Metacognition Learning (2015) 10:279–292
DOI 10.1007/s11409-014-9125-z

If it is stored in my memory I will surely retrieve it: anatomy of a metacognitive belief

Nate Kornell

Received: 17 August 2013 / Accepted: 24 September 2014 / Published online: 5 October 2014

Abstract Retrieval failures—moments when a memory will not come to mind—are a universal human experience. Yet many laypeople believe human memory is a reliable storage system in which a stored memory should be accessible. I predicted that people would see retrieval failures as aberrations and predict that fewer retrieval failures would happen in the future. After responding to a set of trivia questions, participants were asked whether they would do better, about the same, or worse if they were given a different, but equally difficult, set of questions to answer. The majority of participants said they would do about the same, but more participants said they would do better next time than said they would do worse, although these participants did not actually do better. This finding was especially pronounced when participants were given feedback, suggesting that hindsight bias—the feeling, which emerges when an answer is presented, that one knew it all along—contributed to participants' belief that they had underperformed on the first set of questions. The finding that metacognitive judgments were influenced by beliefs stands out in a literature full of studies in which beliefs fail to influence judgments.

Keywords Memory · Metacognition · Overconfidence · Hindsight bias · Optimism

Having an accurate understanding of how memory works is can help a person in many ways, including with mundane problems like remembering a shopping list as well as with life-and-death decisions by jurors. Many studies have examined people's beliefs about how memory works (e.g., Bjork et al. 2013; Magnussen et al. 2006), especially with respect to the influence of metacognitive judgments on decisions about how to study (Kornell and Bjork 2007; Wissman et al. 2012).

Two theoretical issues lie at the heart of research on metacognitive beliefs. How accurate are beliefs? And when and how much do beliefs affect metacognitive judgments (and subsequent behaviors)? There is no single answer to the first question. Some beliefs are accurate; for example, people know they forget over time and learn by studying (although they do not necessarily apply these beliefs, Koriat et al. 2004; Kornell and Bjork 2009). Other beliefs are not at all accurate; for example, Guilmette and Paglia (2004) found that 41 % of respondents

N. Kornell (✉)
Department of Psychology, Williams College, Williamstown, MA 01267, USA
e-mail: nkornell@gmail.com

agreed with the statement, "sometimes a second blow to the head can help a person remember things that were forgotten [as a result of a first blow to the head]."

The present research focused on the second question, specifically with respect to when, and how much, beliefs about memory affect metacognitive judgments. Although there is no single answer to this question either, it is a topic of great interest, and considerable disagreement, in the metacognition literature. Some argue that for the most part, metacognitive judgments are guided by subjective experience (such as the fluency with which an answer comes to mind; e.g., Benjamin and Bjork 1996; Benjamin et al. 1998) and beliefs do not influence judgments unless the beliefs are made salient (e.g., Besken and Mulligan 2013; Bjork et al. 2013; Rhodes and Castel 2008, 2009; Jacoby and Kelley 1987; Koriat 1993, 1997; Koriat et al. 2004; Kornell et al. 2011; Schwartz et al. 1997). Recent research, though, suggests that beliefs do affect judgments even when the beliefs are not made salient, and, in fact, argues that some previous findings attributed to the effects of subjective experience were actually driven by beliefs (e.g., Mueller et al. 2013, 2014).

The present research

The research reported here examined the implications of a common belief in the infallibility of memory. Memory is the way people "record" their experiences, but the way memory functions, and its fallibility, make it almost entirely unlike a video camera (e.g., Bjork 1989; Loftus 1996). Yet in a nationally representative sample of 1,500 people in the United States, 63 % of respondents agreed that "human memory works like a video camera, accurately recording the events we see and hear so that we can review and inspect them later" (Simons and Chabris 2011). This misunderstanding is one among many, but it has important implications for the present research: If people think their memories are largely reliable, then retrieval failures—moments when a memory is stored but cannot be accessed—should not occur very often. The reality of memory is that retrieval failures are not uncommon. I predicted that if people experience retrieval failures, they will perceive a mismatch between what they expected to happen (minimal retrieval failures) and what actually happened (non-minimal retrieval failures) and they will assume the retrieval failures must have been aberrations that won't happen again, and thus expect to experience fewer retrieval failures—and, thus, do better—next time.

In the studies reported here, participants were asked 40 trivia questions and then they were asked if they thought they would do better, the same, or worse if they tried the same task again with a new set of questions. As Fig. 1 shows, I hypothesized that many participants would experience retrieval failures during the trivia task, assume their retrieval failures were aberrations because of the implicit belief that their memories are reliable and that a memory that is stored should be retrievable, and thus predict that they would do better next time.

