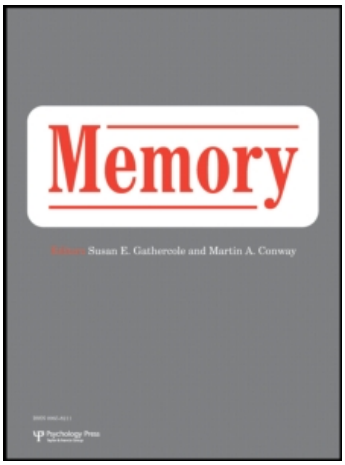

Memory
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683358

Fact learning: How information accuracy, delay, and repeated testing change retention and retrieval experience

Sarah J. Barber [a]; Suparna Rajaram [a]; Elizabeth J. Marsh [b]
[a] Stony Brook University, NY, USA [b] Duke University, Durham, NC, USA

Online Publication Date: 01 November 2008

To cite this Article Barber, Sarah J., Rajaram, Suparna and Marsh, Elizabeth J.(2008)'Fact learning: How information accuracy, delay, and repeated testing change retention and retrieval experience',Memory,16:8,934 — 946
To link to this Article: DOI: 10.1080/09658210802360603
URL: http://dx.doi.org/10.1080/09658210802360603

MEMORY, 2008, 16 (8), 934–946

Psychology Press
Taylor & Francis Group

Fact learning: How information accuracy, delay, and repeated testing change retention and retrieval experience

Sarah J. Barber and Suparna Rajaram
Stony Brook University, NY, USA

Elizabeth J. Marsh
Duke University, Durham, NC, USA

Previous classroom studies have shown that the phenomenology of studied facts changes over time. However, pedagogical needs preclude both the study of errors and the separation of the effects that delay and repeated testing have on retention and retrieval experience. We addressed these issues together in an experiment where participants read stories containing correct and misleading information and provided Remember, Just Know, and Familiar judgements on immediate and delayed general knowledge tests. After 2 days, information learned from the stories shifted from Remembered to Just Known, but repeated testing attenuated this shift. Interestingly, similar patterns of retrieval and phenomenology were observed for correct and misleading information with one important difference—the shift over time to Just Knowing was significantly greater for correct than for misleading information. Together, these findings show the roles of information accuracy, delay, and testing in determining both retention and the subjective experience of retrieval.

Keywords: Episodic memory; Familiarity; Remember–Know paradigm; Semantic memory; Source memory; False memory.

Much knowledge is gained through personal experience, such as when a student listens to a lecture on DNA technology, a reader reads a novel about the civil war, and a traveller learns Spanish vocabulary as she navigates Barcelona. Such knowledge appears to fall into the domain of semantic memory, which is a repository of knowledge about concepts, words, people, and their interrelationships in the world. Semantic memory is often contrasted with episodic memory, which refers to conscious recollection of specific episodes that are accessible through mental travel back in time and space to the original event (e.g., Tulving, 1972, 1995). One of the key distinctions between episodic and semantic memory is that they differ in their phenomenology, with episodic but not semantic memories associated with recollective detail. From this perspective, for example, the traveller will later use the Spanish words without mentally re-experiencing her trip, but will mentally re-experience walking through Gaudi's apartment building when remembering her trip to Barcelona.

Given the importance of phenomenology in theoretical distinctions between episodic and semantic memories, much research effort has been directed at studying the phenomenology of retrieval. But most of this work has used words as stimuli and there are surprisingly few findings

Address correspondence to: Suparna Rajaram, Department of Psychology, Stony Brook University, NY 11794-2500, USA. E-mail: suparna.rajaram@sunysb.edu

http://www.psypress.com/memory DOI:10.1080/09658210802360603

about the phenomenology of retrieved facts. The most relevant studies have been conducted in actual classrooms, with the first done in undergraduate psychology classes by Conway and colleagues (Conway, Gardiner, Perfect, Anderson, & Cohen, 1997). Students took a series of lecture and methods courses, each of which lasted 6 weeks. At the end of each 6-week block students took a multiple-choice test on target facts, and some of these facts were retested later in the year. Critically, after selecting an answer from the three alternatives the students judged how they knew each answer. The memory awareness instructions were based on Tulving's (1985) Remember–Know instructions, with one important variation. ''Remember'' instructions paralleled those typically given in episodic memory studies, with an answer labelled as ''Remembered'' if details associated with learning the fact were recollected, such as remembering reading the fact in the textbook or recalling how the lecturer said a fact. In contrast to prototypical ''Know'' instructions (whereby something is ''known'' if it is recognised as having occurred in a prior study list even though nothing is consciously recollected about its actual occurrence), Conway et al. separated two different senses of ''knowing''. Specifically, participants labelled a fact as ''just known'' if they simply knew the answer and did not recall any particular episode associated with it. In contrast, a fact was called ''familiar'' if the answer was neither recollected nor known, but was chosen because it felt more familiar than the other options. Consider how these ''just know'' and ''familiar'' instructions compare to the phenomenology typically captured in episodic memory studies, in which ''knowing'' means attributing a memory to a particular source (e.g., the study list) even if the details of that source are not recollected (see reviews by Gardiner, 2008; Rajaram, 1999; Roediger, Rajaram, & Geraci, 2007). In contrast to Tulving's original conceptualisation of ''know'' as an index of semantic memory, Know judgements are usually treated as a measure of familiarity or fluency in episodic memory experiments (e.g., Jacoby, Yonelinas, & Jennings, 1997; Yonelinas, 2002). In the Conway et al. study, ''familiar'' was similar to how ''know'' is typically instantiated in episodic memory studies, and ''just know'' captured Tulving's original sense of the concept. Although the concept of semantic memory is not without controversy, ''just knowing'' is useful as a judgement that indexes knowledge about the world, as opposed to a general sense of familiarity associated with past events.

