

The effect of cross-examination style questions on adult eyewitness accuracy depends on question type and eyewitness confidence

Kimberley A. Wade & Emily R. Spearing

To cite this article: Kimberley A. Wade & Emily R. Spearing (2022): The effect of cross-examination style questions on adult eyewitness accuracy depends on question type and eyewitness confidence, *Memory*, DOI: [10.1080/09658211.2022.2129066](https://doi.org/10.1080/09658211.2022.2129066)

To link to this article: <https://doi.org/10.1080/09658211.2022.2129066>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 02 Oct 2022.



[Submit your article to this journal](#)



Article views: 18



[View related articles](#)



[View Crossmark data](#)

The effect of cross-examination style questions on adult eyewitness accuracy depends on question type and eyewitness confidence

Kimberley A. Wade  and Emily R. Spearing 

Department of Psychology, University of Warwick, Coventry, United Kingdom

ABSTRACT

In adversarial legal systems across the world, witnesses in criminal trials are subjected to cross-examination. The questions that cross-examiners pose to witnesses are often complex and confusing; they might include negatives, double negatives, leading questions, closed questions, either/or questions, or complex syntax and vocabulary. Few psycholegal studies have explored the impact of such questions on the accuracy of adult witnesses' reports. In two experiments, we adapted the standard investigative interview procedure to examine the effect of five types of cross-examination style questions on witness accuracy and confidence. Participants watched a mock crime video and answered simple-style questions about the event. Following a delay, participants answered both cross-examination style questions and simple questions about the event. Negative and Double negative questions sometimes impaired the accuracy of witnesses' responses during cross-examination, whereas Leading and Leading-with-feedback questions did not impair – but sometimes enhanced – the accuracy of witnesses' responses. Participants who were better at discriminating between correct and incorrect responses on the initial memory test were more likely to improve the accuracy of their reports during cross-examination. Our findings suggest that the effect of cross-examination style questions on eyewitness accuracy depends on question type and witnesses' confidence in their responses.

ARTICLE HISTORY

Received 5 April 2022
Accepted 21 September 2022

KEYWORDS

Cross-examination;
eyewitness memory;
confidence; accuracy;
leading questions

In adversarial legal systems across the world, witnesses in criminal trials are subjected to cross-examination. The eminent legal scholar, John Henry Wigmore, described cross-examination as “beyond any doubt the greatest legal engine ever invented for the discovery of truth.”¹ Some legal scholars continue to claim that cross-examination is one of the legal system's essential means for testing the evidence, accessing the truth, revealing lies and exposing unreliable witnesses (for discussion, see Hickey, 1993; Wheatcroft et al., 2004). Yet other scholars posit that cross-examination is not an effective tool for determining eyewitness accuracy and instead only serves to undermine a witness's credibility (Plotnikoff & Woolfson, 2009; Spencer, 2012). Although legal scholars have long debated the value and limits of cross-examination, few psycholegal studies have contributed to this debate. The aim of the current research was to build on existing findings to extend our understanding of how cross-examination affects adult eyewitness testimony.

In criminal trials, each party has the right to cross-examine the other side's witnesses. The key aims of cross-examination are threefold: to elicit evidence in support of one's case, to cast doubt on or to undermine the witness's evidence and credibility, or to challenge

disputed evidence (Laver, n.d.). A common cross-examination strategy is to use question formats that can hinder the accuracy, consistency, or completeness of the witness's responses, which can lead fact-finders to infer that the testimony is unreliable (Kapardis, 2010). The questions that cross-examiners pose to witnesses are often complex and confusing; they may include the use of negatives or double negatives, leading questions, closed questions, either/or questions, yes/no questions, multiple questions, and complex syntax or vocabulary (Kapardis, 2010; Kebbell et al., 2003, 2004). Indeed, texts often advocate the use of such techniques to discredit witnesses (Defence-Barrister, n.d.; Levy, 1991; Stone, 1988) and lawyers prefer these types of questions because they enable the lawyer to maintain control over the witness during cross-examination (Ellison, 2001). Furthermore, research shows that lawyers continue to use complex language when they cross-examine witnesses who have intellectual disabilities and tend to be more susceptible to suggestive influences than other witnesses (Kebbell et al., 2004). It is important to note, however, that lawyers' choice of complex language is not necessarily malevolent, and they may simply adopt a questioning style that they believe to be effective in all cases.

CONTACT Kimberley A. Wade  k.a.wade@warwick.ac.uk  Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Although hundreds of studies have examined investigative interviewing techniques, including the effects of suggestive questioning techniques on witness accuracy (for reviews see Gabbert & Hope, 2018; Hope & Gabbert, 2019), there are fundamental differences between investigative interviews and cross-examination (Valentine & Maras, 2011). In a typical eyewitness interview study, mock-witnesses view a target event then they may or may not be exposed to some misleading information before their memories for the target event are tested. In contrast, cross-examination occurs after a witness has viewed an event and already provided an initial memory report, and typically involves active persuasion to undermine the witness's credibility. During cross-examination, witnesses are likely to face direct challenges to their initial report (Valentine & Maras, 2011). For these reasons, the standard eyewitness interview paradigm is not a good analogue for investigating or understanding the impact of cross-examination on memory.

To date, only a handful of published studies have explored the effect of cross-examination on the reliability of adult witness testimony. In one recent study, participants watched a mock crime video then half of the participants were exposed to misinformation via a discussion with a fellow mock-witness participant (Valentine & Maras, 2011). In the same session participants were asked to recall everything they could remember about the target event and to answer some specific questions – this memory test served as participants' initial statement. Four weeks later, all participants were interviewed by a trainee barrister who was told that it would benefit the (fictitious) client's case if the participant was to change their testimony on four critical points. The barrister used the same techniques they would normally employ in court to elicit the desired responses. Under cross-examination, 73% of participants acquiesced to the barrister, changing their answer on at least one of the four critical points (14% changed three answers, but none changed all four). Overall, cross-examination led to a 26% reduction in statement accuracy.

Other studies have shown that cross-examination style questioning can substantially impair adults' ability to respond accurately, even when their memory would otherwise allow them to do so (Jack & Zajac, 2014; Kebbell & Johnson, 2000). In one study, participants watched a video of an assault before answering either simple questions (e.g., "Is it true that the woman went into the house?") or cross-examination style questions (e.g., "Is it not true that the woman did not go into the house?" Kebbell et al., 2010). Participants who responded to cross-examination questions were 41% less accurate than participants who answered simple questions. Likewise, another study found that the decline in accuracy between two interviews was greater when participants answered cross-examination style questions in the second interview than when they answered simple questions (Jack & Zajac, 2014). Taken together, these findings

converge to suggest that witnesses struggle to provide accurate reports when they are subjected to cross-examination questioning.