Hindsight bias—when people think they "knew it all along"—is included alongside retrieval failure in the first box in Fig. 1 because when people are shown the answer to a

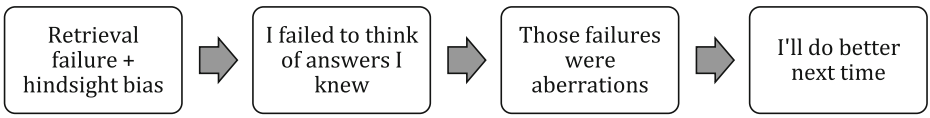

Fig. 1 Hypothesized process that leads people to think they will do better next time after answering trivia questions

trivia question, they often believe they could have thought of the answer even if they could not (Fischhoff 1975; Guilbault et al. 2004). Hindsight bias does not tend to occur unless someone is shown the correct answer (i.e., given feedback). If hindsight bias plays a role in initiating the process depicted in Fig. 1, then feedback should increase the frequency with which participants say they will do better next time. Experiments 1 and 2 tested this hypothesis.

The reason why the psychological process depicted in Fig. 1 is likely to lead people to draw incorrect conclusions is captured in the third box. While it is true that a specific retrieval failure might not occur again, it is a mistake to assume retrieval failures will not occur in aggregate. Making this faulty assumption should cause people to expect that they will do better next time.

The effects of metacognitive beliefs

Most previous research on the influence of metacognitive beliefs has examined situations in which a) there was a fairly clear connection between the relevant belief and the metacognitive judgment being made and b) applying one's beliefs could be expected to make metacognitive judgments more accurate. For example, Koriat et al. (2004) asked people to predict how they would do on a future memory test that would take place after one of a variety of retention intervals. Retention interval was mentioned in the judgment prompt (e.g., you will be tested in a week) and judgments were far more accurate when people applied their beliefs about forgetting.

The present study examined the influence of beliefs using an approach that was novel in three ways. First, the question participants were asked (how will you do on the next set of trivia questions?) does not make any reference to the belief at hand (because memory is reliable, retrieval failures are aberrations). Second, applying this incorrect belief is likely to make metacognitive judgments less accurate. Third, there was a compelling reason for participants to ignore their belief and say they would do the same as they had done before. The task participants were asked to judge was no different than the task they had completed mere seconds before, and there was no reason to expect that they had improved at the task (given that answering trivia questions does not improve one's knowledge about other trivia questions). For a participant to predict that he or she would do better requires participants to be optimistic to a degree that seems somewhat irrational (as the General discussion explains further.)

In short, previous studies have shown that people fail to apply their beliefs when a) it is somewhat obvious that they should and b) they have an incentive to do so in order to improve their metacognitive accuracy. The present experiment tested the opposite extreme, in the sense that a) the relevant metacognitive belief was not obvious and b) there was an incentive to *not* apply one's metacognitive beliefs (stemming from the high likelihood that participants would do equally well on the next set of questions). If participants in the present studies consistently say they will do better on the next set of trivia questions, it would suggest that beliefs can have a powerful influence on metacognitive judgments. Such a finding would provide balance to a literature that currently paints a picture in which people obstinately fail to apply metacognitive beliefs even when they should.

Experiment 1

In Experiment 1 participants were asked 40 trivia questions. Half of the participants were given correct answer feedback after each response and half were not. At the end of the experiment, participants were asked whether they thought they would do better, worse, or about the same if they were asked another equally difficult set of questions.

Method

Participants Eight hundred participants were recruited from Amazon's Mechanical Turk. Half were given feedback after responding to each question and half were not. Twenty-nine participants were excluded because they indicated that they had encountered the same learning materials in another study, and 12 additional participants were excluded because they did not complete the entire experiment. The data analyses included 759 participants; 381 participants were given feedback and 378 were not.

Participants were paid $1.00 for completing the study. They all reported that they were fluent English speakers living in the United States. Two did not report their gender, 370 were female and 387 were male. The average age was 31.6 years (range: 18–72).

Materials and design The materials were 40 trivia questions taken from Nelson and Narens (1980). Examples include, "what is the last name of the villainous captain in the story 'Peter Pan'?" and, "what is the last name of the commander who lost the battle of Little Bighorn?"

There were two between-participant groups; one was given feedback and one was not. Participants in the feedback and no feedback condition were run on two separate, consecutive days.

