BRIEF REPORT

Ordered questions bias eyewitnesses and jurors

Robert B. Michael[1] · **Maryanne Garry**[1]

Abstract Eyewitnesses play an important role in the justice system. But suggestive questioning can distort eyewitness memory and confidence, and these distorted beliefs influence jurors (Loftus, *Learning & Memory, 12*, 361–366, 2005; Penrod & Culter, *Psychology, Public Policy, and Law, 1*, 817–845, 1995). Recent research, however, hints that suggestion is not necessary: Simply changing the order of a set of trivia questions altered people's beliefs about their accuracy on those questions (Weinstein & Roediger, *Memory & Cognition, 38*, 366–376, 2010, *Memory & Cognition, 40*, 727–735, 2012). We wondered to what degree eyewitnesses' beliefs—and in turn the jurors who evaluate them—would be affected by this simple change to the order in which they answer questions. Across six experiments, we show that the order of questions matters. Eyewitnesses reported higher accuracy and were more confident about their memory when questions seemed initially easy, than when they seemed initially difficult. Moreover, jurors' beliefs about eyewitnesses closely matched those of the eyewitnesses themselves. These findings have implications for eyewitness metacognition and for eyewitness questioning procedures.

Keywords Memory · Eyewitness · Juror · Question order

Memory plays a pivotal role in the justice system. But eyewitness memories are easily distorted (Loftus, 2005). Moreover, eyewitnesses who confidently report their distorted memories are persuasive to jurors (Cutler, Penrod, & Dexter, 1990; Douglass, Neuschatz, Imrich, & Wilkinson, 2010). In many cases, these distortions are the product of suggestive questioning (Loftus & Palmer, 1974; Loftus & Zanni, 1975). But what if questions cause trouble even without being suggestive? Could simply changing the order in which eyewitnesses answer questions affect how they appraise their memory, and in turn how jurors appraise those eyewitnesses? You might think changing the order in which eyewitnesses answer questions would have little effect—after all, the questions are still the same overall. But across the experiments presented here, we show that the order of questions matters.

In fact, we already know some seemingly trivial features of questions create problems for eyewitnesses. Take the striking results from changing just one word: In one study, witnesses reported cars in an accident traveled faster when a question suggested the cars *smashed* into rather than *hit* each other (Loftus & Palmer, 1974). In another study, witnesses were more likely to report seeing a nonexistent broken headlight when a question suggested its presence using the word *the*, rather than the more ambiguous *a* (Loftus & Zanni, 1975). More than three decades of research now shows that questions can transmit misleading suggestions that distort memory (see Loftus, 2005, for a review).

But questions can distort more than the details of memories; they can exert equally interesting influences on metacognition. For instance, eyewitnesses incorrectly answer misleading questions quickly and confidently (Loftus, Donders, Hoffman, & Schooler, 1989), and people generally provide more information—but monitor less for accuracy—when forced to answer questions compared with when they decide themselves what to report (Koriat & Goldsmith, 1996). These studies show that questions can change not only the content of eyewitnesses' memories—but also what eyewitnesses think about their memory.

✉ Maryanne Garry
  Maryanne.Garry@vuw.ac.nz

[1] School of Psychology, Victoria University of Wellington,
  PO Box 600, Wellington, New Zealand 6140

Springer

Research in educational psychology has revealed another important property of questions: The order in which they are asked. Across a series of experiments, people who answered trivia questions from the easiest to most difficult believed they answered more questions correctly than people who answered questions the other way around, even though everyone actually got about the same number correct (Jackson & Greene, 2014; Weinstein & Roediger, 2010, 2012).

We were intrigued by these findings and wondered to what extent the order of questions would influence eyewitnesses' beliefs about the accuracy and quality of their memory. But questions asked of eyewitnesses are different from trivia questions. Eyewitness interviewers use questions to gather information, often not knowing whether answers are correct. In that context—where accuracy is rarely known—the eyewitness's subjective experience becomes especially interesting. We were therefore particularly interested to examine what happens when questions are arranged to produce subjective experiences of increasing or decreasing eyewitness confidence.