Intriguingly, Conway et al. found different phenomenologies associated with retrieval of facts from lectures versus research methods classes. The methodology courses emphasised the integration of information, and yielded more Just Know than Remember responses. In contrast, the lecture classes emphasised the retention of many specific details, with the result that all but the students with the lowest grades gave more Remember than Just Know responses. However, when retested at the end of the Introductory Psychology course, students gave more Just Know than Remember responses. Thus, there was a shift from Remembering to Just Knowing over time, possibly due to repeated testing or the passage of time, or both. These findings have been replicated in other classroom settings (e.g., Herbert & Burt, 2001, 2003). However, it is critical to understand the phenomenology of retrieved facts under experimentally controlled conditions because this approach can help specify when Remembering, Knowing (Familiarity), or Just Knowing would characterise retrieval. Classroom studies preclude systematic manipulations of multiple variables but their findings motivate the selection and testing of such variables.

We could not find a laboratory experiment that paralleled these intriguing classroom experiments. The closest study was one in which participants read short stories that contained correct and misleading information, some of which could be used to answer questions on a later general knowledge test (Marsh, Meade, & Roediger, 2003). For example, one student might have read a story stating that ''Moscow is the capital of Russia'' (a correct fact) whereas another read ''St. Petersburg is the capital of Russia'' (a misleading statement), and of interest was students' later answers to the question ''What is the capital of Russia?'' Critically, participants made two source attributions about each answer on the final general knowledge test: whether or not the answer had been read in one of the preceding stories, and whether or not the answer was part of their general world knowledge (meaning that they could have answered the question before participating in the experiment). Participants were explicitly instructed that they could attribute an answer to both sources; an answer could have been read in one of the stories and still have been known prior to the experiment. Reading both correct and misleading facts increased

their production on the later general knowledge test; of critical interest for present purposes is the phenomenology accompanying the retrieval of the answers. On immediate tests, participants were quite good at remembering that they had read both correct and misleading information in the stories, and memory for the story source was better when the stories had been read twice rather than once. This is consistent with Conway's findings that participants sometimes ''remember'' having learned specific facts in lecture courses. However, reading the stories also led participants to believe that they knew many of the answers prior to the experiment, and this was not limited to correct information. Rather, it also occurred for misinformation that was unlikely to have been believed prior to the experiment. Thus although participants showed good memory for the stories, they also showed an illusion of prior knowledge, which is similar to claiming to ''just know'' something. However, this study did not allow conclusions about whether knowledge shifted from being ''remembered'' to ''just known'' over time, as source judgements were not recorded on both immediate and delayed tests in this study.

In short, there is no laboratory study that examines the phenomenology of retrieved facts and how that may shift over time for correct as well as for misleading information. To this end, we conducted a large, laboratory study using the paradigm of Marsh et al. (2003). Participants read stories containing references to real people, places, and objects, and we systematically manipulated three variables, including one encoding variable (information accuracy: correct versus misleading), one retention variable (delay: immediate versus 48 hours), and one retrieval variable (repeated testing: one or no test before the delayed test), and asked participants to make Remember, Familiar, and Just Know judgements for answers produced in response to general knowledge questions.

If retrieved facts are initially associated with recollective detail, they should be judged as ''Remembered'' on the initial test. The key question is whether the phenomenology of retrieved facts changes over a delay of 2 days. A shift from more Remember to more Just Know responses would be in concordance with, and serve as a laboratory demonstration of, the findings of Conway and colleagues (1997). In addition, our design allowed us to simultaneously assess the roles of several variables in shifting the phenomenology of retrieved facts. For example, in the classroom studies all items on the final test were also tested earlier in the term, obscuring whether delay alone or repeated testing alone is sufficient, or whether both are needed to increase Just Know responses. A laboratory design allowed us to systematically test this issue. Repeated testing may enhance the shift from more Remember to more Just Know responses (as suggested by classroom findings) or it may attenuate the shift because repeated testing over a delay of 2 days (as compared to 25 weeks in Conway et al.'s study) may increase memory for episodic detail through rehearsal.

In addition to separating the effects of testing and delay, the laboratory paradigm allowed us to examine the phenomenology of errors, which cannot be examined in classroom studies (as teachers do not deliberately introduce errors into their teaching). People's knowledge about the world is not perfect, and a complete understanding of semantic memory must encompass errors as well as correct beliefs. The phenomenology of these errors may also have implications for correcting errors in knowledge bases; if errors are ''remembered'' as having occurred in a particular source then source-monitoring instructions may help participants avoid the influence of unreliable sources, but if errors are ''just known'' then other strategies may be recommended.

It is not clear what to predict about the phenomenology of errors in the knowledge base. In general, less is known about when false memories will be ''remembered'' than is true of veridical memories. Depending on the particular false memory paradigm being used, sometimes episodic memory errors are ''remembered'' at rates similar to veridical memories (e.g., Chan & McDermott, 2006; Roediger & McDermott, 1995) and other times they are less likely to be ''remembered'' and instead are familiar or known (e.g., Lane & Zaragoza, 1995). Even less can be said about illusions of knowledge than about episodic false memories; in general, anything that increases the availability of false facts increases people's confidence that they are true (e.g., Hasher, Goldstein, & Toppino, 1977; Kelley & Lindsay, 1993).