A slightly larger body of research has investigated the influence of cross-examination style questions on the accuracy of child witnesses and other vulnerable witnesses (e.g., people with intellectual disabilities, see Jack & Zajac, 2014; O'Neill & Zajac, 2013; Righarts et al., 2015; Zajac et al., 2003, 2009; Zajac & Hayne, 2003, 2006). In one study, for instance, 5–6 year old children engaged in activities at a local police station, such as having their photos and fingerprints taken (Zajac & Hayne, 2003). Half of the children were exposed to misleading information about their experiences. Six weeks later, the children underwent an interview in which they were asked to recall everything they could about the visit and to answer yes/no questions about their experience. Eight months after that, the children were cross-examined – they watched a video of their original interview and answered questions about the visit that were designed to persuade them to change their responses (e.g., "I don't think you really got your photo taken. I think someone told you to say that. That's what really happened, isn't it?"). The vast majority of children (85%) made changes to their original statements and were just as likely to alter originally correct responses as incorrect responses. Even when the children were not exposed to misinformation, cross-examination served to reduce the overall accuracy of their reports by approximately 20%. In follow-up research, Zajac and Hayne (2006) found that older children (9-10 year olds) were less likely to change accurate than inaccurate responses under cross-examination, but they still altered 40% of correct answers.

The small body of available research on cross-examination and witness memory shows that cross-examination style questions tend to reduce the accuracy of witness evidence compared to simpler forms of questioning. The question arises, however, of how different types of question formats (e.g., negatives, double-negatives, leading, etc) influence witness accuracy and whether certain types of questions are particularly detrimental. So far, the available evidence is mixed. In one experiment, participants provided fewer correct answers when responding to leading questions (e.g., "The young woman who answered the door had long hair, didn't she?") than when responding to simple questions (Wheatcroft & Woods, 2010). Conversely, another experiment has shown that questions containing negatives and double negatives (e.g., "Did the woman not have black hair?") produced fewer correct responses than did simple questions, while leading questions did not impair accuracy compared to simple questions (Kebbell & Giles, 2000).² These findings highlight the possibility that while some question types may impair witness accuracy during cross-examination, others may have little or no effect. The first aim of Experiment 1 was to systematically explore the effect of five commonly used cross-examination style questions on eyewitness accuracy.

Another important question is whether a witness's confidence in their memory of the target event affects how they respond under cross-examination. Common-sense, along with empirical research, suggests that witnesses who lack confidence in their memory will be more likely to alter their statement under cross-examination than witnesses who are more confident. Indeed, people who lack confidence in their memory are more likely to fall prey to suggestive, external influences. For example, studies have shown that people with relatively low levels of self-esteem are more likely to accept misinformation from an external source than those with higher levels of self-esteem (Thorley & Kumar, 2017), and people who mistrust their own memory are more likely to accept misinformation than those who feel optimistic about their memory ability (van Bergen et al., 2010). Furthermore, witnesses who receive negative feedback about their memory performance subsequently provide lower confidence judgements and are more likely to alter their memory reports compared to those who receive neutral feedback (Henkel, 2017). Based on these findings, we might expect that witnesses will be more likely to fall prey to the persuasive influence of cross-examination style questions, and therefore change their responses, when they are not confident in those initial responses.

New research on the confidence-accuracy relationship in witness recall leads to a further prediction about the effect of cross-examination on witness behaviour: Cross-examination may enhance memory performance for some witnesses, but not others. We know that the relationship between how confident a witness feels about their memory and the accuracy of that memory is reasonably strong in basic recall tasks, particularly when participants (mock witnesses) are questioned appropriately, without suggestion (Paulo et al., 2019; Roberts & Higham, 2002; Sauer & Hope, 2016; Spearing & Wade, 2021; Spearing & Wade, 2022; Vredeveldt & Sauer, 2015; Wixted et al., 2018). If witnesses are good at discriminating between correct and incorrect responses when they initially recall an event, then subsequently, when under cross-examination, witnesses should be more likely to change responses that they initially answered incorrectly than responses that they initially answered correctly, thus increasing their overall accuracy. By contrast, witnesses who are poor at discriminating between correct and incorrect responses will be more likely to change their correct responses, thus decreasing their overall accuracy. The second aim of Experiment 1 was to test this prediction.

We adapted the standard investigative interview procedure to examine the effect of cross-examination style questions on eyewitness accuracy, and whether question type and witness confidence play a role. Participants watched a mock crime video and following a short filler task they answered simple-style questions about the event (e.g., "Were the lockers that the perpetrator tried to break into red?") and rated their confidence in each response. After a 3–5-day delay, participants answered a

second memory test consisting of both cross-examination style questions and simple questions, or a memory test consisting of only simple questions.

Experiment 1

Method

Transparency and openness

For both experiments, we report how we determined the sample sizes, all data exclusions, all manipulations, and all measures in the study. The pre-registrations, numeric data, and corresponding R code are available on the Open Science Framework (https://osf.io/tv9pg/?view_only=002448b94461447b9425ddb4f5e8be9) and materials can be accessed on request from the corresponding author. Data were analyzed using R, version 4.1.2 (R Core Team, 2022) and calibration statistics were calculated using the legalPsych package (Van Boeijen & Saraiva, 2018).

Participants & design

Participants were allocated to the cross-examination or control condition. In the cross-examination phase, cross-examination participants answered 5 types of question (*Simple, Negative, Double negative, Leading, Leading-with-feedback*), whereas control participants answered only simple questions. The motivation for this design was two-fold. First, we wanted to examine the effect of question type on memory performance as a within-participants comparison to maximise statistical power (i.e., a within-participants analysis in the cross-examination condition). Second, we wanted to examine whether the mere act of undergoing cross-examination affects how participants respond to simple questions and including the control condition enabled us to do this (i.e., a between-participants analysis of performance on simple questions).

An a priori power analysis for linear regression indicated that approximately 130 participants would be sufficient to detect a medium effect size ($f^2 = .15$) with .95 power ($\alpha = .05$). Guided by this and previous research, we aimed to recruit 150 participants to the cross-examination condition. We aimed to recruit 30 participants to the control condition, so that both conditions produced at least 600 observations for simple questions.

In total, we recruited 233 participants. Of these, 147 were first year psychology students at the University of Warwick who participated in partial fulfilment of course requirements. The remaining 86 participants were recruited through Prolific and received £2.50 upon completing the experiment. We excluded those who did not complete part 2 within 3–5 days of completing part 1 ($n = 21$), failed to comply with the criteria outlined in the experiment ($n = 20$), experienced technical difficulties ($n = 7$), or answered an attention check question incorrectly ($n = 4$). The final sample consisted of 181 participants (144 women, 35 men, 2 other/undisclosed, $M = 27.6$

years, $SD = 13.4$, range: 18–67).³ There were 151 participants in the cross-examination condition and 30 in the control condition. The Department of Psychology Research Ethics Committee at the University of Warwick approved this research.

Materials

We used the “Mischievous Melanie” mock crime sequence which contains 97 still coloured images that depict a woman stealing various items from an academic department on a university campus (Rasor et al., 2021). The sequence is 6 minutes and 28 s long, with each image being presented for 4 s. The images are accompanied by an audio narrative outlining the story. More information about the stimuli is available at <https://osf.io/znj2e/>.