Procedure The study was conducted online. Participants were asked 40 trivia questions, one by one. They were given unlimited time to answer each question. In the feedback condition, after submitting their response, participants were shown the correct answer for 2 s and then the next trial began. In the no feedback condition, the next trial began immediately. Question order was randomized uniquely for each participant.

After completing the trivia task, participants were asked the following question: "How do you think you'd do if we asked you another set of questions like those? Assume none of the questions would be the same, but you'd be asked the same number of questions and they'd be about equally difficult. Compared to how you just did, do you think you'd do better, worse, or the same? (Type better, worse, or same.)"

I expected participants would select "better" more often than "worse." If such a pattern emerged, one explanation would be hindsight bias and the belief that retrieval failures are aberrations. I also tested two alternate explanations, namely that some people a) hold a general belief that luck controls their fate, or b) have an expectation that a sequence of bad outcomes will be followed by a sequence of good outcomes even if nothing has changed. Thus, participants were asked three additional questions at the end of Experiment 1. The first question was, "imagine that you were flipping a fair coin and you got tails eight times in a row. Do you think the next flip would be more likely to be heads or tails, or do you think they're equally probable?" A sizeable proportion of people say "heads" in response to questions like this, despite the fact that coins do not have memories (Tversky and Kahneman 1971). I hypothesized that participants who gave an answer other than "they're equally probable" might believe their fates are controlled by luck or fate, which would increase their willingness to endorse the notion that they had been unlucky on the first list of trivia questions, and lead them to predict that they would do better on the next set of trivia questions. The second and third questions were "have you purchased a lottery ticket in the past 30 days?" and "outside of lottery tickets have you done any gambling or betting in the last 30 days?" These questions were included based on the hypothesis that people who play the lottery seem to remain optimistic about their future prospects of winning even in the face of past evidence that they (like almost everyone) have a losing record playing the lottery, and that such participants might make a similar prediction that their future prospects in the trivia task were

better than their past performance. If participants' answers to one or more of these three questions were correlated with their tendency to say they would do better next time, it would provide support for the hypothesis that global beliefs about luck and fate played an important role in their decision making, which might also suggest that their metacognitive beliefs about memory played a less important role.

Results

The primary question was whether participants were biased to say they were going to do better on the next set of questions. The rate at which participants said "better" is not informative in isolation because unbiased participants would say "better" and "worse" at equal rates. Instead, I examined the ratio of "better" to "worse" responses because the more participants lean toward saying better, the greater this ratio should become.

When participants were asked how they would do if they tried the task again, they said better more often than they said worse (Table 1). The difference between better and worse (excluding "same" responses) was significant in the feedback condition $X^2=30.12$, $p<0.0001$. It was not significant in the no feedback condition, $X^2=2.44$, $p=0.118$. The ratio of better to worse responses was greater in the feedback condition than in the no feedback condition, $\phi=0.218$, $p=0.013$.

Table 2 shows responses to the other three questions asked at the end of Experiment 1. Participants' responses to the questions about trivia questions (in Table 1) were not significantly related to their responses to any of the three questions in Table 2 (unlike the analyses in the previous paragraph, these analysis included all response categories, including "the same"). Thus, there was no support for the hypothesis that beliefs about luck would help explain participants' tendency to say they would do better on the next set of trivia questions.

Proportion correct in the trivia task is displayed in Table 3. Responses were scored as correct even if they were not spelled correctly. There was a main effect of responses to the better/same/worse question, $F(2, 753)=5.71$, $p=0.003$, $\eta_p^2=0.012$. Participants who said "same" did the best on the trivia questions, followed by participants who said "better" and

Table 1 Number (and percentage) of participants who said they would do better, the same, or worse if they did the study again with new questions, in Experiments 1–4

	Better	The same	Worse	Total
Experiment 1				
Feedback	67 (18 %)	298 (78 %)	16 (4 %)	381
No feedback	36 (10 %)	319 (84 %)	23 (6 %)	378
Experiment 2				
Feedback	14 (40 %)	21 (60 %)	0 (0 %)	35
No feedback	7 (14 %)	37 (74 %)	6 (12 %)	50
Experiment 3				
Feedback	17 (21 %)	60 (73 %)	5 (6 %)	82
Experiment 4				
Feedback/Self	9 (24 %)	28 (74 %)	1 (3 %)	38
Feedback/Others	8 (21 %)	16 (42 %)	14 (37 %)	38

In Experiment 4, participants were asked how they would do if they did the study again (Self) and how they did compared to their peers (Other)

Table 2 Number (and percentage) of participants who chose each option in response to the final three questions in Experiment 1

Following a streak of tails, what is likely to come next?			
Heads	Equally probable	Tails	Total
147 (19 %)	561 (74 %)	51 (7 %)	759
Have you played the lottery in the last 30 days?			
Yes	No		Total
144 (19 %)	615 (81 %)		759
Have you done other gambling in the last 30 days?			
Yes	No		Total
98 (13 %)	661 (87 %)		759

then participants who said "worse." Post-hoc Tukey tests showed that only the difference between "same" and "worse" was significant.