For our purpose, of particular interest are studies in which testing and delay were manipulated and remember–know judgements were made. Consider the Deese-Roediger-McDermott paradigm for creating false memories of non-presented words: participants are likely to falsely recall and recognise a critical non-presented word

such as *sleep* after studying a list of related words such as *bed*, *rest*, *wake*, *tired* (Deese, 1959; Roediger & McDermott, 1995). An initial attempt to recall the list (a test) boosts later ''R'' judgements for both studied items and critical lures (Roediger & McDermott, 1995, exp. 2). Similar effects of testing have been observed in the eyewitness studies in which participants witness an event and are later misled on its details. Producing the misinformation on an initial test increases the likelihood it will be ''remembered'' on a later test (Roediger, Jacoby, & McDermott, 1996). Thus the false memory literature suggests that testing has similar effects on the phenomenology of true and false facts, although it is not a given that episodic false memories and illusions of knowledge will behave similarly (e.g., Fazio & Marsh, 2008). As already described, Marsh et al. (2003) did collect source attributions for misleading information learned while reading passages: repeated testing increased misinformation production and story attributions. However, no conclusions about delay were possible as source judgements were only collected on a single test, which is unfortunate since the picture becomes more complicated when thinking about the effects of delay on phenomenology. In the DRM paradigm, ''remember'' responses for false memories do not always drop over a delay (e.g., Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996). And at least sometimes, ''remember'' responses *increase* over a delay in the eyewitness misinformation paradigm (Frost, 2000).

In short, our participants read several stories that contained correct and misleading information that could be used to answer questions on later general knowledge tests. The design allowed us to separate the effects of repeated testing and delay on the phenomenology of answers produced on the final test, for both correct and misleading information.

METHOD

Participants

A total of 288 Stony Brook undergraduates volunteered for partial fulfilment of a course requirement. Of these, 192 completed a group norming study and 96 participated individually in the experiment.

Design

Information accuracy (correct, misleading), and story status (read, not read) were manipulated within participants. Delay (immediate, 48 hours) and repeated testing (delayed test: single, repeated) were manipulated between participants. Half of the participants took the immediate test, and all returned for the second session. To pinpoint the role of delay, performance on the immediate test was compared to the delayed test performance of participants who had not previously been tested. To pinpoint the role of repeated testing, performance on the single delayed test was again examined, but instead compared to performance observed on the repeated delayed test. The dependent measures were the proportion of general knowledge questions answered correctly versus with misinformation, and the classification of answers as *Remember*, *Familiar*, or *Just Know*.

Materials

In the norming study participants answered 36 general knowledge questions related to facts from two stories from Marsh (2004; ''Summer Star Search'' and ''The Inventor''). There was no penalty for guessing and participants were encouraged to answer even if unsure. In all, 32 questions were selected to span a range of difficulty (the hardest question was correctly answered by 0.5% of students while the easiest was correctly answered by 81.2% of students).

We matched the difficulty of the questions across the two stories. Each story contained two pairs of facts also equated for difficulty: one pair was presented in correct format (e.g., ''The Pacific is the largest ocean'') and the other was incorrect or misleading (e.g., ''The Atlantic is the largest ocean''). Across participants, both stories were read equally often and each fact appeared equally often in correct and misleading formats. Finally, we adapted the story ''The Art Thief'' (Marsh, 2004) for use as a distractor story. All critical facts were removed from this story to prevent unintended interference.

The general knowledge test contained 48 questions, 32 of which corresponded to the critical studied and non-studied facts; 16 questions were fillers and also spanned a wide range of difficulty (Nelson & Narens, 1980; mean difficulty = 46.1,

$SD = 27.49$). Each question was in cued-recall format with a line for recording one's answer and three small boxes labelled "R", "F", and "JK" for Remember, Familiar, and Just Know.

Procedure

The experiment had six phases: story reading, distractor activity, cued-recall test #1, delay, cued-recall test #2, and final questionnaires.

In Phase 1, story reading, participants were allowed up to 5 minutes (with a time warning at 2.5 minutes) to read one of the critical stories. They were instructed to read the story only once and to cover it when finished. Afterwards, all participants answered a brief reading comprehension questionnaire, which began with "Did you finish reading the story?" All participants answered affirmatively.

In Phase 2, distractor activity, participants read the filler story. This was included to increase the plausibility of the cover story (that the experiment was about reading comprehension). Participants again received 5 minutes for reading, and answered comprehension questions afterwards.

Phase 3, cued-recall test #1, was only completed by the participants in the Repeated Test condition. Participants were instructed to answer the questions in sequential order and to only answer each question if they were reasonably sure of the answer (i.e., not to guess). Before taking this test, participants were instructed about the Remember, Familiar, and Just Know classifications. These instructions were adapted and combined from Rajaram (1993, 1996) and Conway et al. (1997). *Remember* responses were described as those for which the participant had a vivid or conscious recollection of an answer being in the story (e.g., could remember which character said it, where it had occurred in the story, what the participant was thinking while reading that piece, etc.). *Familiar* responses were described as responses for which the participant knew the answer had been in the story, but did not have a vivid or conscious recollection of it being in the story. Thus, Familiar responses here parallel the "Know" instructions in most previous studies. Finally, *Just Know* responses were described as answers that were "just known" and could be answered based on one's general world knowledge (i.e., answers known prior to the beginning of the experiment).[1]

It was made clear to participants that while both the Remember and Just Know boxes could be checked for a given answer, or both the Familiar and Just Know boxes, it was impossible to check both the Remember and Familiar boxes. It was also emphasised that although more than one box could be checked, it was equally acceptable to check only one. As each participant was tested alone, the experimenter ensured that all were able to correctly explain (in their own words) the distinctions between the three responses before beginning the test. Participants received as long as they needed to complete this phase, and took approximately 20 minutes.

Phase 4, delay, involved a 48-hour delay. Participants received no instructions regarding the second day of the experiment beyond the fact that they would be fully debriefed at the end of the study.

In Phase 5, cued-recall test #2, each participant took the same cued-recall test as was used in Phase #3 (including the Remember, Familiar, and Just Know instructions). Thus this was a repeated test for half of the participants (delayed test: repeated) and the first test for the other half (delayed test: single).