The sequence contains 20 critical items. We chose to examine 5 types of cross-examination style questions that have been explored in previous research and are commonly used by lawyers (e.g., Kebbell et al., 2004, 2010; Kebbell & Johnson, 2000; Wheatcroft et al., 2004):

1. Simple (e.g., “Is it right that the lockers that the perpetrator tried to break into were blue?”, correct answer = “yes”).
2. Negative: included the word “not” once (e.g., “Is it not right that the lockers that the perpetrator tried to break into were blue?”, correct answer = “no”).
3. Double negative: included the word “not” twice (e.g., “Is it not right that the lockers that the perpetrator tried to break into were not blue?”, correct answer = “yes”).
4. Leading: strongly suggested the expected answer (e.g., “It is right that the lockers that the perpetrator tried to break into were purple, isn’t it?”, correct answer = “no”). Half of leading questions suggested a correct answer (correct answer = “yes”), and half suggested an incorrect answer (correct answer = “no”).
5. Leading-with-feedback: A leading question, followed by the suggestion that the correct response to the leading question was “yes” (e.g., “Most people said that the lockers that the perpetrator tried to break into were purple. They’re right about that, aren’t they?”, correct answer = “no”). Feedback was only given when participants answered “no” to the leading question, and the participant’s final response was included in the analysis. As Leading-with-feedback questions were the most complex question type, we have provided a diagram in the Appendix that shows how participants could respond to these questions and the feedback they would receive.

In the cross-examination condition, questions were divided into 5 sets of 20 questions with each question targeting a different critical item. Each set contained 4 simple, 4 negative, 4 double negative, 4 leading and 4 leading-with-feedback questions. These sets were fully counterbalanced such that each question type was presented the same number of times and the 20 critical items featured

equally often in every question type. Control participants always answered the same 20 simple questions. For each set and question type, the correct answer was “yes” to approximately half of the questions and “no” to the other half of the questions. The questions were tested to ensure that participants could not guess the correct answer based on the question format. The counterbalancing scheme is available on the OSF page.

Procedure

A general overview of the three-phase procedure is presented in Figure 1. Participants completed the study individually, online, and were randomly assigned to the cross-examination condition or the control condition. Participants were told that the study was about “learning styles and perception of still images” and were asked to comply with several requirements during the experiment (e.g., “Please do not speak to anyone during the experiment”). In Phase 1 (target event), all participants watched the mock crime sequence. The video was followed by two attention check questions and a 10-minute filler task involving logic problems.

In Phase 2, all participants completed the first memory test containing 20 forced-choice, yes/no questions each referring to a different critical item (e.g., “Were the lockers that the perpetrator tried breaking into blue?”). The questions were presented in chronological order and participants had unlimited time to respond. Participants were told that they must select Yes or No for each response, but that they could elaborate in the textbox next to their chosen response. Participants rated their confidence immediately after each response on an 11-point scale, ranging from 0% (*not at all confident*) to 100% (*very confident*).

Phase 3 began 3–5 days later ($M = 3.46$, $SD = 0.67$). Participants were randomly assigned to the cross-examination or control group. Cross-examination participants answered 20 questions including 4 of each question type (simple, negative, double negative, leading, and leading-with-feedback), whereas control participants answered 20 simple questions. Participants could elaborate on their responses in the box next to their chosen response and rated their confidence on the same 11-point scale. Once again, the questions were presented in chronological order and there was no time limit for responding. Finally, participants were asked if they had complied with the criteria outlined in Phase 1, answered demographic questions, and were debriefed.

Coding

The second author (ES) coded the responses for both memory tests, blind to question type and condition. For both memory tests, responses were coded as correct if participants selected the correct yes/no response, or if they correctly described the critical item regardless of the specificity of their response. Therefore, if the answer was “light green” then “green” and “lightly coloured” would

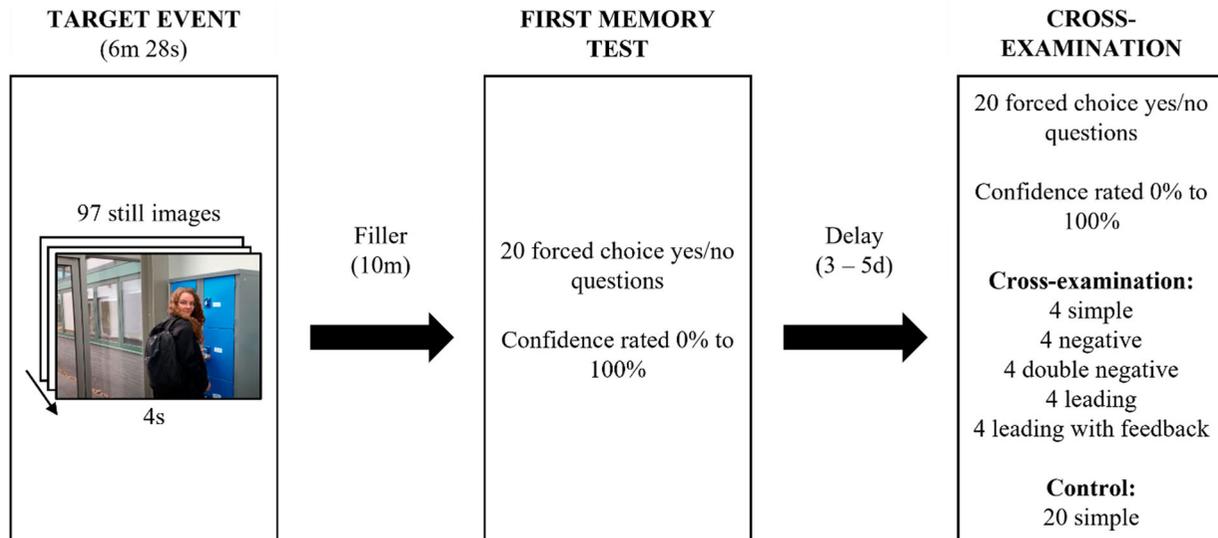


Figure 1. The general procedure for Experiment 1.

also be coded as correct. Accuracy was not coded for responses where participants wrote “don’t remember” next to their chosen response, because their confidence judgements did not always reflect their lack of certainty (Responses excluded on the first memory test: $n = 6$, $M = 45.00$, $SD = 44.61$, range = 0-100, Responses excluded on the cross-examination test: $n = 57$, $M = 29.30$, $SD = 23.44$, range = 0-90). Overall, 3,558 observations were coded and included in the analyses.

In line with previous research on cross-examination (Kebbell & Johnson, 2000), the correct yes/no response for negative questions was the opposite to the equivalent simple question, whereas the correct response to double negative questions was the same as the correct response for the equivalent simple question. For example, the correct answer to “Is it not right that the lockers that the perpetrator tried to break into were blue?” would be “no” because the lockers were blue, whereas the correct answer to the simple question “Is it right that the lockers that the perpetrator tried to break into were blue?” or the double negative question “Is it not right that the lockers that the perpetrator tried to break into were not blue?” would be “yes.”

For leading-with-feedback questions, only the final answer was included in the analysis. Feedback was only given when participants gave a “no” response to the initial leading question, such that the feedback always supported the same response as the leading question. When no feedback was given (i.e., when participants gave a “yes” response to the leading question), then the question was recoded as a leading question (rather than “leading-with-feedback”) for the analyses. In total, 857 leading questions and 351 leading-with-feedback questions were included in the analyses.