Of course, it is not obvious that the order of questions should influence eyewitnesses at all. Whereas trivia questions can be drawn from a virtually infinite pool, questions put to eyewitnesses are typically from a more limited set, addressing a specific and recent event. This relative constraint should provide fewer opportunities for uncertainty, reducing eyewitnesses' reliance on heuristic processing—cognitive shortcuts that can result in biased judgments from seemingly innocuous manipulations (Tversky & Kahneman, 1974). It would be surprising and worrying if a simple change to the order of questions put to eyewitnesses could change how they appraise their memories.

But across six experiments, we show that the order of questions matters. In Experiments 1a, 1b, and 1c, we asked people to watch a video of a crime, then take an eyewitness memory test. We arranged test questions to produce one of two experiences, based on previously normed confidence ratings. In one version, the test began with the question that elicited the greatest confidence and ended with the question that elicited the least, so that people became progressively less confident. In the other version, we reversed this ordering. These different experiences—produced by an identical set of questions—affected how many questions eyewitnesses thought they answered correctly, and how confident they were about the accuracy of their memory. In Experiments 2a, 2b, and 2c, we show that these biases have consequences beyond the eyewitnesses: Jurors' impressions of eyewitnesses matched the eyewitnesses' own impressions, a finding in line with research showing that jurors find confident eyewitnesses to be credible eyewitnesses (Penrod & Cutler, 1995).

Experiment 1a

Method

Subjects Through pilot work, we determined a sample size of 100 (50 per between subjects cell). We ultimately recruited a total of 102 Amazon Mechanical Turk workers (www.mturk.com), because Mechanical Turk and Qualtrics—our experimental software—interact such that it is possible to unintentionally collect more data points than requested.

Design We used a simple two groups design with Question Order (low-to high confidence, high-to-low confidence) manipulated between subjects.

Procedure The experiment had four phases. First, we told subjects the study was examining learning styles. They watched one of two similar videos of a tradesman who stole items from the unoccupied house he was working in (Takarangi, Parker, & Garry, 2006). We counterbalanced versions across subjects and conditions.

The second phase began when the video ended. To mirror real-life memory decay, subjects solved Sudoku puzzles for 10 minutes.

In the third phase, subjects took a surprise memory test consisting of 30 two-alternative forced choice (2AFC) questions about the video. As we noted earlier, we were particularly interested in examining what happens when questions are arranged to produce experiences of increasing or decreasing eyewitness confidence. Accordingly, we ordered the test questions by confidence rather than accuracy, because accuracy is typically unknown in an eyewitness context.[1]

We constructed the order of test items using data from an earlier, separate group of 107 subjects who followed the same procedure, except the 30 questions were ordered randomly. Then, on the basis of mean confidence, we ordered the 30 questions from the lowest confidence ($M = 1.73$, $SD = 1.06$) to highest ($M = 4.79$, $SD = 0.63$) to produce the low-to-high confidence test. We reversed this order to create the high-to-low confidence version. Subjects in the current experiment were randomly assigned one of these versions.

---

[1] To consider the possibility that the subjective experience of confidence tracked subjective difficulty, we asked a group of 141 people to rate the difficulty of each randomly ordered question on a scale from 1 ("Very easy") to 5 ("Very difficult"). When we compared items in this group with items in the group of 107 subjects, we found that reported difficulty closely matched reported confidence ($r = -.82$, 95 % CI [−.66, −.91], $p < .001$, treating items as cases), suggesting that confidence was a good proxy for subjective difficulty.

For each question, subjects used a scale from 1 ("Not at all confident") to 5 ("Very confident") to report their confidence they had selected the correct answer.

The fourth phase followed the test. Subjects answered two randomly ordered questions: [1] "The memory test about Eric the Electrician consisted of 30 questions. How many of those questions do you think you answered correctly?" Subjects responded with a number between 0 and 30; [2] "Suppose that you were asked to testify as an eyewitness. How confident would you be in your memory of the events you saw in the video of Eric the Electrician?" Subjects responded on a scale from 1 ("Not at all confident") to 5 ("Very confident").