Finally, in Phase 6 participants were probed about their beliefs about the experiment's purpose. They were explicitly asked whether they noticed

[1] Although our phenomenology instructions differed slightly from those of Conway and colleagues, we do not think this affected our results. We modelled our Just Know instructions on those of Conway and colleagues (1997), albeit a few differences between the two sets of instructions were needed primarily because of the respective testing formats. Specifically, the Conway et al. study utilised recognition tests, and participants selected Just Know when one of the multiple-choice answers simply stood out without any recollection of a specific study episode associated with learning. In contrast, the current study used a cued-recall test, and participants indicated Just Knowing when the answer they provided was based on their general world knowledge. Despite these differences, the core essence of the two definitions is the same as Conway et al. (1997), who told participants that " ... In this case you would not recall a specific episode and instead you would simply know the answer" (p. 398). Thus, in both studies, a Just Know response should have been selected when an answer was chosen or provided based on general knowledge without a specific recollection of the learning episode. Finally, a "guess" response category that Conway et al. (1997) included (see also Gardiner, Java, & Richardson-Klavehn, 1996; Mantyla, 1993) was not included here because participants were explicitly instructed *not* to guess; as such, our instructions for guessing were similar to those of Conway et al.

errors within the story and whether or not they had looked up any of the facts during the 48-hour delay. Finally, all participants received the correct facts as part of the debriefing. Each was thanked, and asked not to discuss this experiment with others.

RESULTS

Our primary aim was to assess the phenomenology of retrieved facts as a function of three key variables—accuracy of information read in the stories, delay, and repeated testing. However, to situate these findings in the context of prior reports, we first report whether these variables affected answers to the general knowledge questions, before turning to a discussion of the phenomenology of retrieved information.

As noted in the Introduction, to pinpoint the role of delay we compared performance on two tests: the immediate test and the single delayed test. To pinpoint the role of repeated testing, we again examined performance on the single delayed test, but instead compared it to the performance observed on the repeated delayed test. This approach was used when examining both the production and the phenomenology of answers to general knowledge questions. Results are significant at the .05 level unless otherwise noted.

Production of answers

Because the studied story contained both correct information and misinformation, it was possible for participants to produce correct or misleading answers on the general knowledge test. We begin with an analysis of *correct answers*: these data appear in the left half of Table 1. Did reading information in the stories (either correct or misleading information) influence participants' abilities to correctly answer general knowledge questions, and if so did this depend on the timing of the test? We first conducted a 2 (information accuracy: correct, misleading) × 2 (story reading: read, not read) × 2 (delay: immediate, 48 hours) ANOVA on the proportion of questions correctly answered. Replicating Marsh et al. (2003), participants answered more questions correctly after having read correct facts rather than misinformation in the stories, $F(1, 94) = 66.02$, $MSE = .02$ (see Table 1). Further, there was an interaction between information accuracy, story reading, and delay, $F(1, 94) = 16.33$, $MSE = .02$.

Reading the correct answer increased correct responding both immediately, $F(1, 47) = 65.51$, $MSE = .03$, and after a delay, $F(1, 47) = 13.36$, $MSE = .02$, but this effect was greater on the immediate test ($M = .53$) than on the delayed test ($M = .37$), $F(1, 94) = 12.10$, $MSE = .02$. This pattern was reversed after reading misinformation. Reading the misleading answer decreased correct responding immediately ($M = .18$), $F(1, 47) = 18.78$, $MSE = .01$, but not after a delay ($M = .24$), $F < 1$. Thus, reading correct facts boosted correct responding on both immediate and delayed tests. In contrast, reading misinformation only lowered correct responding immediately and this adverse effect dissipated over time.

We next asked if repeated testing impacted the ability to correctly answer delayed general knowledge questions. To answer this, we conducted a 2 (information accuracy: correct, misleading) × 2 (story reading: read, not read) × 2 (delayed test: single, repeated) ANOVA on the proportion of questions correctly answered. This analysis revealed that, overall, participants answered more

TABLE 1
Mean proportion of questions answered with the correct and misleading information as a function of story reading, information accuracy, and time of test

	Production of correct information		*Production of misleading information*	
	Story read	*Story not read*	*Story read*	*Story not read*
Correct information in story				
Immediate test	.53	.26	.03	.06
Delay, single test	.37	.26	.07	.06
Delay, repeated test	.51	.28	.04	.06
Misleading information in story				
Immediate test	.18	.28	.39	.04
Delay, single test	.24	.26	.14	.06
Delay, repeated test	.19	.29	.37	.04

questions correctly after reading correct facts rather than misinformation, $F(1, 94) = 55.16$, $MSE = .02$. However, this depended on whether participants had taken a prior test, as revealed by a significant interaction among information accuracy, story reading, and delayed test condition, $F(1, 94) = 12.28$, $MSE = .02$. Follow-up analyses confirmed that reading a correct fact increased correct responding on both the repeated delayed test, $F(1, 47) = 42.36$, $MSE = .03$, and on the single delayed test, $F(1, 47) = 13.37$, $MSE = .02$, but this effect was greater on the repeated ($M = .51$) than on the single delayed test ($M = .37$), $F(1, 94) = 7.13$, $MSE = .03$. In contrast, after reading misinformation in a story, the effects of testing were negative. Reading misinformation in the story decreased correct responding on the repeated delayed test (.19), $F(1, 47) = 22.25$, $MSE = .01$, but did not decrease correct responding for participants who had not taken the immediate test ($M = .24$), $F < 1$. In summary, reading correct facts boosted performance both immediately and after a delay, and repeated testing enhanced these effects. In contrast, reading misinformation decreased correct responding on the delayed test only if participants had also been tested earlier.