Participants were coded as changing their responses under cross-examination if they (1) answered correctly

on the first memory test and incorrectly on the cross-examination test or (2) answered incorrectly on the first memory test and correctly on the cross-examination test.

Results

Preliminary analyses

Before conducting our main analyses, we explored whether performance on simple questions, during the cross-examination phase, was similar between the cross-examination and control conditions. For each condition, we calculated mean accuracy on simple questions in the cross-examination phase, and this was calculated as the number of correct responses divided by the total number of correct and incorrect responses. The non-overlapping CIs indicated that cross-examination participants ($M = 72.74$, 95% CI [69.12, 76.36]) were significantly more accurate on simple questions than were control participants ($M = 64.76$, 95% CI [61.36, 68.15]). This result provides preliminary evidence that the act of responding to cross-examination style questions may influence witnesses’ responding on a global level. Cross-examination may serve to encourage participants to monitor the accuracy of their memories more carefully as they report them. Consistent with this interpretation, accuracy on the first memory test was similar for the cross-examination ($M = 76.73$, 95% CI [74.65, 78.81]) and control ($M = 73.49$, 95% CI [69.33, 77.64]) conditions, which suggests the differences observed on the simple questions between the two conditions were induced by the cross-examination process.

Main analyses

How do different types of cross-examination style questions affect witness accuracy? To test this, we calculated accuracy on the cross-examination test for each question type. The means and CIs for each condition are presented

in Table 1. Participants were significantly more accurate on simple questions than negative and double negative questions. Surprisingly, leading and leading-with-feedback questions produced similar accuracy to simple questions. These findings suggest that negative and double negative questions may impair witnesses' ability to provide an accurate report compared to simple questions, whereas leading and leading-with-feedback questions may have little to no effect on eyewitness accuracy under cross-examination. We return to these findings on memory accuracy in the exploratory analyses.

How do cross-examination style questions affect the relationship between witness confidence and accuracy? To answer this, we first created calibration curves by plotting participants' memory accuracy during the cross-examination test against their confidence. We used 3 confidence bins (0–40, 50–80, 90–100) which were chosen such that each confidence bin contained at least 100 observations. Figure 2 shows the confidence-accuracy calibration plots for each question type. At the lowest level of confidence (0–40) all question types produced a similar, moderate, level of accuracy. At medium (50–80) to high (90–100) levels of confidence, however, negative questions and double negative questions produced significantly lower accuracy than simple, leading, and leading-with-feedback questions. These results suggest that questions containing negatives and double negatives are more likely to impair eyewitness performance at the highest levels of confidence than simple questions, but performance at the lowest level of confidence is similar across question types. One important practical application of this result is that confidence judgements may be less informative about eyewitness accuracy when witnesses answer negative and double negative questions, than when they answer other types of cross-examination style questions.

Next, we examined whether witnesses were more likely to change their responses during cross-examination when (1) their initial answers were inaccurate than when their initial answers were accurate and (2) they were not very confident in their responses than when they were highly confident in their responses. We conducted a binomial logistic regression on response change (change vs no change) with binary accuracy (correct vs incorrect) and confidence on the first memory test as predictors.

Table 1. Means and 95% confidence intervals for accuracy and confidence by condition and question type in Experiment 1.

	Accuracy		Confidence	
	M	CI	M	CI
Control				
Simple	64.76	61.36, 68.15	60.83	54.18, 67.48
Cross-Examination				
Simple	72.74	69.12, 76.36	65.69	62.37, 69.01
Negative	46.85	42.89, 50.82	62.68	59.28, 66.09
Double negative	54.03	49.72, 58.34	61.89	58.41, 65.36
Leading	76.45	73.67, 79.23	66.77	63.57, 69.97
Leading-with-feedback	74.22	69.54, 78.90	62.02	57.84, 66.21

Responses were coded as changed when participants' responses changed in either direction (i.e., accurate to inaccurate or inaccurate to accurate) between the first memory test and the cross-examination test.

The model revealed that accuracy on the first memory was a significant predictor of response change, $\chi^2(1) = 42.62$, $p < .001$, and the odds ratios was 1.84, indicating that participants were 84% more likely to change initially inaccurate responses than initially accurate responses. This did not mean, however, that cross-examination improved overall memory accuracy in the current study. Note that although participants changed proportionally more initially incorrect responses (53.1%) than initially correct responses (31.9%), the overall number of correct responses changed ($n = 724$) was greater than the overall number of incorrect responses changed ($n = 369$). Given that participants changed more correct than incorrect responses, their overall memory accuracy decreased at cross-examination relative to the initial memory test. Of course, had participants performed more poorly on the initial memory test (reporting a much greater proportion of incorrect than correct answers) then we may have observed an increase in overall memory accuracy following cross-examination.

We predicted that participants would be more likely to change their responses that were given initially with low confidence, than responses given initially with high confidence. In line with that prediction, the model revealed that confidence during the first memory test was a significant predictor of response change under cross-examination, $\chi^2(1) = 107.31$, $p < .001$. The odds ratio was 1.13 indicating that each 10% reduction in confidence increased the likelihood that participants would change their response by 13%. In sum, participants were more likely to change their response under cross-examination when their initial response was inaccurate or when they were not very confident in their response.

Are witnesses who successfully monitor the accuracy of their responses more likely to improve their accuracy during cross-examination? To answer this question, we calculated accuracy change for each participant by subtracting the proportion of correct responses in the cross-examination test from the proportion of correct responses in the initial memory test. Thus, positive values reflected an increase in accuracy under cross-examination and negative values reflected a decrease in accuracy under cross-examination. We then conducted a multiple linear regression on accuracy change with resolution (ANRI) and calibration on the initial memory test and the cross-examination test as predictors. ANRI could not be calculated for four participants because they achieved 100% accuracy in the first memory test ($n = 3$) or in the cross-examination test ($n = 1$), so these participants were excluded from the analysis. The remaining 147 cross-examination participants were included in the analysis.

The model was a significant predictor of accuracy change $F(4, 142) = 3.74$, $p < .01$, $R^2 = .10$. Participants

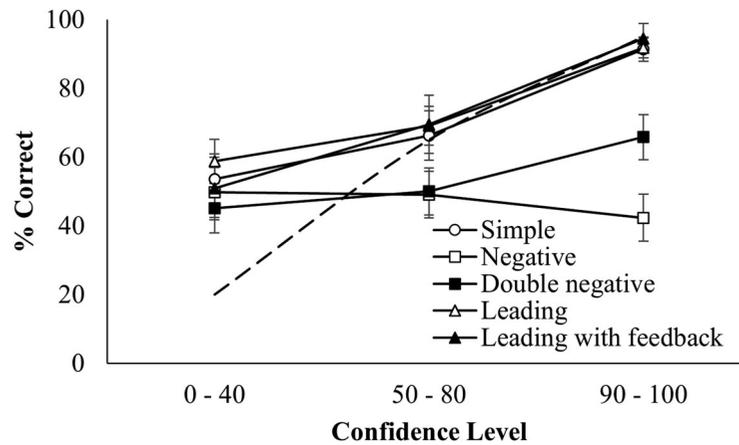


Figure 2. Calibration for each cross-examination question type in the cross-examination condition in Experiment 1. Note. The dashed line represents perfect calibration. Error bars denote the 95% CI around the mean.

with higher resolution (ANRI) scores on the first memory test were more likely to improve their accuracy during the cross-examination phase ($\beta = 0.16$, $p < .01$). None of the other predictors were significant ($p > .05$). Thus, participants who were better able to discriminate between correct and incorrect responses on the initial memory test were more likely to improve their overall accuracy under cross-examination.