Results and discussion

We first performed a manipulation check by examining mean confidence ratings for individual test questions. These data appear in the top panel of Fig. 1 and show that our manipulation worked: "low-to-high" subjects were increasingly confident, and "high-to-low" subjects were the opposite. The middle panel of Fig. 1 displays accuracy for individual test questions and shows a similar pattern—although less cleanly, as a consequence of 2AFC scoring. The bottom panel of Fig. 1 displays confidence-accuracy relationships for individual test questions and suggests that the order of questions did not affect subjects' insight into their own accuracy. We also found that the order of questions had little effect on overall test performance, $M_{diff} = 0.65$ (2.17 %), 95 % confidence interval (CI) [−0.46, 1.76]; $t(100) = 1.15$, $p = .252$.

We now address our primary questions: To what extent did the order of questions [1] bias subjects' retrospective estimates of their test performance, and [2] affect their confidence in their memory? To answer [1], we subtracted subjects' test scores from their retrospective estimates to produce bias scores. Positive bias scores represent subjects who thought they performed better on the test than they truly did, and negative bias scores represents the opposite. We present actual and retrospective estimates of test scores in the top panel of Fig. 2 and bias scores in the middle panel. These data show that low-to-high confidence subjects were more pessimistic than high-to-low confidence subjects, $M_{diff} = 2.32$ (7.73 %), 95 % CI [0.33, 4.32]; $t(100) = 2.31$, $p = .023$. To answer [2], we examined subjects' post-test reports of memory confidence. These data appear in the bottom panel of Fig. 2 and show that low-to-high confidence subjects were less confident about the accuracy of their memory: $M_{diff} = 0.46$ (11.50 %), 95 % CI [0.08, 0.84]; $t(100) = 2.41$, $p = .018$ (for all experiments, we report cell means and SDs in Tables 1 and 2).

To determine the extent to which these effects would generalize to the more real-world situation of open-ended questions, we conducted Experiment 1b.

Experiment 1b

Method

Subjects To boost precision, we recruited a larger sample of 220 Mechanical Turk workers.

Design and procedure Experiment 1b followed the design and procedure of Experiment 1a, except we converted each 2AFC question into a cued-recall question.

Results and discussion

We scored responses by a keyword search. A blind rater also hand-scored a random 20 % of responses; electronic and hand scores were highly correlated, $r = 0.96$, $p < .001$.

This new format replicated the earlier results: low-to-high confidence subjects were more pessimistic, $M_{diff} = 3.65$ (12.17 %), 95 % CI [2.33, 4.98]; $t(218) = 5.43$, $p < .001$; and were less confident about the accuracy of their memory, $M_{diff} = 0.37$ (9.25 %), 95 % CI [0.10, 0.64]; $t(218) = 2.73$, $p = .007$. We next ran Experiment 1c to ensure these effects were not tied to specific materials.

Experiment 1c

Method

Subjects We recruited a new sample of 205 Mechanical Turk workers.

Design and procedure The design and procedure was the same as Experiment 1a, except subjects viewed a different video and answered a different set of twenty 2AFC questions (French, Garry, & Mori, 2011). We again arranged questions in the two orders, based on data from an earlier 106 subjects who rated each randomly ordered question for its difficulty.

Results and discussion

As before, we found that low-to-high confidence subjects were more pessimistic, $M_{diff} = 1.88$ (9.40 %), 95 % CI [0.86, 2.90]; $t(203) = 3.62$, $p < .001$; they also were less confident about the accuracy of their memory, $M_{diff} = 0.28$ (7.00 %), 95 % CI [0.01, 0.55]; $t(203) = 2.05$, $p = .042$. These data show that the influence of the order of questions generalizes to novel materials.