There was also a significant main effect of feedback on proportion correct, $F(1, 753)=9.18$, $p=0.003$, $\eta_p^2=0.015$, with better performance in the no feedback condition. The reason for this difference is unclear; it is impossible to rule out the possibility that there were preexisting differences between the groups, given that the participants in the feedback condition all completed the study before any of the participants in the no feedback condition. This explanation seems unlikely given that this result was replicated in Experiment 2, in which feedback condition was assigned randomly. It seems more likely that participants in the no feedback condition, who knew that they would not find out the answer unless they thought of it by themselves, tried harder because they were curious about some of the answers. More

Table 3 Proportion of trivia questions answered correctly (standard deviation in parentheses) in Experiments 1–4 as a function of whether participants thought they would do better, the same, or worse if they did the study again with new questions

	Better	The same	Worse	Total
Experiment 1				
Feedback	0.74 (0.16)	0.77 (0.17)	0.65 (0.15)	0.76 (0.17)
No feedback	0.78 (0.15)	0.81 (0.15)	0.76 (0.18)	0.80 (0.15)
Experiment 2				
Feedback	0.45 (0.23)	0.63 (0.27)	–	0.56 (0.26)
No feedback	0.49 (0.28)	0.66 (0.18)	0.72 (0.15)	0.64 (0.20)
Experiment 3				
Feedback/1st half	0.48 (0.23)	0.52 (0.20)	0.46 (0.34)	0.51 (0.22)
Feedback/2nd half	0.46 (0.21)	0.51 (0.22)	0.40 (0.26)	0.49 (0.22)
Experiment 4				
Feedback/Self	0.66 (0.14)	0.57 (0.15)	0.45 (–)	0.58 (0.15)
Feedback/Others	0.70 (0.14)	0.61 (0.14)	0.48 (0.12)	0.58 (0.15)

In Experiment 2 there were no observations in the Feedback/Worse condition. In Experiment 3 data from the first and second half of the trivia questions are presented. In Experiment 4 the data are separated by how people thought they would do if they did the study again (Self) and by how they thought they did compared to other participants in the study nOthers); because both rows are based on the same dataset, the values in the total column are necessarily the same. Also, in Experiment 4, there was only one observation in the Feedback/Self/Worse condition, so standard deviation could not be computed

effort should translate into longer time spent trying to retrieve answers. The data did not provide support for this hypothesis, however: Median response time, computed for each participant, was numerically higher on average in the No Feedback condition (M=8.56, SD=4.43) than in the Feedback condition (M=8.10, SD=3.01), but the difference was not significant, $t(757)$= 1.657, p=0.098.

In summary, the majority of participants expected to do about the same if they answered another set of questions, but among those who did not expect to do the same, more said they would do better than said they would do worse, especially when participants were given feedback. These findings suggest that people may see retrieval failures as aberrations, especially when feedback awakens their hindsight bias. An alternate hypothesis, that beliefs about fate or luck contributed to participants' tendency to say "better," was not supported.

Experiment 2

Many participants in Experiment 1 seemed to believe they would do better if they tried again even though their task was going to be the same. Because this attitude seems irrationally overconfident, I replicated Experiment 1 using a different sample (UCLA students) and a different phrasing of the question of primary interest.

Method

Participants Eighty-five UCLA students were given course credit for participating. Due to random assignment, there were 35 in the feedback condition and 50 in the no feedback condition. Because of an experimental error, demographic information about age and gender were not recorded. Although this oversight is unfortunate, it does not impact the central conclusions of the research, because neither age nor gender was a variable of concern in any of the predictions or analyses.

Procedure The materials and design were the same as in Experiment 1. The procedure was also the same, with the following exceptions. First, the questions about coin tosses, the lottery, and gambling were not asked. Second, feedback was shown for 3 s instead of 2. Third, the primary question was reworded as follows: "How do you think you'd do if you did this experiment again with an equivalent set of questions (assuming the total number of questions was the same, but none of the questions you were just asked were repeated)? Compared to how you just did, do you think you'd do better, worse, or the same?" Participants were shown three buttons, in a vertical arrangement, labeled "Better," "The same," and "Worse."