Exploratory analyses

Recall that the leading and leading-with-feedback questions suggested the correct response to participants half of the time and the incorrect response half of the time (and the feedback always implied that the suggestion was correct). Therefore, it is possible that the effect of leading questions on eyewitness accuracy depended on the accuracy of the suggestion. Naturally we might expect witnesses to be more accurate on leading questions that contained a correct suggestion rather than an incorrect suggestion.

To examine this possibility, we calculated the proportion of correct responses by the accuracy of the suggestion in leading questions and leading-with-feedback questions. For leading questions, accuracy was relatively high regardless of whether the question contained a correct ($M = 80.24$, 95% CI [76.66, 83.83]) or incorrect suggestion ($M = 72.57$, 95% CI [67.94, 77.20]). Even more surprisingly, participants were more accurate on leading questions that included feedback when those questions contained an incorrect suggestion ($M = 89.24$, 95% CI [85.01, 93.46]) than when they contained a correct suggestion ($M = 27.65$, 95% CI [18.44, 36.85]). It is worth noting however, that participants only received feedback when they initially disagreed with the suggestion (i.e., answered “no”) in the leading question. Given that participants displayed relatively high accuracy on leading questions, it was relatively rare for participants to receive feedback on questions containing a correct suggestion because they usually answered “yes” to the initial question. Out of 341

leading-with-feedback questions containing a correct suggestion, participants received feedback only 29% of the time ($n = 99$, excluding questions where participants answered “don’t know”). These findings suggest that participants rarely changed their responses after feedback, even when it steered them towards the correct response. This may partly explain why leading-with-feedback questions did not affect eyewitness accuracy during cross-examination style questioning.

Whereas leading questions had little to no effect on eyewitness accuracy, negative and double negative questions significantly impaired the accuracy of witnesses’ reports. One explanation for this finding is that participants were confused about how to answer negative and double questions and, as a result, often gave the “wrong” answer even when they could remember the critical detail accurately. If participants are confused by negative and double negative questions, we might expect them to take longer to comprehend these questions than other types of cross-examination style questions. We calculated the mean response time for each question type, except leading-with-feedback questions because these questions consisted of two parts (the initial leading question and the subsequent feedback). To control for unmeasured variables that might affect response times (e.g., lapses in attention), response times more than 3 times above the mean were excluded from the analysis ($n = 61$, 1.71%). As expected, participants took longer to answer negative ($M = 10.23$, 95% CI [9.32, 11.14]) and double negative ($M = 13.64$, 95% CI [12.53, 14.75]) questions than simple ($M = 7.64$, 95% CI [7.01, 8.27]) and leading ($M = 8.11$, 95% CI [7.57, 8.65]) questions. These results provide some evidence that participants may have found negative and double negative questions confusing and took longer to comprehend these questions than other types of cross-examination style questions.

If participants answered negative and double negative questions incorrectly, despite having an accurate memory of the critical details in question, then we might also

expect participants to be more accurate when they elaborated on their responses than when they did not elaborate. In other words, if witnesses remember a detail accurately then they should be able to describe it correctly, even if they do not know whether “yes” or “no” is the correct response to give. Overall, cross-examination participants only elaborated on 10.89% ($n_{elaborated} = 323$, $n_{non-elaborated} = 2643$) of responses. Figure 3 shows that elaborated responses were more accurate than non-elaborated responses. This difference is largest for negative and double negative questions, with a smaller but significant difference for simple questions. There was no significant difference, however, for leading and leading-with-feedback questions. These results suggest that witnesses may have sometimes answered negative and double negative questions incorrectly because they were unsure of the correct response, even when they remembered the critical details accurately. Furthermore, these data suggest that encouraging witnesses to elaborate on their yes/no responses – which rarely happens under cross-examination (Ellison, 2001) – may be important for maximising the accuracy of eyewitness reports.

Conclusion

Experiment 1 suggests that the effect of cross-examination style questions on the accuracy of witness memory may depend on question type and witnesses’ confidence in their initial responses. Whereas negative and double negative questions impaired participants’ ability to provide accurate responses compared to simple questions, leading and leading-with-feedback questions did not significantly affect the accuracy of their responses. Negative and double negative questions produced a weaker confidence-accuracy relationship than other types of cross-examination style questions, producing the lowest accuracy and greatest overconfidence. Furthermore, participants were more likely to change their initially incorrect and low confidence responses than their initially correct and high confidence responses. Finally, participants who were better at discriminating between correct and incorrect responses on the first memory test were more likely to improve their accuracy under cross-examination.

The finding that leading and leading-with-feedback questions did not impair memory accuracy is particularly difficult to reconcile with previous research that has illustrated the robust, negative effects of leading questions on witness memory (Holst & Pezdek, 1992; Weinberg et al., 1983; Zaragoza et al., 2019). One potential explanation for this unexpected result could be that participants who were unsure of the correct response inferred that the suggestions contained in the leading questions were likely to be incorrect. Recall that during the cross-examination phase, the cross-examination participants were faced with a high portion of confusing, cross-examination style questions (16 out of 20 questions). This high proportion of complex questions to simple questions may have

heightened participants’ suspicion of the cross-examination test. Thus, when participants did not remember a suggested detail, they may have inferred that the leading question was highly likely to suggest the wrong answer. Such a mechanism would explain why participants rarely changed their responses after receiving feedback. That is, if participants believed that the leading questions and feedback were steering them away from the correct response, then it is unsurprising that they often stayed with their original “no” response after feedback.

The main aim of Experiment 2 was to explore this scepticism account by exposing participants to fewer confusing cross-examination style questions. If the high proportion of confusing questions in Experiment 1 induced participants to monitor their responses more carefully, then we might expect memory performance on leading and leading-with-feedback questions to be poorer if participants are exposed to fewer confusing cross-examination style questions. Indeed, in real-world legal settings, witnesses are likely to encounter only a small portion of questions in the “leading” or “leading-with-feedback” format while under cross-examination so a design in which participants encounter fewer complex questions better mimics the cross-examination scenario. Unlike in Experiment 1, participants in Experiment 2 answered cued-recall questions in the first memory test to better reflect the process of an initial police interview.

Experiment 2

Method

Participants & design

In total, we recruited 378 participants: 249 were recruited through Prolific and received £2.20 for completing the experiment, and the remaining 129 were first year psychology students at the University of Warwick who participated in partial fulfilment of course requirements.⁴ We excluded participants who experienced technical difficulties ($n = 11$), failed to comply with the criteria outlined in the experiment ($n = 10$) or answered an attention check question incorrectly ($n = 5$). The final sample consisted of 352 participants (266 women, 79 men, 7 other/undisclosed, $M = 33.98$, $SD = 14.74$, range:17-74). The Department of Psychology Research Ethics Committee at the University of Warwick approved this research. Unlike in Experiment 1, we used a within-subjects design and did not include a separate control group. Thus, all participants answered cross-examination style questions in the cross-examination memory test.