We now turn to the production of *misleading answers*. Did reading correct or misleading information in the stories influence participants' tendency to answer general knowledge questions with misinformation, and if so did this depend on the timing of the test? To be clear, a response was scored as a misinformation answer if it was the specific incorrect answer that was used as misinformation in the stories; other wrong answers were not scored as misinformation answers. The data on misinformation answers are on the right side of Table 1. We first conducted a 2 (information accuracy: correct, misleading) × 2 (story reading: read, not read) × 2 (delay: immediate, 48 hours) ANOVA on the proportion of general knowledge questions answered with the misinformation. Replicating previous findings (Marsh et al., 2003), participants answered more questions with the misinformation after reading misleading facts rather than correct information in the stories, $F(1, 94) = 76.97$, $MSE = .02$. However, there was an interaction between information accuracy, story reading, and delay, $F(1, 94) = 37.65$, $MSE = .02$).

Reading the misleading answer increased the likelihood that participants would answer general knowledge questions with the misinformation both immediately, $F(1, 47) = 82.07$, $MSE = .04$, and after a delay, $F(1, 47) = 16.53$, $MSE = .01$, but this effect was greater on the immediate test ($M = .39$) than on the delayed test ($M = .14$), $F(1, 94) = 6.03$, $MSE = .02$. Thus, reading misinformation hindered performance on both tests, but the effects were stronger on the immediate test. In contrast, reading correct information only decreased misleading responding immediately ($M = .03$), $F(1, 47) = 4.16$, $MSE = .01$, not after a delay ($M = .07$), $F < 1$.

We next examined how repeated testing impacted production of misinformation as answers to delayed general knowledge questions. We conducted a 2 (information accuracy: correct, misleading) × 2 (story reading: read, not read) × 2 (delayed test: single, repeated) ANOVA on the proportion of questions answered with the misleading information. In general, participants answered more questions with the misleading information after reading misleading statements rather than correct facts, $F(1, 94) = 64.64$, $MSE = .02$. However, this depended on whether participants had taken a prior test, as revealed by a significant interaction among information accuracy, story reading, and delayed test condition, $F(1, 94) = 28.91$, $MSE = .02$.

Follow-up analyses confirmed that reading the misleading facts increased the tendency to respond with the misinformation on both the repeated delayed test, $F(1, 47) = 69.71$, $MSE = .04$, and on the single delayed test, $F(1, 47) = 16.53$, $MSE = .01$, but this effect was greater on the repeated ($M = .37$) than on the single delayed test ($M = .14$), $F(1, 94) = 30.23$, $MSE = .02$. In contrast, reading a correct fact did not decrease misleading responding on either the repeated delayed test, $F < 1$, or on the single delayed test, $F < 1$, and this did not vary as a function of prior testing. Thus, the effects of testing can be negative: when participants had both read and been previously tested on misinformation, testing hindered final memory performance (in that it increased misinformation answers).

Phenomenology of retrieval

We now turn to the main question of interest in this study: What was the phenomenology of correct and misleading answers? The critical comparison is between memory that includes source information (both "Remember" and "Familiar" judgements) and memory that does not include source information ("Just Know"

judgements). Thus, the Remember response category in all subsequent analyses includes instances where only Remember was selected as well as instances where both Remember and Just Know were selected. Similarly, the Familiar response category includes instances where only Familiar was selected as well as instances where both Familiar and Just Know were selected. Both the Remember and the Familiar definitions include all occasions in which the participant had a specific type of recollective experience (regardless of whether the participant also believed the information to be previously known). As the same patterns of data were obtained when we analysed Remember and Familiar responses given in isolation, we report the more comprehensive measures of recollective experience that also include the Just Know responses. Critically, and in contrast, the Just Know response category includes only instances in which participants selected Just Know in the absence of source memory (i.e., neither Remember nor Familiar was selected).

As expected, when participants provided responses that had *not* been read in the story, they were overwhelmingly judged as "just known". Accordingly our focus is on the phenomenology of the correct answers and of the misleading answers that had been read in the story: Would participants remember reading these answers in the stories, and if so would this change as a function of delay or repeated testing?

We report here conditional probabilities: Given that a question was answered with the information read in the story (either correct or misleading), was it labelled as Remembered, Familiar, or Just Known? Computing these probabilities requires a participant to have at least one correct and one misleading answer (after having read the corresponding facts in the stories), as otherwise the denominator in the calculation is zero. On the basis of this criterion, we excluded 5 participants on the immediate test, 16 participants on the delayed, single test, and 4 participants on the delayed, repeated test. The following analyses therefore include 42 participants in the immediate test condition, 32 participants in the delayed, single test condition, and 43 participants in the delayed, repeated test condition. The patterns of means were consistent when all participants were included. The relevant data are shown in Figure 1.

Phenomenology of retrieval: Effects of delay and accuracy of information

How did Remember judgements change over time and was this similar for correct and misleading answers? We first conducted a 2 (delay: immediate, 48 hours) × 2 (information accuracy: correct, misleading) ANOVA on the probability of giving a Remember response. This analysis revealed a main effect of delay such that participants gave more Remember responses on the immediate test than on the delayed test, $F(1, 72) = 18.69$, $MSE = .16$. Simple effects analyses showed that Remember responses *decreased* over time for both correct, $F(1, 72) = 27.59$, $MSE = .08$, and misleading answers, $F(1, 72) = 5.42$, $MSE = .15$. As can be seen in Figure 1, Remember responses (the black bars) decreased from the immediate test to the delayed, single test for both correct answers (the top portion of the figure) and for misleading answers (the bottom portion of the figure). While Remember responses tended to decrease more over time for correct answers than for misleading answers, this numerical pattern failed to reach significance in a test of interaction between information accuracy and delay, $F(1, 72) = 2.43$, $MSE = .08$, $p = .12$.