Results

As Table 1 shows, participants in the feedback condition said they would do better more often than they said they would do worse, X^2=12.08, p=0.0005. This effect did not occur in the no feedback condition, X^2=0, p=1. The feedback and no feedback conditions differed significantly with respect to the ratio of "better" to "worse" responses, ϕ=0.555, p=0.006.

The recall data also mostly replicated Experiment 1, although UCLA students tended to score lower than Amazon's Mechanical Turk participants (Table 3). It was not possible to conduct a 2 (feedback) x 3 (final question response) ANOVA because no participant in the feedback condition chose the "worse" response. Instead, I conducted two separate ANOVAs.

There was a significant effect of response type, $F(2, 82)=6.44$, $p=0.003$, $\eta_p^2=0.136$. Unlike Experiment 1, participants who said they would do worse (of whom there were only 6) did the best ($M=0.721$, $SD=0.147$), followed by participants who thought they would do the same ($M=0.647$, $SD=0.214$), and then participants who thought they would do better ($M=0.462$, $SD=0.242$). Tukey post-hoc tests showed that the group who said "better" performed significantly worse than the other two groups, which did not differ from each other. The main effect of feedback was not significant, $F(1, 83)=2.90$, $p=0.093$ (numerically, however, it was in the same direction as Experiment 1, with participants who were not given feedback achieving numerically higher means than those who were given feedback; see Table 3).

In sum, Experiment 2 conceptually replicated Experiment 1 using a rephrased final question and different participant population. Participants thought they would do better more often than they thought they would do worse, although this effect only occurred in the feedback condition.

Experiment 3

Predicting that you will do better next time, when the situation has not changed, seems irrational. However, it is not irrational if it is true, and it is possible that participants who predicted they would do better had good reasons for doing so (e.g., they had intentionally answered incorrectly, or they were not paying attention, etc.). Thus, Experiment 3 tested whether participants who expected to get better actually got better. Participants were asked 40 trivia questions and then they were asked how they would do next time; unlike the previous studies, they were then asked another 40 questions.

Method

Participants Eighty-two UCLA students were given course credit for participating. Due to an experimental error their gender and age were not recorded.

Materials The materials consisted of 80 trivia questions, 40 of which were the same ones used in Experiments 1 and 2. The other 40 were drawn from the same pool of questions (Nelson and Narens 1980). Because the first set consisted of mainly easy questions, the pool of items did not include another 40 items of similar difficulty, so the new set of questions had to be more difficult. (This difference was confirmed by the average accuracy during the question phase, which was 0.64 for the old set of questions and 0.36 for the new set.) Because the 80 questions were mixed together and ordered randomly for each participant, the first set of 40 questions that a given participant saw was not, on average, more or less difficult than the second set.

Procedure The procedure was identical to Experiment 2 with three exceptions. 1) As mentioned above, the questions were more difficult. 2) All participants were given feedback. 3) After participants made their predictions about how they would do on a second set of questions, they were asked a second set of questions.

Results

Participants said they would do better significantly more often than they said they would do worse, $X^2=5.50$, $p=0.019$ (see Table 1). Performance on the trivia questions (Table 3) was not significantly related to the participants' better/same/worse responses, $F(2, 79)=0.68$, $p=0.509$,

η_p^2=0.017, or to whether they were responding during the first or second half of the study, $F(1, 79)=2.013$, $p=0.160$, $\eta_p^2=0.025$, nor was there an interaction, $F(2, 79)=0.52$, $p=0.597$, $\eta_p^2=0.012$. Moreover, planned comparisons showed that participants who said they would do better did not do significantly better, $t(16)=0.45$, $p=0.656$, $d=0.067$, nor did participants who said they would do worse actually do significantly worse, $t(4)=1.13$, $p=0.320$, $d=0.218$.

One might hypothesize that the reason participants in these studies said they would do better on a second set of trivia questions is that they planned to try harder. When people try harder to remember something, they usually spend more time trying to remember. Thus, if this hypothesis is true, participants in Experiment 3 should have spent more time trying to retrieve in the second half of the study than in the first. To test this hypothesis I computed a median reaction time for each participant for the first and second half of the study. Participants spent more seconds trying to retrieve in the first half of the study ($M=13.19$, $SD=14.13$) than the second half ($M=10.69$, $SD=9.68$), $t(81)=3.405$, $p=0.001$. This finding suggests that the reason participants said they would do better was not because they decided to try harder next time.