In line with Cumming's (2012) recommendations, we obtained more precise estimates of these effect sizes by meta-analysing the results of Experiments 1a, 1b, and 1c, using ESCI software to run two random effects model meta-analyses. These analyses estimate that "low-to-high" eyewitnesses
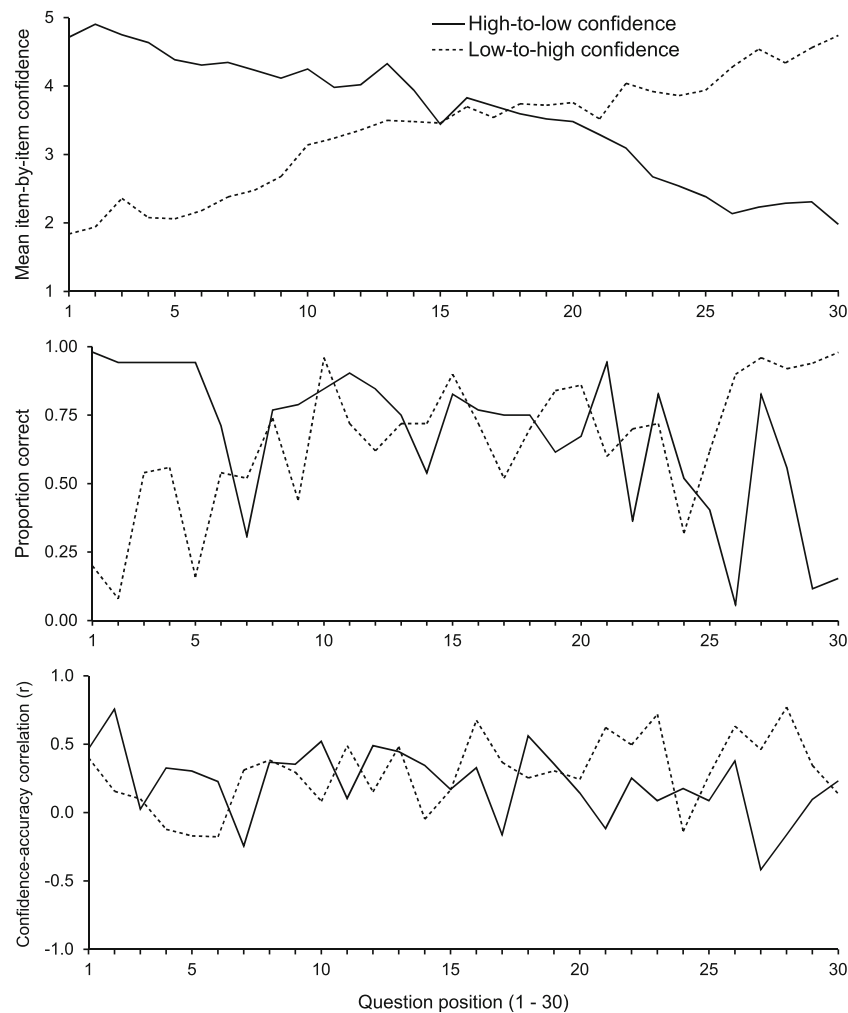
Fig. 1 *Top panel*: Mean confidence of a correct answer for each test question, ordered by position on test. *Middle panel*: Proportion of subjects who answered each test question correctly, ordered by position on test. *Bottom panel*: Pearson correlations between confidence and accuracy ratings for each test question, ordered by position on test. Note that the test versions are symmetric, i.e., question 1 in one condition is the same as question 30 in the other condition. Data are from Experiment 1a

would be 10.33 % more pessimistic about their performance than "high-to-low" eyewitnesses, M_{diff} = 10.33 %, 95 % CI [7.36, 13.30], z = 6.82, p < .001. These "low-to-high" eyewitnesses also would be 0.36 points, or 9.00 %, less confident about what they remember, M_{diff} = 0.36, 95 % CI [0.19, 0.52], z = 4.10, p < .001.

The results of Experiments 1a, 1b, and 1c show that the order of questions shapes what eyewitnesses believe. Specifically, when people answered questions that initially seemed difficult and then became easy, they were more pessimistic and less confident about their memory compared with others who answered questions that initially seemed easy and then became difficult.

In changing how eyewitnesses appraise their memories, one possible consequence is that jurors will appraise the eyewitness's credibility in the same direction (Douglass

et al., 2010). Such a result would have disturbing implications for the justice system. Because jurors tend to rely on eyewitness confidence as a signal of accuracy (Penrod & Cutler, 1995), we asked subjects in Experiments 2a, 2b, and 2c to take on the role of a juror, evaluating an eyewitness whose confidence systematically changed over the course of questioning.

