory performance. When people currently remember the answer, they may be more likely to think that they had remembered it in the past. Likewise, when people currently fail to remember the answer, they may be more likely to think that they had forgotten it in the past. To explore these issues, we examined error responses to the Phase 3 memory question (*Did you remember the answer in Phase 2?*). Error responses are those in which the participant answered *no* to this question but had remembered accurately in Phase 2, and vice versa, those who answered *yes* but had not remembered the answer in Phase 2. *Yes* and *no* error responses were categorized for congruence with Phase 3 accuracy. Accuracy, here, refers to whether participants answered the factual questions (*What time was your original testing session?*) correctly in Phase 3. In other words, participants who gave the correct answer in Phase 3 (10 a.m.) and assessed their past performance as being correct (*yes*) were classified as congruent. Similarly, participants who gave the incorrect answer in Phase 3 (12 a.m.) and assessed their past performance as being incorrect (*no*) were also classified as congruent. Mismatches between Phase 3 accuracy and assessment were categorized as incongruent. The proportion of congruent and incongruent, yes and no responses was calculated for each participant. The average percentages of congruent (shaded) and incongruent (unshaded), yes and no responses across participants are shown in Figure 1.

In order to explore the impact of congruence and response type (*yes* or *no*) on retrospective metamemory, a repeated measures ANOVA was conducted. Ten participants were omitted from this analysis because there was at least one category in which they made no response. There was a reliable tendency for congruence [$F(1,31) = 22.11$, $MS_e = 1.45$, $p < .0001$], as well as a reliable tendency to say *yes* indicating a participant's having remembered in the past [$F(1,31) = 11.23$, $MS_e = 2.05$, $p < .01$]. There was also a reliable response type $\times$ congruence interaction [$F(1,31) = 4.39$, $MS_e = 0.38$, $p < .05$].

Figure 1. Mean percent yes and no error responses to the Phase 3 recall question *Did you remember the answer to this question during your first phone interview?* Shaded bars represent responses congruent with current memory performance (i.e., assessing past memory as successful when remembering, and assessing past memory as failed when forgetting).

The tendency toward congruence was much stronger when responding *no*, indicating forgetting in the past. People were inclined to say *yes*, indicating that they had remembered, whether the response was congruent with present performance or not.

Answer Recognition

Next we analyzed the answer recognition judgments. After hearing the correct answer in Phase 3, participants indicated whether they had provided the answer in Phase 2. Comparing answer recognition to recall allowed us to estimate the proportion of recall errors that resulted from matching judgments as opposed to accuracy judgments.

The mean percent accuracy for answer recognition was 83% ($SD = 11\%$) when remembering a successful memory, and 76% ($SD = 18\%$) when remembering memory failures. For recall, the mean percent accuracy had been 85% ($SD = 11\%$) when remembering memory success, and 54% ($SD = 22\%$) when remembering memory failure. Because we were also curious about the impact of Phase 2 accuracy (success and failure) on the relationship between the two types of test (answer recognition and recall), a repeated measures ANOVA was conducted. There was a reliable main effect for test type [$F(1,41) = 58.02, MS_e = 0.45, p < .0001$] and a reliable test type $\times$ Phase 2 accuracy interaction [$F(1,41) = 73.63, MS_e = 0.74, p < .0001$]. Tests of simple effects, using Bonferroni adjusted critical values, revealed that the improvement of answer recognition over recall occurred mainly when participants were attempting to remember Phase 2 memory failures [$t(41) = 9.57, p < .0001$]. Answer recognition accuracy was not reliably different from recall accuracy when participants remembered successful Phase 2 remembering [$t(41) = -1.38, p > .025$]. Again, there was also a main effect for accu-

racy [$F(1,41) = 24.91, MS_e = 1.48, p < .0001$]; remembering past memory success was more accurate than remembering past memory failure.

The predicted improvement in answer recognition over recall appeared to be exclusively in memory for past errors. Altogether, about half of the recall errors were corrected in the answer-recognition test. This suggests that approximately half of the errors in recalling past memory failures arose because of invalid accuracy judgments. People remembered the answer that they had given before, but erroneously thought it had been correct.