In sum, participants in Experiment 3 predicted that they would do better more often than they predicted that they would do worse, but these predictions were not accurate.

Experiment 4

People are frequently overconfident and overly optimistic (e.g., Shepperd et al. 2013). Some forms of overconfidence are comparative; people tend to think they are above average, compared to their peers, in all sorts of situations (e.g., Alicke et al. 1995; Chambers and Windschitl 2004; Kruger and Dunning 1999). Other forms of overconfidence are absolute, not comparative; for example, people underestimate the time it will take them to complete a new task—the so-called planning fallacy—even if they have fallen victim to the planning fallacy repeatedly in the past (Kruger and Evans 2004; Kahneman and Tversky 1979). The current study is a hybrid of these two types of judgments: Participants made comparative, not absolute, judgments, but they did not compare themselves to a peer group. Instead, each participant compared himself with himself or herself with herself.

In Experiment 4, participants were asked to predict how well they would do if they did the task again, but they were also asked to judge their performance compared to their peers. Experiment 4 tested two competing hypotheses. If simple overconfidence were at work, one would expect participants to rate themselves as better than their peers. However, if my earlier hypothesis was correct, participants should believe they underperformed on the questions they just finished answering, which might lead them to think they did *worse* than their peers.

Method

Participants Thirty-eight UCLA students were given course credit for participating. Due to an experiment experimental error their gender and age were not recorded.

Procedure The materials and procedure were the same as those in Experiment 2, with three exceptions. First, all participants were given feedback. Second, after indicating whether they expected to do better, the same, or worse if they were asked another set of questions, participants were asked another question: "Compared to other students in your class who have participated in this experiment, how do you think you did?" The options were better, the same, or worse. Third, during the trivia phase of the study, when participants could not think of

an answer, they were asked to type in "dk" if they did not know the answer or "cr" if they knew the answer but could not remember it at the moment. In retrospect, this measure was not well conceived because it was meant to examine the effect of hindsight bias, but participants were forced to respond before seeing feedback, and thus before hindsight bias could set in. Thus, this rating will not be discussed further. There is no obvious reason to expect that asking would bias, or even affect, the key dependent variables in the study, namely, participants' responses to the two questions at the end of the study.

Results

Participants said they would do better on significantly more often than they said they would do worse when asked how they would do on a second list, $X^2=4.9$, $p=0.027$ (see Table 1). When asked how they had done compared to their peers, however, there was no significant difference between the rate at which they said they had done better versus worse, $X^2=1.14$, $p=0.286$ (see Table 1).

Recall performance (Table 3) was not significantly related to whether participants said they would perform better, the same, or worse if they did the task again, $F(2, 35)=1.61$, $p=0.214$, $\eta_p^2=0.084$. How participants thought they had done compared to their peers did have a significant relationship with trivia accuracy, $F(2, 35)=7.40$, $p=0.002$, $\eta_p^2=0.297$. Post-hoc Tukey tests showed that participants who said they had done better actually did do better than the other two groups of participants, which did not differ significantly.

In sum, when asked how they would do if they tried the trivia task again, participants said they would do better more often than they said they would do worse, replicating the prior studies. They were not simply overconfident, however: They said they had done worse than their peers more often than they said they had done better, although not significantly. This finding is consistent with the hypothesis that retrieval failures and hindsight bias made participants believe they had underperformed. Previous studies have shown that if people think they are good at a task, they rate themselves as above average (even if everyone else is also good at the task) but if they think they have not performed well, they rate themselves as below average (Kruger 1999).

General discussion

After responding to a set of trivia questions, participants in four experiments were asked how they would do if they were asked a new set of questions of similar difficulty. In all studies, the majority of participants said they would do about the same, but more participants said they would do better than said they would do worse (although they did not actually do significantly better). The combined data are presented in Fig. 2. When participants were given feedback this effect was strong. When they were not given feedback it was weak and not significant. The pattern of overconfidence was not all encompassing; participants did not tend to report that they had done better than their peers.

The findings seem to suggest that metacognitive judgments were influenced by participants' beliefs. The belief in question is that in human memory, retrieval failures are aberrations. The findings are consistent with the sequence of events depicted in Fig. 1: when participants could not think of an answer and then received feedback, they experienced hindsight bias, decided that they had suffered from a retrieval failure, and therefore predicted they would do better the next time because they believed that their retrieval failures were aberrations.