Experiment 2a

Method

Subjects We aimed to collect data from 200 people but ultimately recruited 261 Mechanical Turk workers.
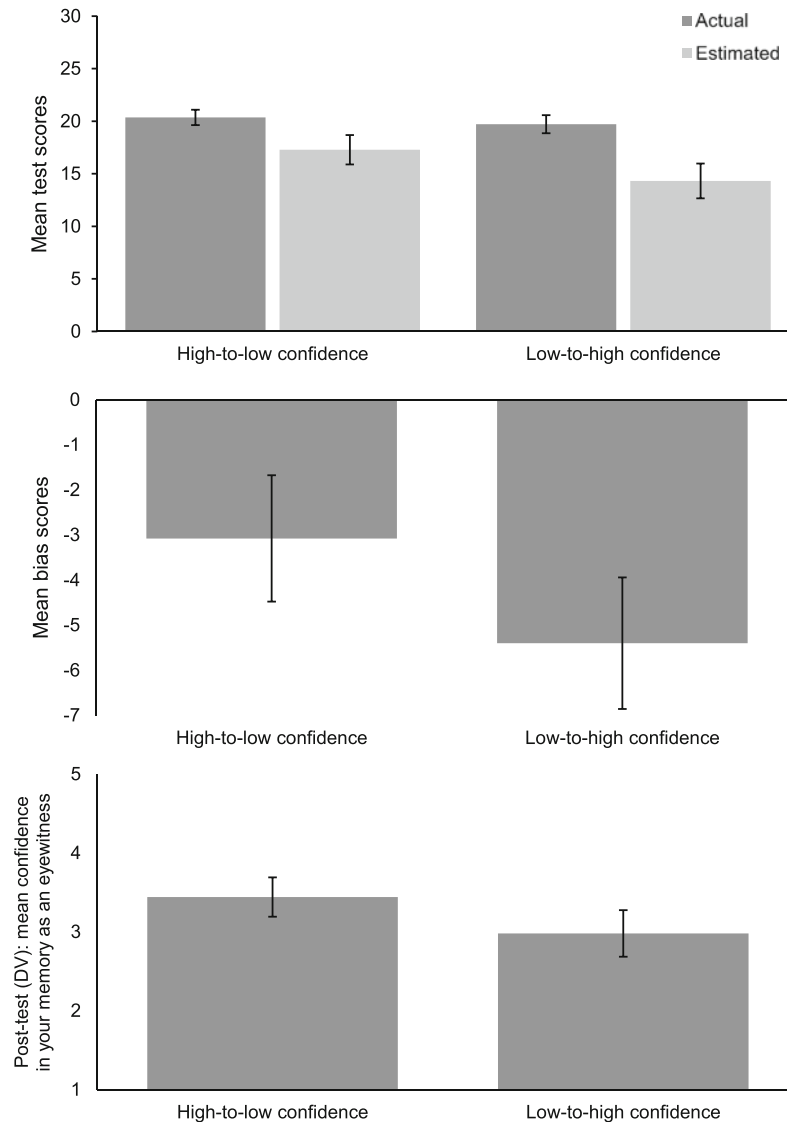
QUESTION ORDER BIASES EYEWITNESSES AND JURORS

Fig. 2 *Top panel*: Mean actual and estimated test scores by condition. *Middle panel*: Mean bias (estimated test score - actual test score) by condition. Positive bias scores represent subjects who thought they performed better than they truly did; negative bias scores represent the opposite. *Bottom panel*: Mean post-test memory confidence by condition. Error bars represent 95 % confidence intervals of cell means. Data are from Experiment 1a

Table 1 Experiment 1 mean scores for Bias and Confidence by condition

	Bias				Confidence			
	Condition				Condition			
	High-to-low	Low-to-high	M_{Diff}	95 % CI	High-to-low	Low-to-high	M_{Diff}	95 % CI
Experiment 1a ($N = 102$)	-3.08 (5.03)	-5.40 (5.13)	2.32	0.33, 4.32	3.44 (0.89)	2.98 (1.04)	0.46	0.08, 0.84
Experiment 1b ($N = 220$)	1.64 (5.49)	-2.02 (4.43)	3.65	2.33, 4.98	2.69 (1.08)	2.32 (0.94)	0.37	0.10, 0.64
Experiment 1c ($N = 205$)	-3.25 (4.00)	-5.13 (3.43)	1.88	0.86, 2.90	2.99 (1.01)	2.71 (0.97)	0.28	0.01, 0.55
Meta-analysis			10.33 %	7.36, 13.30			0.36	0.19, 0.52