Curiously, hearing the correct answer did not improve participants' memory for successful remembering. When we looked at *text* and *event* answers separately, we discovered that most of the errors were made on *text* questions. Participants were, on average, 6% less accurate in recognizing *text* answers than in recalling them. This difference was reliable [$t(40) = 1.8, p < .05$]. Some participants made an accurate assessment of past memory in the recall test and then changed their minds after hearing the correct answer. Because we thought this was an odd sequence of responses, we examined the answers of these 24 participants individually. In approximately a third of the cases, there were slight differences in the wording. For example, 1 participant's answer to question 8, *When horseback riding, from what is the proper position built?* in Phase 2 was "from the stirrups up" which was considered correct although the actual answer provided by the experimenter in Phase 3 was "from the feet up." In the rest of the cases, however, the answer given by the participant in Phase 2 and the correct answer provided by the experimenter in Phase 3 were identical. It is interesting to note, that, for over half of these questions (11/18, 61%), although the Phase 2 answer was correct, the Phase 3 answer provided by participants was wrong. This suggests that some participants who had provided the correct answer in Phase 2 and had said that they recalled having answered correctly, were committing both matching and accuracy judgment errors. They erroneously thought that both the Phase 3 answer had been correct and that it had been given in Phase 2. After the Phase 3 answer was revealed to be wrong, they altered their accuracy judgments by saying that they had been wrong but had retained their matching judgment, assuming that they had given the same answer in Phase 2. In fact they had given the correct, albeit different, answer 6 weeks earlier.

To summarize, the predicted improvement in answer recognition over recall resides predominantly in memory for past errors, suggesting accuracy judgment errors in recall. There was no answer-recognition advantage for correct responses, suggesting that forgotten memory success is not retrievable, even with very specific cues. In some cases, answer recognition was actually worse than recall. This is more difficult to explain, although some of these responses may be due to a combination of accuracy and matching judgment errors.

DISCUSSION

How well do people remember past memory performance? Our results suggest that people remember their memory successes fairly well. For example, when participants had been accurate in Phase 2, they recalled those successes about 80% of the time. Memory failures were much less well remembered. Participants who had been wrong in Phase 2 remembered poorly: Approximately half the time they said that they remembered items that they had previously forgotten.

One possible explanation for this pattern of results is that memory successes are more readily accessible because they are positive experiences. Another possible explanation is that memory successes are especially coherent experiences because the correct answer constitutes a better match to both the original experience and the question. Moreover, remembering the correct answer provides an additional rehearsal of identical material.

It is important to note, however, that despite all of our precautions, high accuracy for correct responses is not necessarily evidence for explicit memory. People consistently judged their memory to be more accurate than it actually was. Indeed, approximately half the memory failures were judged to be successes. This suggests that, in many cases, people do not access explicit memories when they claim to have remembered, but rather they may make inferences from an inflated analysis of their own abilities. They merely assume that they must have remembered. Such a strategy works in their favor when they have actually provided the correct answer on the previous occasion. However, the same bias works against them when they attempt to recall past errors.

Recalling past memory errors (i.e., recalling forgetting) was difficult. Some of these mistakes might have been accuracy judgment errors. When claiming, in Phase 3, to have remembered the answer in Phase 2, some participants might have failed to realize that the answer they had given was wrong. In other cases, participants may have been unable to remember the answer they had given. When participants responded that they did not know the answer in Phase 2, in Phase 3 they were searching memory for what was essentially a nonevent—that is, nothing was retrieved. However, because we strongly encouraged participants to provide an answer to the initial questions, this accounts for only a small proportion of the misremembered errors. Retrieving the wrong answer in Phase 2 might also be less memorable. A wrong answer fits less well with the question and lacks the salient details accompanying the original learning experience. Thus, recalling the wrong answer may constitute a less detailed, less coherent, and hence less memorable experience. In cases like these, when memory for the original remembering event was weak, participants might have been swayed by their bias to claim successful memory.

It is important to note that, although memory for memory success was generally more accurate than mem-

ory for memory failure, 13% of the correct event answers and 21% of the correct text answers were remembered as failures. When people made errors about memory success, they were biased toward congruence with present performance. When they failed to remember the correct answer in Phase 3, participants were more likely to answer *no*, indicating that they had forgotten in Phase 2 (although they had not), than when they remembered the answer in Phase 3. This too may have been due to an over-reliance on inference. Participants may have assumed that, because the question was difficult to answer in Phase 3, they must have erred 6 weeks earlier.

A question-difficulty heuristic is also suggested by the fact that participants were less accurate when they remembered successfully remembered text answers, which were less accurate to begin with. Regarding the text questions as difficult, some participants may have had a tendency to assume that they were answered incorrectly. Many of us would assume, for instance, that we could not provide a thorough explanation for string theory. If asked about providing the explanation on a past occasion, we would not bother to consult memory and simply answer “no.” The question-difficulty heuristic resembles the cognitive-effort heuristic discussed earlier (Belli et al., 1998). When increased effort is required, past memory is assumed to be less complete. Heuristics such as these may well bias retrospective metamemory judgments toward a claim of forgetting in some cases. Surprisingly, memory for memory successes did not improve in the answer recognition test, suggesting that forgotten memory success is especially difficult to access. People may actually be more likely to infer forgetting when their most recent remembering effort was facilitated by specific cues (Padilla & Poole, 1999).