Given that the Remember responses decreased over time, how did the Just Know responses for correct and misleading answers change over time? To answer this question, we conducted a 2 (delay: immediate, 48 hours) × 2 (information accuracy: correct, misleading) ANOVA on the probability of a Just Know response. There was a main effect of delay, but in contrast to the previous analyses, participants here gave *more* Just Know responses on the delayed test than on the immediate test, $F(1, 72) = 32.20$, $MSE = .13$. Simple effects tests revealed that this increase over time occurred for both correct, $F(1, 72) = 46.30$, $MSE = .08$, and misleading answers, $F(1, 72) = 9.40$, $MSE = .12$. In other words, unlike Remember responses, Just Know responses *increased* over time for both correct and misleading answers. Interestingly, the interaction between information accuracy and delay was significant, $F(1, 72) = 5.57$, $MSE = .06$, such that Just Know responses increased more over time for correct answers than for misleading answers.

In summary, as can be seen in Figure 1, Just Know responses (the white bars) increased from the immediate test to the delayed, single test for both correct answers (the top portion of the

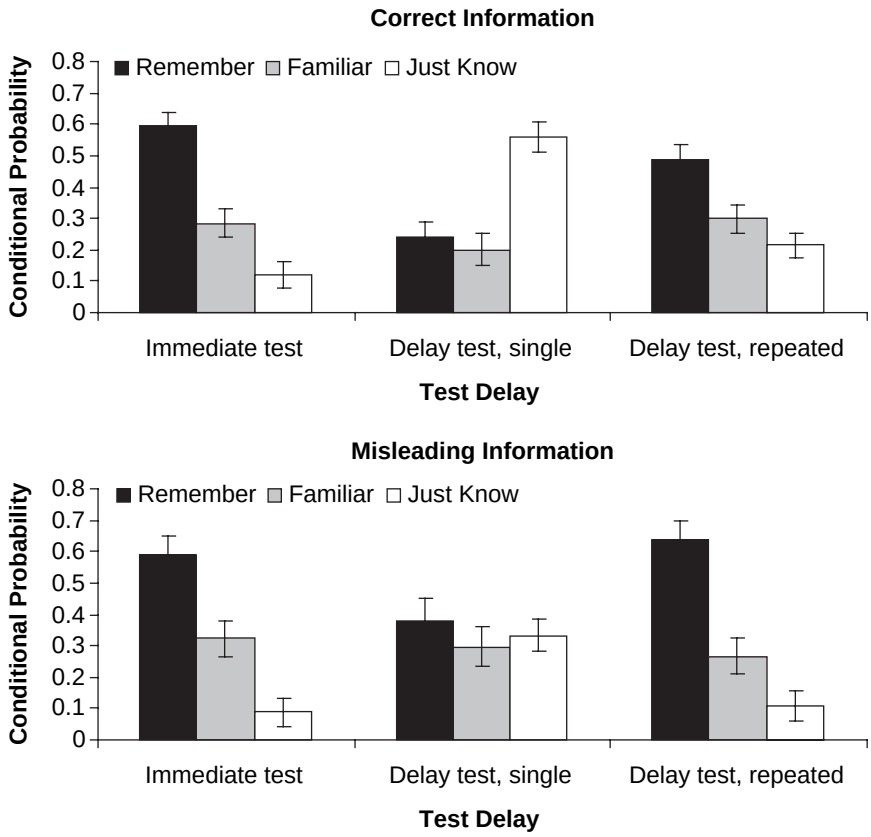

Figure 1. Conditional probabilities of giving a Remember (R), Familiar (F), or Just Know (JK) response for correct answers (after reading the correct answer; top panel) and misleading answers (after reading the misinformation; bottom panel).

figure) and for misleading answers (the bottom portion of the figure). However, this increase was greater for the correct answers than for the misleading answers. Thus, while Remember judgements *decreased* over time, Just Know judgements *increased* over time. Additionally, Remember responses were similar for correct and misleading answers whereas the pattern for Just Know was more pronounced for correct than for misleading answers.

Finally, Familiar judgements showed yet another pattern as a function of delay and accuracy of information (correct or misleading). We conducted a 2 (delay: immediate, 48 hours) $\times$ 2 (information accuracy: correct, misleading) ANOVA on the probability of giving a Familiar response. In contrast to the previous analyses, this test revealed no significant effects such that there was neither a main effect of delay, $F < 1$, nor an interaction between delay and information accuracy, $F < 1$. As can be seen in Figure 1, there is no change in the level of Familiar responses (the grey bars) from the immediate test to the delayed, single test for either correct answers (the top portion of the figure) or for misleading answers (the bottom portion of the figure).

In summary, the effects of delay were different for the three phenomenologies for both the correct and the misleading responses. Remember judgements *decreased*, Just Know judgements *increased*, and Familiar judgements *stayed the same*. Correct and misleading answers showed similar patterns for Remember and Familiar Judgements, but the increase in Just Know responses over time was greater for correct than for misleading answers.

Phenomenology of retrieval: Effects of repeated testing and accuracy of information

We now turn to the effects of repeated testing on the phenomenology experienced on the delayed test. This set of analyses again involves conditional probabilities: Given that a question was answered with the information read in the story

(either correct or misleading), was it labelled as Remembered, Familiar, or Just Known?