Standard deviations in parentheses. Note: The meta-analysed bias difference is a percentage, because the number of test questions differed in Experiment 1c

Table 2 Experiment 2 mean scores for Estimate and Confidence by condition

	Estimate				Confidence			
	Condition				Condition			
	High-to-low	Low-to-high	M_{Diff}	95 % CI	High-to-low	Low-to-high	M_{Diff}	95 % CI
Experiment 2a ($N = 261$)	17.47 (4.81)	14.23 (4.29)	3.23	2.12, 4.34	3.15 (0.84)	2.67 (0.84)	0.47	0.27, 0.68
Experiment 2b ($N = 305$)	17.79 (4.52)	13.62 (5.13)	4.18	3.09, 5.27	3.06 (0.85)	2.34 (0.92)	0.72	0.52, 0.92
Experiment 2c ($N = 316$)	12.33 (3.30)	10.46 (3.38)	1.88	1.14, 2.62	3.06 (0.91)	2.64 (0.89)	0.42	0.22, 0.62
Meta-analysis			11.38 %	8.77, 14.00			0.54	0.36, 0.72

Standard deviations in parentheses. Note: The meta-analysed estimate difference is a percentage, because the number of test questions differed in Experiment 2c

Design We used a two groups design with Question Order (low-to-high confidence, high-to-low confidence) manipulated between subjects.

Procedure We asked people to take on the role of a juror and answer questions about an eyewitness who had been in a previous study. We told these "jurors" that in the previous study, the eyewitness had taken a memory test after watching the video of Eric the Electrician. The juror's task was not to watch the video but to carefully read the eyewitness's memory test and then answer some questions.

To mirror the real-world scenario where a group of jurors evaluate one eyewitness, all jurors within a group actually read a single eyewitness's test that we secretly created. In the high-to-low confidence version, the eyewitness's answers were initially confident but became less confident over the test. In the low-to-high confidence version, this pattern reversed. We created these two versions using data from Experiment 1a. We calculated mean confidence ratings for each of the 30 questions, rounding each mean to an integer so it could be represented on the Likert scale of confidence the eyewitness had ostensibly used. We also randomly selected, for each test question, which answer the eyewitness had ostensibly chosen.

Subjects randomly received either the low-to-high confidence or high-to-low confidence eyewitness test, formatted exactly like the test in Experiment 1a. Immediately after reading, subjects answered two randomly ordered questions: [1] "The memory test about Eric the Electrician consisted of 30 questions. How many of those questions do you think the eyewitness answered correctly?" Subjects responded with a number between 0 and 30; [2] "How confident are you about the accuracy of the eyewitness's memory?" Subjects responded on a scale from 1 ("Not at all confident") to 5 ("Very confident").

Results and discussion

Jurors believed that an initially confident eyewitness was more accurate, estimating that these eyewitnesses answered more questions correctly, $M_{diff} = 3.23$ (10.77 %), 95 % CI [2.12, 4.34]; $t(259) = 5.73$, $p < .001$. Jurors also reported more confidence in these eyewitnesses' memories, $M_{diff} = 0.47$ (11.75 %), 95 % CI [0.27, 0.68]; $t(259) = 4.54$, $p < .001$.

Note, however, that each of the 30 test questions always appeared with the same confidence rating. This confound leaves open the possibility that jurors were influenced not by the eyewitness's confidence, but by the content of the questions. We ran Experiment 2b to address this counter explanation.

Experiment 2b

Method

Subjects We aimed to boost precision by increasing observations to 150 per between subjects cell, ultimately recruiting 305 Mechanical Turk workers.

Design and procedure The design and procedure was the same as in Experiment 2a, except that we decoupled questions from their associated confidence ratings while maintaining the ascending or descending pattern of confidence, by randomly assigning questions to each confidence rating.

Results and discussion

We found again that subjects believed high-to-low confidence eyewitnesses answered more questions correctly, $M_{diff} = 4.18$ (13.93 %), 95 % CI [3.09, 5.27]; $t(303) = 7.54$, $p < .001$, and were more confident about the accuracy of these eywitnesses' memories, $M_{diff} = 0.72$ (18.00 %), 95 % CI [0.52, 0.92]; $t(303) = 7.10$, $p < .001$.