What about the notion of repression? Obviously, laboratory-based work must be cautiously applied to real-world autobiographical memory and CSA. However, in light of the fact that much of the research investigating repression is based on unverifiable periods of self-reported, prolonged forgetting, our results may have relevance. They suggest that people are not always accurate when remembering past memory performance. Thus, it is possible, under some circumstances, for people to remember an event in the present and believe that it was forgotten in the past when in fact it was not. Schooler, Bendixen, and Ambadar (1997) describe just such a situation, calling it the *forgot it all along effect*, in which an individual erroneously claims to be remembering an abusive event for the first time in a therapeutic context. However, other people close to her report that she had, in fact, discussed the event several times in the past. Schooler et al. suggest that realizing that one has been the victim of CSA can generate a rush of emotion which may, in turn, be used as a heuristic indicating that the memory was previously forgotten. This sequence of events could give the appearance of lifting “total repression,” when actually the specific incident was consistently available to memory.

The work presented here contributes to our understanding of retrospective metamemory and highlights issues related to CSA and repression that demand further investigation. Clearly more work is required to explore the factors that impact retrospective metamemory accuracy, especially as it concerns erroneously reported forgetting. Although there is widespread acceptance of the notion of repression in some circles, much of the evidence for it is based on self-report, in which, often, neither the event nor the forgetting is verifiable. There has been a healthy debate of issues relating to verification of the event. Until recently, however, the issue of memory for forgetting has been little discussed. We believe that this work sheds light on that important issue. We have shown that, like human memory in general, memory for memory can be quite fallible even after a short duration. Thus, evidence based on self-reported periods of past forgetting must be scrutinized, for it is entirely possible that some reported forgetting never took place.

REFERENCES

- ANDERSON, J. R., & BOWER, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*, 97-123.
- ARNOLD, M. M., & LINDSAY, D. S. (May, 2000). *Remembering remembering*. Paper presented at meeting of NOWCAM, Victoria, BC.
- ATKINSON, R. C., & SHIFFRIN, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.
- BAHRICK, H. P. (1970). Two phase model for prompted recall. *Psychological Review*, *77*, 215-222.
- BAHRICK, H. P., HALL, L. K., & BERGER, S. A. (1996). Accuracy and distortion in memory for high school grades. *Psychological Science*, *7*, 265-271.
- 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.
- CHRISTIANSON, S. A., & LOFTUS, E. F. (1987). Memory for traumatic events. *Applied Cognitive Psychology*, *1*, 225-239.
- DIETRICH, D., & OLSON, M. (1993). A demonstration of hindsight bias using the Thomas confirmation vote. *Psychological Reports*, *72*, 377-378.
- DUNLOSKY, J., & NELSON, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL. *Journal of Memory & Language*, *36*, 34-49.
- FEHER, E. P., LARRABEE, G. J., SUDILOVSKY, A., & CROOK, T. H. (1994). Memory self-report in Alzheimer's Disease and in age-associated memory impairment. *Journal of Geriatric Psychiatry & Neurology*, *7*, 58-65.
- FISCHHOFF, B. (1975). Hindsight does not equal foresight. *Journal of Experimental Psychology: Human Perception & Performance*, *1*, 288-299.
- FISCHHOFF, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception & Performance*, *3*, 349-358.
- FOOS, P. W., & FISHER, R. P. (1988). Using tests as learning opportunities. *Journal of Educational Psychology*, *80*, 179-183.
- GERVASIO, A. H., & BLUSEWICZ, M. J. (1988). Prediction and evaluation of everyday memory in neurological patients. *Bulletin of the Psychonomic Society*, *26*, 339-342.
- GILBERTSON, L. J., DIETRICH, D., OLSON, M., & GUENTHER, R. K. (1994). A study of hindsight bias: The Rodney King case in retrospect. *Psychological Reports*, *74*, 383-386.
- HERRMANN, D. J., GRUBS, L., SIGMUNDI, R., & GRUENEICH, R. (1986). Awareness of memory ability as a function of memory experience. *Human Learning*, *5*, 91-107.
- HYDE, T. S., & JENKINS, J. J. (1973). Recall for words as a function of semantic, graphic and syntactic orienting tasks. *Journal of Verbal Learning & Verbal Behavior*, *17*, 649-667.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, *113*, 1-28.
- KINTSCH, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 331-373). New York: Academic Press.
- LINTON, M. (1986). Ways of searching the contents of memory. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 50-69). Cambridge: Cambridge University Press.
- MADIGAN, S. (1976). Reminiscence and item recovery in free recall. *Memory & Cognition*, *4*, 233-236.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271.
- METCALFE, J. (1999). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality & Social Psychological Review*, *2*, 100-110.
- PADILLA, L., & POOLE, D. (1999, August). *Memory for previous recall: A comparison of free recall and recognition*. Poster session presented at the American Psychological Society meeting, Denver.
- PARKS, T. E. (1999). On one aspect of the evidence for recovered memories. *American Journal of Psychology*, *112*, 365-370.
- 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. Conway (Ed.), *False and recovered memories* (pp. 251-292). Oxford: Oxford University Press.
- WAGENAAR, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, *18*, 225-252.
- WHEELER, M. A., & ROEDIGER, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*, 240-245.
- WHEELER, M. A., STUSS, D. T., & TULVING, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin*, *121*, 331-354.
- ZELINSKI, E. M., GILEWISKI, M. J., & THOMPSON, L. W. (1980). Do laboratory tests relate to self-assessment of memory ability in the young and old? In L. W. Poon & J. L. Fozard (Eds.), *New directions in memory and aging: Proceedings of the George A. Talland Memorial Conference* (pp. 519-544). Hillsdale, NJ: Erlbaum.