Materials

This experiment used the “Mischievous Melanie” mock crime sequence as in Experiment 1. Of the 20 items in the video, 5 served as critical items and 15 served as filler items. Critical items were chosen such that they were roughly equally distributed throughout the video

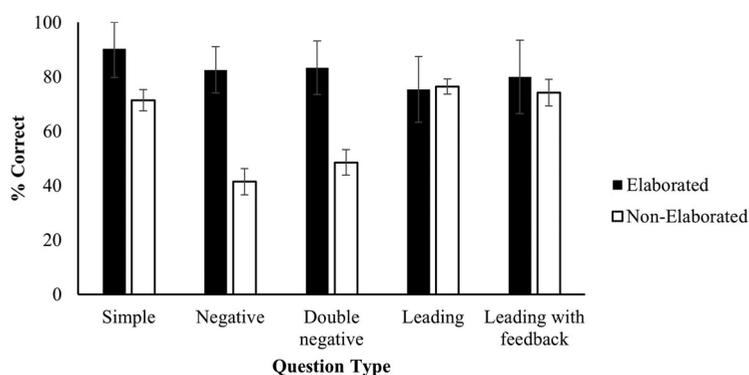


Figure 3. Mean accuracy for elaborated and non-elaborated responses by question type in the cross-examination condition in Experiment 1.

Note. Error bars denote the 95% CI around the mean.

and memory tests (i.e., items 3, 7, 11, 15 and 19). We created 5 questions for each critical item, including one of each type of cross-examination style question (simple, negative, double negative, leading and leading-with-feedback). Filler items were always targeted by simple questions. The questions were created the same way as in Experiment 1, except that leading and leading-with-feedback questions always suggested an incorrect response (see diagram in Appendix).

Questions were divided into 5 sets of 20 questions. Each set contained 5 cross-examination style questions each targeting a different critical item, and 15 filler questions. The critical items were counterbalanced so that each item appeared once in each question type, and each question type appeared once in each set. Put simply, participants answered one of each type of cross-examination style question and each targeted a different critical item. The counterbalancing scheme is available on the OSF page for this experiment.

Procedure

Participants completed the three-phase study online, individually, and were told that the study was about “learning styles and perception of still images”. They were also asked to comply with several criteria during the experiment (e.g., “Please do not speak to anyone during the experiment”). In Phase 1 (target event), participants watched the mock crime sequence, answered two attention check questions, and completed a 10-minute filler task involving logic problems.

In Phase 2 (first memory test), participants completed the first memory test consisting of 20 cued-recall questions each referring to a different item in the video (e.g., “What colour were the lockers that the perpetrator tried to break into?”) and rated their confidence immediately after each response on an 11-point scale, ranging from 0% (*not at all confident*) to 100% (*very confident*). As in Experiment 1, questions were presented in chronological order and there was no time limit for participants to give their responses.

Phase 3 (cross-examination) began 3–5 days later ($M = 3.35$, $SD = 0.58$), and participants completed a 20-item forced-choice, yes/no memory test which was similar to that used in Experiment 1, except that it consisted of 5 cross-examination style questions and 15 filler questions. Participants could elaborate on their responses in the box next to their chosen response and they were asked to rate their confidence on the same 11-point scale. Finally, participants indicated if they had complied with the criteria outlined in Phase 1, answered demographic questions, and were debriefed.

Coding

The second author (ES) coded the responses for both memory tests, blind to question type. For the first memory test, responses were coded as correct if participants correctly described the critical item, regardless of the specificity of their response. Responses in the cross-examination test were coded in the same way as Experiment 1. Again, responses were coded as correct if participants gave the correct yes/no response or if they correctly described the critical item, regardless of the specificity of their response. Accuracy was not coded for responses where participants wrote “don’t know” in either memory test, because their confidence judgements did not always reflect their lack of certainty (Responses excluded on the first memory test: $n = 857$, $M = 30.94$, $SD = 44.57$, range = 0-100; Responses excluded on the cross-examination test: $n = 59$, $M = 33.05$, $SD = 35.63$, range = 0-100). Overall, 6,124 observations were coded and included in the analyses.

As in Experiment 1, the correct yes/no response to negative questions was the opposite to the equivalent simple question, whereas the correct response to double negative questions was the same as the correct response for the equivalent simple question. In contrast to Experiment 1, however, the correct answer to leading and leading-with-feedback questions was always “no”, because the question always suggested an incorrect response.

For leading-with-feedback questions, feedback was only given when participants correctly answered “no” to the initial leading question and only the final, post-feedback answer was included in the analyses. No feedback was given when participants incorrectly answered “yes” to the initial leading question. “No feedback” responses were excluded from the analysis because participants’ initial response was always incorrect (i.e., they agreed with the leading suggestion), so recoding them as leading questions (as we did in Experiment 1) would artificially reduce accuracy on leading questions (see deviations from the pre-registration on the OSF page). In total, 66 “no feedback” responses were removed and 255 leading-with-feedback questions were included in the analysis.

Results

Main analyses

Recall that the main aims were to examine the effect of leading questions when the cross-examination phase included fewer confusing cross-examination style questions, and to replicate the findings of Experiment 1. Table 2 shows the mean accuracy scores and CIs for each question type. Similar to Experiment 1, negative and double negative questions produced the lowest accuracy. However, only negative questions significantly impaired accuracy compared to simple questions. In contrast to Experiment 1, leading and leading-with-feedback questions produced significantly higher accuracy than simple questions. In fact, leading-with-feedback questions produced the highest accuracy of any question type during cross-examination. We return to this finding in the exploratory analyses. These findings suggest that negative questions may impair the accuracy of witnesses’ reports compared to other types of cross-examination style questions, whereas leading and leading-with-feedback questions may enhance eyewitness accuracy.

Turning now to participants’ confidence ratings during cross-examination. Recall that in Experiment 1, the confidence-accuracy relationship was relatively strong for simple, leading and leading-with-feedback questions, but negative and double negative questions impaired this relationship. Specifically, negative and double negative questions impaired eyewitness accuracy at the highest levels of confidence compared to simple questions, but all question types produced a similar level of accuracy at the lowest level of confidence. Figure 4 shows that, in

contrast to Experiment 1, leading and leading-with-feedback questions produced significantly higher accuracy and greater underconfidence at the lowest level of confidence than other types of cross-examination style questions. However, simple, negative and double negative questions produced a similar level of accuracy at low (0–40) to medium (50–80) levels of confidence. Consistent with Experiment 1, negative questions produced the lowest accuracy and greatest overconfidence at the highest level (90–100) of confidence. Double negative questions also produced significant overconfidence, but accuracy did not significantly differ from that for simple or leading questions. Together, these findings suggest that leading and leading-with-feedback questions may produce relatively high accuracy across all levels of confidence, whereas negative and double negative questions may tend to impair accuracy at the highest level of confidence. Given that participants answered fewer confusing cross-examination questions than in Experiment 1, and still performed relatively well on leading and leading-with-feedback questions, these findings do not support the idea that answering a relatively high number of confusing questions led participants to monitor their memories more carefully.