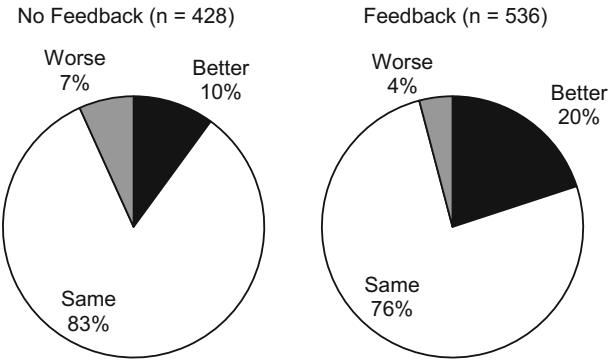

Fig. 2 The percentage of participants who said they thought they would do better, the same, or worse if they were asked another, similar set of trivia questions, summed across Experiments 1–4

The present studies cannot prove conclusively that participants' beliefs caused them to expect to do better, but three lines of evidence support this causal connection. First, if participants had been blindly optimistic, the presence or absence of feedback would not necessarily have affected their predictions. Feedback did affect predictions, however, probably because feedback caused people to believe they knew answers all along—that is, to show a hindsight bias—even if they did not. In other words, feedback caused participants to see themselves as having suffered from retrieval failures and it also increased their expectation that they would do better. Second, participants were not overconfident with respect to their peers, which is consistent with the possibility that retrieval failures caused some participants to believe that their performance on the first set of questions underestimated their true ability. Third, alternate explanations of the findings, having to do with participants' beliefs about luck and fate, were not supported.

The finding that beliefs affect judgments is noteworthy in the context of the literature on metacognition. Previous studies have demonstrated a disconnect between metacognitive beliefs and metacognitive judgments, whereby beliefs that should affect judgments fail to do so (e.g., Koriat et al. 2004; Kornell et al. 2011). The present results add to a small recent literature demonstrating situations in which beliefs do affect judgments (e.g., Mueller et al. 2013, 2014). The present results are especially novel because beliefs affected judgments even though a) the questions posed in the study did not make reference to, or obviously prompt, the relevant belief, and b) the belief was incorrect, meaning that it actually decreased metacognitive accuracy, and moreover there was a compelling reason for participants to ignore their beliefs and say they would do the same (as described next).

Irrational overconfidence

Previous studies have shown that people are overconfident about their memories in some situations (for reviews see Bjork 1999; Metcalfe 1998) but not overconfident, or even underconfident, in other situations (e.g., Finn and Metcalfe 2008; Koriat et al. 2006; Koriat et al. 2002; Kornell and Bjork 2009). Research with trivia questions has shown that people tend to be overconfident about specific responses they have given, if they are asked for the confidence ratings before receiving feedback (Fischhoff et al. 1977; Koriat et al. 1980). The present studies represent a departure from past research in that participants were overconfident when making judgments about their performance on new, future questions, not about specific

questions they had already answered. Also, their overconfidence increased, instead of decreasing, when they were given feedback about their answers.

The findings are also a departure in the sense that it is not necessarily irrational to be highly confident in one's memory, but it is, arguably, irrational to be confident that one will do be better at remembering next time when the task has not become any easier. It is well established that people can be irrationally optimistic when making other predictions, including predictions about their future grades or starting salaries, or their risk of breast cancer, heart attack, severe alcohol problems, sexually transmitted disease, getting fired, or getting sued, among other things (Shepperd et al. 2013). The present studies seem to add recalling information from memory to the list of situations in which people are susceptible to irrational overconfidence.

The fact that the task was not going to get any easier probably made it difficult for some participants to justify saying they would do better on the next set of questions. If some (or many) participants who said they would do the same actually believed they would do better, then the data presented in this article might underestimate the number of participants who believed they would do better. The fact that many participants said they would do better, despite it being a response that is difficult to justify, is a testament to the strength of their conviction that they would improve.

"I am not good at multiple choice tests"

The present results suggest an explanation for a common phenomenon in school: Students often believe that a test they have taken underestimated their true ability. If a student suffers from retrieval failures on a test, and then suffers from hindsight bias when the test is handed back, she might react the same way participants in the present study did: She might believe that her test score is an aberration—and an underestimate of her true ability—because of the belief that retrieval failures are aberrations. (Students can often be heard saying that they did poorly because of "stupid mistakes," which usually means getting an answer wrong that in hindsight they know how to answer correctly.)