NOTES

1. We acknowledge that our asking the questions a second time might affect accuracy. Some evidence suggests that repeated questioning, especially at shorter intervals, can lead to hypermnnesia (i.e., more information available with each subsequent test, Madigan, 1976; Wheeler & Roediger, 1992). Other evidence suggests less forgetting in repeated testing conditions, as compared with single test conditions (Foos & Fisher, 1988). Alternatively, repeated questions might consolidate memory errors. Although it is difficult to predict the exact effect of having asked the same question 6 weeks previously, memory accuracy per se was not our focus. Because we were interested in how well participants remembered previous performance, rather than the performance itself, and because repetition was held constant, the impact of repeated questioning is not a crucial issue.
2. Several event questions had few answer possibilities (2, 6, 10, 13, 14, and 18). For these questions, guessing alone might have led to higher accuracy. However, the superiority of event over text accuracy remains when these questions are omitted from the analysis [$t(41) = 4.87$, $p < .0001$].

APPENDIX

Questions Asked in Phase 2 and 3

When the experimenter telephoned the participant, she explained who she was and why she was calling. Then she explained that the questions she was about to ask referred to the original, in person, experimental session.

1. What time was your original testing session? (event)
2. What day of the week was your testing session? (event)
3. Why do red crabs march from the forest to the sea? (text)
4. What was the experimenter's name? (event)
5. What was the first topic you read about? (event)
6. Was the experimenter willing to answer questions during the experiment? (event)
7. How many other subjects were in the testing session? (event)
8. When horseback riding, from what is the proper position built? (text)
9. What was the most important part of Dennis Grossini's ritual? (text)
10. Was the experimenter wearing glasses? (event)
11. What vegetable is kohlrabi most similar to? (event)
12. When Germans speak of a *beautiful flower* to what are they referring? (event)
13. Was there a clock on the wall of the room during your testing session? (event)
14. Was there a window in the room? (event)
15. What religion considered drinking coffee sinful at first? (text)
16. What did Diane Argus do for a living? (text)
17. In Glen Rose, in the 1980's, paleontologists tried to find evidence to support what? (text)
18. Were there more male subjects or more female subjects, or an equal number of male and female subjects? (event)
19. What was the name of the experiment? (event)
20. In what ocean is Christmas Island located? (text)
21. Did you see something in the classroom that was unusual for a classroom? What? (event)
22. How many cups of coffee did Voltaire 'reportedly' drink a day? (text)
23. In which building was the experiment held? (event)
24. Was the day of the experiment rainy, sunny, or overcast? (event)
25. Bouquet is most important to what style of beer? (text)
26. Several residents of Glen Rose supplemented their incomes by selling what to tourists? (text)
27. How did Diane Arbus die? (text)
28. What color hair did the experimenter have? (event)
29. Why do baseball players adopt rituals? (text)
30. When was kohlrabi first noted in the US? (text)

In Phase 3, after each of the questions above, the experimenter asked these questions:

Did you remember the answer to this question during your first telephone interview?

The correct answer is (the correct answer was provided).

Did you remember the answer to this question during your first telephone interview?

(Manuscript received June 30, 1998;
revision accepted for publication April 11, 2001.)