Were participants more likely to change initially accurate and low confidence responses than initially accurate and high confidence responses? We conducted a binomial logistic regression on response change (change vs no change) with binary accuracy (correct vs incorrect) and confidence on the first memory test as predictors. Again, participants were coded as having changed their responses when the accuracy of their responses changed between the first memory test and the cross-examination test. Consistent with Experiment 1, accuracy on the first memory test, $\chi^2(1) = 41.38$, $p < .001$ was a significant predictor of accuracy change. The odds ratio was 2.33, suggesting that participants were more than twice as likely to change an initially incorrect response than a correct response. Confidence on the first memory test also significantly predicted response change, $\chi^2(1) = 29.33$, $p < .001$, and the odds ratio was 1.10, indicating that each 10% decrease in confidence increased the likelihood that participants would change their responses by 10%. Thus, in line with Experiment 1, participants were more likely to change their initial responses when their responses were not very accurate, or when they were not very confident in their responses.

Are witnesses who are better at discriminating between correct and incorrect responses more likely to improve their accuracy during cross-examination? To test this, we conducted a multiple linear regression on accuracy change with resolution (ANRI) and calibration (C) on both the initial memory test and the cross-examination test as predictors. ANRI could not be calculated for seven participants because they achieved 100% accuracy in the first memory test ($n = 4$) or in the cross-examination test ($n = 3$). The remaining 345 participants were included in

Table 2. Means and 95% confidence intervals for accuracy and confidence by question type in Experiment 2.

	Accuracy		Confidence	
	M	CI	M	CI
Simple	60.13	54.68, 65.58	52.77	48.89, 56.64
Negative	47.45	41.92, 52.98	51.75	47.91, 55.60
Double negative	52.40	46.85, 57.94	52.36	48.46, 56.27
Leading	72.61	67.67, 77.55	52.58	48.85, 56.31
Leading-with-feedback	91.76	88.38, 95.15	51.18	46.87, 55.48

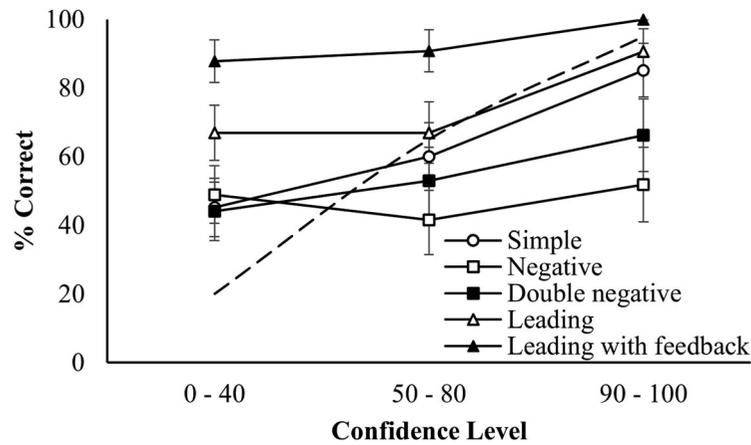


Figure 4. Calibration for each question type in the cross-examination test in Experiment 2.

Note. The dashed line represents perfect calibration. Error bars denote the 95% CI around the mean.

the analyses. The model significantly predicted accuracy change, $F(4, 340) = 7.55$, $p < .001$, $R^2 = .08$. Participants who achieved higher resolution scores in the first memory test, ($\beta = 0.16$, $p < .01$), or relatively strong calibration in the cross-examination memory test, ($\beta = 0.85$, $p < .001$), were more likely to improve their accuracy during cross-examination. Neither calibration on the first memory test or resolution on the cross-examination test significantly predicted accuracy change ($p > .05$). Thus, in line with Experiment 1, participants who were good at discriminating between accurate and inaccurate responses on the first memory test, or whose confidence judgements were closely aligned with their actual accuracy during the cross-examination phase, were more likely to improve their accuracy during cross-examination.

Exploratory analyses

Recall that leading-with-feedback questions produced the highest accuracy and the greatest underconfidence during cross-examination. One possible explanation for these findings is that participants rarely changed their responses after receiving feedback. In the current experiment, leading questions always suggested an incorrect response,

and participants only received feedback when they disagreed with this suggestion (i.e., when they answered “no” to the preceding leading question). Put another way, participants only received feedback when they answered the initial leading question correctly. One consequence of this is that accuracy will be very high if the feedback fails to convince participants to change their (correct) response to the initial leading question. To examine whether this could explain the high accuracy on leading-with-feedback questions, we calculated how often participants changed their responses to the leading question after feedback. Out of 255 leading-with-feedback questions, participants only changed their responses after feedback 5% ($n = 13$) of the time. These results suggest that the feedback rarely shifted participants’ responses and goes some way to explaining why accuracy was particularly high on leading-with-feedback questions.

In Experiment 1, we found that participants were more accurate when they elaborated on their yes/no responses than when they did not elaborate, and this difference was largest for negative and double negative questions. Consistent with this result, Figure 5 shows that accuracy on negative and double negative questions was significantly

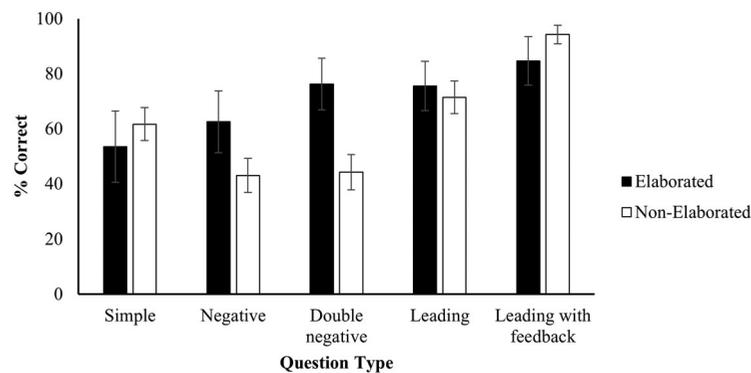


Figure 5. Mean accuracy for elaborated and non-elaborated responses by question type in the cross-examination test in Experiment 2.

Note. Error bars denote the 95% CI around the mean.

higher when participants elaborated on their responses, than when they did not elaborate on their responses. Accuracy on simple, leading and leading-with-feedback questions was similar regardless of whether participants elaborated on their responses. These results suggest that participants may have sometimes answered negative and double questions incorrectly during cross-examination even when their memory should have allowed them to answer correctly.

Discussion

Across two experiments, we examined how different types of cross-examination style questions affect the accuracy of witnesses' reports. Negative and double negative questions sometimes impaired the accuracy of witnesses' responses during cross-examination, whereas leading and leading-with-feedback questions did not impair – and sometimes enhanced – the accuracy of witnesses' responses. Participants who were better at discriminating between correct and incorrect responses were more likely to improve the accuracy of their reports during cross-examination. Together, these results suggest that the effect of cross-examination style questions on eyewitness accuracy depends on question type and witnesses' confidence in their responses.