Finally, we ran Experiment 2c to demonstrate that these effects were not tied to specific materials.

Experiment 2c

Method

Subjects We aimed to collect 150 observations per between subjects cell, and ultimately recruited 316 Mechanical Turk workers.

Design and procedure The design and procedure was the same as in Experiment 2b but used the materials from Experiment 1c.

Results and discussion

We found again that jurors believed high-to-low confidence eyewitnesses answered more questions correctly, $M_{diff} = 1.88$ (9.40 %), 95 % CI [1.14, 2.62]; $t(314) = 4.99$, $p < .001$, and jurors were also more confident about the accuracy of these eyewitnesses' memories, $M_{diff} = 0.42$ (10.50 %), 95 % CI [0.22, 0.62]; $t(314) = 4.14$, $p < .001$.

The findings from Experiments 2a, 2b, and 2c fit with those of Experiments 1a, 1b, and 1c, in which eyewitnesses thought they answered more questions correctly and reported higher confidence in their memory if their initial experience was one of high confidence. We meta-analysed the results of Experiments 2a, 2b, and 2c (Cumming, 2012) and estimated that jurors believe "high-to-low" eyewitnesses answer 11.38 % more questions correctly, $M_{diff} = 11.38$ %, 95 % CI [8.77, 14.00], $z = 8.53$, $p < .001$. Moreover, jurors are 0.54 points—or 13.50 %—more confident about the accuracy of a "high-to-low" eyewitness's memory, $M_{diff} = 0.54$, 95 % CI [0.36, 0.72], $z = 5.75$, $p < .001$.

General discussion

Across six experiments, we found that the order in which eyewitnesses answered questions mattered in two key ways. First, the order changed how eyewitnesses appraised themselves. When questions produced an initial experience of high confidence rather than low confidence, eyewitnesses believed that they were more accurate and were more confident about their memory. Second, the order changed how jurors appraised eyewitnesses. Jurors believed eyewitnesses who initially displayed high confidence were more accurate, and jurors were more confident about those eyewitnesses' memories. This collection of results paints a worrying picture of the malleability of beliefs about memory accuracy.

It is surprising that questions produce different beliefs in witnesses when all that changes is the order those questions are asked. Ultimately, everyone answers the same questions, so it seems reasonable to expect no differences in beliefs. But the influence of order shows that beliefs about memory are shaped not only by the content or phrasing of questions, but also by factors that—on the face of it—are trivial.

In fact, our seemingly trivial manipulation produced effects similar in size to more blatant manipulations affecting eyewitness credibility. An eyewitness who claims to be absolutely certain, for example, is rated more credible than an eyewitness who does not (Tenney, MacCoun, Spellman, & Hastie, 2007), and prosecution eyewitnesses who elaborate their testimony with extra details are more credible, and get more guilty verdicts, than eyewitnesses who do not (Bell & Loftus, 1988, 1989). It is worrying that our subtle manipulation produces effects similar in magnitude to these relatively heavy-handed approaches.

How can we explain our effects? One possibility is that people's attention wanes over the test, resulting in impressions influenced most by early experience (Crano, 1977). If this "attention decrement" hypothesis is true, then the same question should be answered with higher accuracy when it appears early rather than late. To address this possibility, we ran a random effects model meta-analysis comprising all three datasets from Experiment 1. This meta-analysis compared accuracy between groups for the subjectively easiest and most difficult test questions, because each appears first for one group and last for the other. We found no support for this attention-based explanation: Accuracy is not notably different when a question appears first rather than last, $M_{diff} = -0.01$, 95 % CI [−0.04, 0.02], $z = -0.40$, $p = .686$.

An alternative explanation is that the effects are driven by early experience and insufficient adjustments: The subjective ease or difficulty of early questions sets an anchor, and to save effort, people adjust from this anchor only until reaching a plausible impression (Epley & Gilovich, 2006). This explanation is consistent with recent research in which subjects held biased impressions of performance throughout a trivia test, and not merely at the end (Weinstein & Roediger, 2010, 2012). Relatedly, Experiments 2a-2c suggest that jurors used early information to create a story about the eyewitness's credibility and were slow to revise that story in the face of new information. This explanation fits with the Story Model of juror decision-making, a model in which juror's verdicts are influenced by the stories they construct to make sense of events (Pennington & Hastie, 1992).