First, we conducted a 2 (delayed test: single, repeated) × 2 (information accuracy: correct, misleading) ANOVA on the probability of a Remember response. This analysis revealed a main effect of delayed test condition such that participants gave more Remember responses on the delayed test if that fact had been previously tested, $F(1, 73) = 12.38$, $MSE = .19$. In other words, Remember responses *increased* with repeated testing. Furthermore, this analysis revealed no significant interaction between information accuracy and delayed test condition, $F < 1$. Planned simple effects analyses indicated that repeated testing increased Remember responses for both correct, $F(1, 73) = 10.86$, $MSE = .10$, and misleading answers, $F(1, 73) = 6.58$, $MSE = .19$. As shown in Figure 1, fewer Remember responses (the black bars) occurred with the delayed, single test than with the delayed, repeated test for both correct answers (the top portion of the figure) and for misleading answers (the bottom portion of the figure). For both correct and misleading information, the conditional probability of giving Remember responses *increased* with repeated testing.

Repeated testing had a different effect on Just Know responses, as revealed by a 2 (delayed test: single, repeated) × 2 (information accuracy: correct, misleading) ANOVA on the probability of a Just Know response given to an answer that was previously read in the story. This analysis revealed a main effect of delayed test condition, but in contrast to Remember responses, participants gave more Just Know responses on the delayed test if that fact had *not* been previously tested, $F(1, 73) = 20.81$, $MSE = .14$. In other words, Just Know responses *decreased* with repeated testing, and simple effects tests showed this occurred for both correct, $F(1, 73) = 22.34$, $MSE = .10$, and misleading answers, $F(1, 73) = 8.31$, $MSE = .11$. The interaction between accuracy of information delayed test condition did not reach significance, $F(1, 73) = 1.96$, $MSE = .07$, $p = .17$. As can be seen in Figure 1, there was a higher level of Just Know responses (the white bars) associated with the delayed, single test than with the delayed, repeated test for both correct answers (the top portion of the figure) and for misleading answers (the bottom portion of the figure). Thus, while Remember judgements *increased* with repeated testing, Just Know judgements *decreased* with repeated testing for both correct and misleading information.

Finally, Familiar responses yielded a third pattern of responses. There were no significant results from the 2 (delayed test: single, repeated) × 2 (information accuracy: correct, misleading) ANOVA on the probability of a Familiar response given to an answer that had been read in the story. Participants gave an equivalent number of Familiar responses on the delayed tests regardless of repeated testing condition, $F < 1$. Similarly, this analysis revealed no significant interaction between repeated testing condition and accuracy of information, $F(1, 73) = 2.25$, $MSE = .07$, $p = .14$. As seen in Figure 1, there is an equivalent level of Familiar responses (the grey bars) associated with the delayed, single test and with the delayed, repeated test for both correct answers (the top portion of the figure) and for misleading answers (the bottom portion of the figure).

In summary, repeated testing *increased* Remember responses on the delayed test; repeated testing served as a re-learning episode (Roediger & Karpicke, 2006) and enhanced source memory. In contrast, repeated testing *decreased* Just Know responses. Familiar responses *did not show a pronounced change* as a function of repeated testing.

All the analyses of phenomenologies involved conditional probabilities; as noted earlier, the use of conditional probabilities meant that not all participants could be included in the analyses. Thus the data were also analysed using joint probabilities: the probability that a question was answered with the studied information *and* was associated with a particular phenomenological state. The conclusions drawn from the conditional and joint probability analyses yielded nearly identical conclusions (a few discrepancies emerged due to floor effects in some of the joint probability analyses). Therefore we focused our discussion on the conditional probabilities to allow direct comparison between the responses associated with correct and misleading information. Since participants answered fewer questions with read misinformation than with read correct information (see Table 1), it is impossible to directly compare the joint probabilities associated with correct and misleading answers. In contrast, the conditional probability analyses take into account the different base rates of responding and allow for direct comparison between the

responses associated with correct and misleading answers.

DISCUSSION

The results of the study were straightforward: the phenomenology of retrieved facts changed as a function of repeated testing and delay. Repeated testing increased the likelihood a fact would be *remembered*, whereas delay was associated with an increase in *just knowing*. The two key variables had different effects on phenomenology, supporting the separation of Know into two distinct senses (Conway et al., 1997). Unlike Remember and Just Know responses, Familiar responses did not change in response to delay or repeated testing; the patterns for Remember and Familiar replicate prior work (e.g., Gardiner & Java, 1991) and contrast with the novel findings for Just Know.

Repeated testing increased the likelihood that correct answers would be *remembered.* This is consistent with laboratory work showing that testing has its effects on recollection rather than familiarity. Consider a recent study by Karpicke, McCabe, and Roediger (2006), in which they used Jacoby's (1991) process dissociation procedure to estimate the contributions of recollection and familiarity to the testing effect. It is well established that re-testing is better than re-studying; students who are asked to recall a passage after a delay do better if they had attempted to recall it repeatedly between the initial study and the final test (the Study-Test-Test-Test condition; STTT) than if they had studied it repeatedly between the initial study and the final test (the Study-Study-Study-Study condition; SSSS), (Roediger & Karpicke, 2006). Of current interest, Karpicke and colleagues used two different final tests (inclusion and exclusion) that together allow for estimations of recollection and familiarity. In the Inclusion condition, participants had to recall 20 studied exemplars of each category, even if they had to guess. In the Exclusion condition, participants were asked to produce new non-studied exemplars. Based on the data from these two conditions, *recollection* and *familiarity* were estimated, and the advantage for the repeated testing (STTT) condition over the repeated study (SSSS) condition at the delay was completely in recollection, not in familiarity.

In our study, delay (in the absence of repeated testing) was associated with an increase in Just Knowing, in contrast to the effects observed with repeated testing. It is especially intriguing that the current work yielded a shift to Just Knowing over time given that we used a much shorter delay (2 days) than did Conway et al. (who retested participants after 25 weeks). Our findings show that a sense of Just Knowing is not limited to information that has been schematised through repeated exposure, considerable integration, and extensive delays between learning and testing, as would be the case for course material (Herbert & Burt, 2003; see also Conway et al., 1997). Our findings support a second sense of Just Knowing, one that can occur for information encountered more recently (in the present case, 2 days). Rather than schematisation, recently encountered information is Just Known when its retrieval fluency is interpreted as retrieval of one's general world knowledge.