Students often attribute their ostensible underperformance to the test itself—that is, a student might decide he is not good at taking certain types of tests (e.g., multiple choice tests) or that he is not good at taking tests in general. In other words, the reason why students often believe they are not good at taking certain kinds of tests may be that they have a history of taking such tests and then concluding that their test score does not reflect their true ability. This belief may be true in some cases, especially for students who suffer from test anxiety (Cassady and Johnson 2002), but for many students it is probably an illusion.

Conclusion

In four experiments, participants expected to do better on an upcoming memory test even though the test would be just as difficult as the one they had just completed—if they were given feedback on the one they had just completed. Of course, they did not actually do better. The cause their misplaced optimism seems to be the false belief that their memories are not subject to retrieval failures. These findings suggest that people do not fully appreciate the difference between the availability of a memory (i.e., whether it is encoded) and the accessibility of a memory (i.e., whether they can access it at the moment; Tulving and Pearlstone 1966). These findings identify a situation in which people seem to be sensitive to their metacognitive beliefs. They also suggest that the answer to a puzzling question—why do people so frequently believe that their test scores underestimate their true knowledge—may lie in their belief that the retrieval failures they experienced on the test were aberrations and not the normal functioning of memory.

Author note This research was supported by a Scholar Award from the James S. McDonnell foundation.

References

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., & Yurak, T. J. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*(5), 804–825. doi:10.1037/0022-3514.68.5.804.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Mahwah: Erlbaum.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55–68.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition, 41*(6), 897–903. doi:10.3758/s13421-013-0307-8.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Hillsdale: Erlbaum.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444. doi:10.1146/annurev-psych-113011-143823.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*(2), 270–295. doi:10.1006/ceps.2001.1094.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130*(5), 813–838. doi:10.1037/0033-2909.130.5.813.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*(1), 19–34. doi:10.1016/j.jml.2007.03.006.

Fischhoff, B. (1975). Hindsight is not equal to foresight: the effects of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: the appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552–564. doi:10.1037/0096-1523.3.4.552.

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology, 26*(2–3), 103–117. doi:10.1080/01973533.2004.9646399.

Guilmette, T. J., & Paglia, M. F. (2004). The public's misconception about traumatic brain injury: a follow up survey. *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists, 19*(2), 183–189. doi:10.1016/S0887-6177(03)00025-8.

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

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: biases and corrective procedures. *TIMS Studies in Management Science, 12*, 313–327.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609–639.

Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. doi:10.1037//0096-3445.126.4.349.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–118.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147–162. doi:10.1037//0096-3445.131.2.147.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*(4), 643–656. doi:10.1037/0096-3445.133.4.643.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 595–608. doi:10.1037/0278-7393.32.3.595.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219–224.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138*, 449–468. doi:10.1037/a0017350.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease of processing heuristic and the stability bias: dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22*(6), 787–794. doi:10.1177/0956797611407929.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*(2), 221–232. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10474208.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134.

Kruger, J., & Evans, M. (2004). If you don't want to be late, enumerate: unpacking reduces the planning fallacy. *Journal of Experimental Social Psychology, 40*(5), 586–598. doi:10.1016/j.jesp.2003.11.001.

Loftus, E. F. (1996). *Eyewitness testimony.* Harvard University Press.

Magnussen, S., Andersson, J., Cornoldi, C., De Beni, R., Endestad, T., Goodman, G. S., et al. (2006). What people believe about memory. *Memory, 14*(5), 595–613. doi:10.1080/09658210600646716.

Metcalfe, J. (1998). Cognitive optimism: self-deception or memory-based processing heuristics? *Personality and Social Psychology Review, 2*(2), 100–110. doi:10.1207/s15327957pspr0202_3.

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review, 20*(2), 378–384. doi:10.3758/s13423-012-0343-6.

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70*, 1–12. doi:10.1016/j.jml.2013.09.007.

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 and Verbal Behavior, 19*, 338–368.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615–625. doi:10.1037/a0013684.

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: effects on monitoring and control. *Psychonomic Bulletin & Review, 16*(3), 550–554. doi:10.3758/PBR.16.3.550.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science, 6*(5), 132–137. doi:10.1111/1467-8721.ep10772899.

Shepperd, J. A., Klein, W. M. P., Waters, E. A., & Weinstein, N. D. (2013). Taking stock of unrealistic optimism. *Perspectives on Psychological Science, 8*(4), 395–411. doi:10.1177/1745691613485247.

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: a representative survey of the U.S. population. *PloS One, 6*(8), e22757.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 381–391. doi:10.1016/S0022-5371(66)80048-8.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105–110. doi:10.1037/h0031322.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory, 20*(6), 568–579. doi:10.1080/09658211.2012.687052.