Our findings have implications for eyewitnesses' metacognition, because they suggest that the order of questions influences eyewitnesses' ability to evaluate what they know about an event. Similarly, our findings are reminiscent of other suggestive techniques that manipulate eyewitness beliefs, such as subtle changes to the wording of questions, or direct feedback

about lineup identifications (Douglass & Steblay, 2006; Loftus & Palmer, 1974; Loftus & Zanni, 1975). But in contrast, we have manipulated what eyewitnesses and jurors believe about memory without using suggestive techniques.

Our findings also raise interesting questions. For instance, does the order of questions influence other related judgments, such as eyewitnesses' estimates of how well they saw the perpetrator? We know that positive post-identification feedback enhances eyewitnesses' beliefs about their memory for a crime, including how well they could see a suspect's face and how much attention they paid (Wells & Bradfield, 1998). Perhaps an initial experience of subjectively easy questions causes similar enhancements. It would also be useful to know if the order of questions produces lasting changes in beliefs or if the influence is fleeting. Finally, it is worth considering that we ordered questions in our experiments either by subjective confidence or subjective difficulty. Earlier work has ordered questions by objective difficulty, calculated as the mean proportion of people who answer a question correctly (Jackson & Greene, 2014; Weinstein & Roediger, 2010, 2012). Our results suggest that the subjective experience of difficulty may underpin the influence of question order—but a future experiment teasing apart subjective and objective difficulty could provide information about their relative contributions.

Eyewitnesses play an undeniably important role in the justice system. But justice requires that we protect the integrity of eyewitness memory as much as possible. That integrity is called into question when eyewitnesses and jurors are swayed by something as trivial as the order in which they answer questions.

References

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology, 18,* 1171–1192. doi:10.1111/j.1559-1816.1988.tb01200.x

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology, 56,* 669–679. doi:10.1037/0022-3514.56.5.669

Crano, W. D. (1977). Primacy versus recency in retention of information and opinion change. *The Journal of Social Psychology, 101,* 87–96. doi:10.1080/00224545.1977.9923987

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14,* 185–191. doi:10.1007/BF01062972

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Douglass, A. B., & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20,* 859–869. doi:10.1002/acp.1237

Douglass, A. B., Neuschatz, J. S., Imrich, J., & Wilkinson, M. (2010). Does post-identification feedback affect evaluations of eyewitness testimony and identification procedures? *Law and Human Behavior, 34,* 282–294. doi:10.1007/s10979-009-9189-5

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17,* 311–318. doi:10.1111/j.1467-9280.2006.01704.x

French, L., Garry, M., & Mori, K. (2011). Relative—not absolute—judgments of credibility affect susceptibility to misinformation conveyed during discussion. *Acta Psychologica, 136,* 119–128. doi:10.1016/j.actpsy.2010.10.009

Jackson, A., & Greene, R. L. (2014). Impression formation of tests: retrospective judgments of performance are higher when easier questions come first. *Memory & Cognition, 42,* 1325–1332. doi:10.3758/s13421-014-0439-5

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological review, 103,* 490–517. doi:10.1037/0033-295X.103.3.490

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12,* 361–366. doi:10.1101/lm.94705

Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition, 17,* 607–616. doi:10.3758/BF03197083

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior, 13,* 585–589. doi:10.1016/S0022-5371(74)80011-3

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society, 5,* 86–88. doi:10.1016/0010-0285(75)90023-7

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of personality and social psychology, 62,* 189–206.

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1,* 817–845. doi:10.1037/1076-8971.1.4.817

Takarangi, M. K., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology, 20,* 583–590. doi:10.1002/acp.1209

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18,* 46–50. doi:10.1111/j.1467-9280.2007.01847.x

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. doi:10.1126/science.185.4157.1124

Weinstein, Y. L., & Roediger, H. L. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition, 38,* 366–376. doi:10.3758/MC.38.3.366

Weinstein, Y. L., & Roediger, H. L. (2012). The effect of question order on evaluations of test performance: How does the bias evolve? *Memory & Cognition, 40,* 727–735. doi:10.3758/s13421-012-0187-3

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83,* 360–376. doi:10.1037/0021-9010.83.3.360