These two senses of Just Knowing can be contrasted with a third sense of Knowing, as typically used in episodic memory experiments; a sense of familiarity in which the general context of the study encounter is known, but the recalled information lacks a sense of immediacy or details that characterise mental travel back to the episode.

Our study design allowed us to examine another novel issue—the phenomenologies of both correct and misleading answers learned from the stories and their shifts as a function of repeated testing and time. As described in the introduction, the literature on the phenomenology of memory errors is relatively small, and our results add to this growing literature. Testing increased Remember responses for misleading answers, which is consistent with results from the DRM and eyewitness paradigms. Of greater interest were the effects of delay, as the false memory literature is mixed regarding the relationship between delay and phenomenology. We found that delay decreased Remember responses for misleading answers, and that this decrement was similar to that observed with correct answers. However, delay led to a larger increase in Just Know responses for correct answers than for misinformation answers. In other words, on the single delayed test, the predominant response for correct answers was ''just know''. In contrast, there was no predominant response for misleading answers. This suggests that the switch from Remembering to Just Knowing over time may be at least partially dependent on pre-existing knowledge (which is assumed to be strongly

present for correct answers and either absent or only weakly present for misinformation; Marsh, Balota, & Roediger, 2005).

We note that these novel findings about the phenomenology of retrieval were supported by several anticipated findings about the effects of story reading on ability to answer general knowledge questions. Consistent with Marsh et al. (2003), we found that reading correct answers during study boosted production of correct answers on both immediate and delayed tests and that testing further enhanced this effect. In contrast, reading misinformation reduced correct responding, but only on the immediate test. On the delayed test, if participants had not been tested in the first session there was a rebound in the production of prior, correct knowledge. On both immediate and delayed tests participants answered with misinformation read in the stories, and testing increased this effect. Together, these findings once again show the power of repeated testing in maintaining both correct and incorrect information over time.

In closing, ''just knowing'' is a useful judgement to index knowledge about the world, as opposed to a general sense of familiarity associated with past events. Our findings show that this sense of Just Knowing increases with the passage of time and, interestingly, occurs more often for material that may already be represented in memory. Although the concept of semantic memory is not without controversy, this description of Just Knowing maps nicely onto the original concept of semantic memory.

Manuscript received 25 February 2008
Revised manuscript received 22 July 2008

REFERENCES

Chan, J. C. K., & McDermott, K. B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology*, *20*, 633–639.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*, 393–413.

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

Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin and Review*, *15*, 1081–1089.

Frost, P. (2000). The quality of false memory over time: Is memory for misinformation 'remembered' or 'known'? *Psychonomic Bulletin & Review*, *7*, 531–536.

Gardiner, J. M. (2008). Remembering and knowing. In H. L. Roediger III (Ed.), *Cognitive psychology of memory* (pp. 285–305) [Vol. 2 of ''Learning and memory: A comprehensive reference'', 4 vols., J. Byrne, Ed.]. Oxford, UK: Elsevier.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*, *19*, 617–623.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How levels of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology*, *50*, 114–122.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, *16*, 107–112.

Herbert, D. M. B., & Burt, J. S. (2001). Memory awareness and schematization: Learning in the university context. *Applied Cognitive Psychology*, *15*, 617–637.

Herbert, D. M. B., & Burt, J. S. (2003). The effects of different review opportunities on schematisation of knowledge. *Learning and Instruction*, *13*, 73–92.

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

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306–340.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Karpicke, J. D., McCabe, D. P., & Roediger, H. L. III. (2006). *Testing enhances recollection: Process dissociation estimates and metamemory judgements.* Poster presented at the annual meeting of the Psychonomic Society, Houston, TX.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*, 1–24.

Lane, S. M., & Zaragoza, M. S. (1995). The recollective experience of cross-modality confusion errors. *Memory & Cognition*, *23*, 607–610.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271.

Mantyla, T. (1993). Knowing but not remembering: Adult age differences in recollective experience. *Memory & Cognition*, *221*, 379–388.

Marsh, E. J. (2004). Story stimuli for creating false beliefs about the world. *Behavior Research Methods, Instruments & Computers*, *36*, 650–655.

Marsh, E. J., Balota, D. A., & Roediger, H. L. III. (2005). The effects of aging and mild dementia on learning facts from fiction. *Neuropsychology*, *19*, 115–129.

Marsh, E. J., Meade, M. L., & Roediger, H. L. III. (2003). Learning facts from fiction. *Journal of Memory and Language*, *49*, 519–536.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgements and memory characteristics questionnaire compared. *Memory & Cognition*, *25*, 826–837.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior*, *19*, 338–368.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*, 261–285.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89–102.

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 365–377.

Rajaram, S. (1999). Assessing the nature of retrieval experience: Advances and challenges. In B. H. Challis & B. Velichkovsky (Eds.), *Stratification in cognition and consciousness* (pp. 255–275). Amsterdam, Netherlands: John Benjamins Publishing Company.

Roediger, H. L. III, Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language*, *35*, 300–318.

Roediger, H. L. III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

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

Roediger, H. L. III, Rajaram, S., & Geraci, L. (2007). Three forms of consciousness in retrieving memories. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *Cambridge handbook of consciousness*. New York: Cambridge University Press.

Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language*, *35*, 319–334.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Oxford, UK: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12.

Tulving, E. (1995). Organization of memory: Quo vadis? In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 839–847). Cambridge, MA: MIT Press.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, 152–176.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.