The finding that negative questions impaired memory accuracy fits with previous research showing that these types of questions produce less accurate responses than simple questions (Kebbell & Giles, 2000; Kebbell & Johnson, 2000). Although we did not set out to explore underlying mechanisms, negative questions may impair witness accuracy because people struggle to determine whether Yes or No is the required answer. Recall that the correct response to negative questions was the opposite to the equivalent simple question, yet some participants may have answered negative questions in the same way as simple questions. For example, if participants believed that the lockers the perpetrator tried to break into were blue, then they may answer “yes” to the question “Is it not right that the lockers that the perpetrator tried to break into were blue?” regardless of whether it contained the negative (Walker, 1998). If participants answered negative questions in the same way as simple questions, then their responses would have been coded as incorrect even if their memory should have enabled them to respond correctly. This linguistic account may, at least partly, explain why negative questions impaired the accuracy of witnesses' reports compared to simple questions. Relatedly, recent research on linguistic differences across cultures suggest that culture may influence how people respond to cross-examination style questions (Hope et al., 2022). This could be a fruitful topic for future research.

Recall that participants tended to take longer to answer negative and double negative questions than simple questions, which is consistent with previous work (Kebbell

et al., 2010). This finding fits with the notion that participants struggled to decide on the correct answer to these questions. It is interesting that participants rarely acknowledged that their responses to negative and double negative questions might be ambiguous by either [a] elaborating on their responses or [b] lowering their confidence to reflect their lack of certainty that they had chosen the correct answer. Indeed, as the results showed, participants rarely elaborated on their responses, and participants' confidence on negative and double negative questions was similar to that on simple questions. If real witnesses fail to recognise when their responses could be misinterpreted, then they may confidently give a response which does not accurately reflect the content of their memory.

Previous research shows that participants who are exposed to misinformation often come to report this information in their memory reports (Flowe et al., 2019; Greene et al., 2021; Ito et al., 2019; Spearing & Wade, 2021; Stark et al., 2010). Why, then, did the leading questions in the current study fail to impair the accuracy of eyewitness reports when those questions suggested an incorrect response? One explanation is that participants' original memories were strengthened and protected – at least to some extent – by the recall task they completed before being exposed to misinformation during the cross-examination phase. In a typical misinformation experiment, mock witnesses are exposed to misinformation before their memory of the critical event is tested and thus are more likely to fall prey to the misinformation. This interpretation is supported by research showing that people who complete a recall task before encountering misinformation are less likely to report misinformation in their later memory reports than those who do not complete an initial recall task (Gabbert et al., 2012).

This initial test mechanism may also explain why participants rarely changed their responses following negative feedback. Indeed, research suggests that feedback is likely to have a greater impact on eyewitness memory reports when a witness's memory is relatively weak than when their memory is relatively strong (Charman et al., 2010; Iida et al., 2020). Thus, completing an initial memory test may have enabled our participants to perform relatively well on leading and leading-with-feedback questions. We should also note that the delay periods (between target event and memory test) in the current studies were short compared to those typically experienced by witnesses in the real world. Future research should examine the effects leading questions and feedback following longer delays when memory has more time to fade.

Consistent with a growing body of research on eyewitness confidence and accuracy (Ackerman & Goldsmith, 2008; Weber & Brewer, 2008; Wixted et al., 2018), we found that people were more likely to change responses that they were not very confident in than responses that they were highly confident in. Furthermore, participants

who were better at discriminating between correct and incorrect (initial) responses were more likely to improve their accuracy, later, during cross-examination. These findings are consistent with previous research showing that participants use metacognitive monitoring to guide their responses during cross-examination (Jack & Zajac, 2014) and suggest that questions that enable witnesses to better estimate their memory accuracy may help to improve the accuracy of their responses during cross-examination. Put simply, if lawyers ask questions that enhance witnesses' ability to discriminate between correct and incorrect responses, then witnesses should be more likely to change inaccurate responses than accurate responses, increasing the accuracy of their reports. If, however, lawyers' questions make it hard for witnesses to discriminate between correct and incorrect responses, then witnesses may change their responses indiscriminately at a cost to their overall accuracy.

The link between confidence and performance under cross-examination has important implications for witness preparation training. A key aim of witness preparation is to enhance witnesses' testimony delivery skills, and a witness's confidence is often a key target during preparation (Boccaccini, 2002; Boccaccini et al., 2004). Preparation training often involves role-playing, where the attorney or lawyer will ask the witness to answer direct and cross-examination questions to see how the witness might improve their testimony delivery style. We know that repeatedly rehearsing an event in memory serves to inflate confidence in the accuracy of that memory (Shaw, 1996; Shaw & McClure, 1996). As such, preparation training may have the unintentional (yet beneficial, from the attorney's perspective) side-effect of protecting witnesses from the deleterious effects of cross-examination. To be clear, we are not advocating "memory confidence" boosting as part of witness preparation training – if a witness's memory is distorted to begin with then inflating their confidence will likely hinder, not help, the fact-finding process. Nevertheless, having a better mechanistic understanding of cross-examination will help researchers to determine the conditions that promote accurate memory retrieval.

Our findings highlight at least two fruitful areas for future research. First, it is not clear whether cross-examination has long-term effects on eyewitness memory. We do not yet know whether cross-examination affects the content of witnesses' memory, or which types of cross-examination style questions are the most likely to damage the accuracy of witnesses' future memory reports. Second, as far as we are aware, no published research into cross-examination and witness behaviour has explored the effect of cross-examination style questions on non-cooperative, reluctant or deceptive witnesses. Future research should explore the extent to which cross-examination style questions influence the responses of witnesses who are not forthcoming with the truth.

To conclude, we report the first systematic investigation of how different types of cross-examination style questions affect eyewitness accuracy and confidence. There is a surprising dearth of research in the applied psychological literature on how cross-examination affects witness behaviour or how confidence affects witnesses' responding under cross-examination. Lawyers often assume that reliable and co-operative witnesses are not affected by the complex and suggestive questioning techniques used during cross-examination (Henderson, 2015). Our findings add to the small but growing body of literature that suggests this assumption is wrong: Cross-examination style questions may be the accepted means of probing a witnesses' evidence, but the answers delivered – even with high levels of confidence – may not be any more accurate.

Notes

1. 3 Wigmore, Evidence §1367, p. 27 (2d ed. 1923).
2. It is worth noting that these studies used a between-participants design with only ~10-20 participants per condition so the results may not be stable.
3. Performance on cross-examination style questions was similar for male and female participants in both experiments. There was insufficient variance in age to examine age differences in Experiment 1, with over 50% of participants being aged 18–19 years. However, an analysis of Experiment 2 suggested that age does not significantly affect accuracy for any type of cross-examination style question.
4. In our original pre-registration, we stated that we would collect data from 200 participants. We later realised that this was a miscalculation and that it would not give us enough data points to plot stable calibration curves. Therefore, we continued to collect data until we had at least 250 usable observations per question type, with at least 50 observations in each confidence bin. A document describing the changes to our pre-registration is available on the OSF page.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The second author was supported by a University of Warwick Departmental PhD Fellowship and a research grant from the European Association of Psychology and Law.

Data availability statement

The pre-registrations, numeric data, and corresponding R code are available on the Open Science Framework (https://osf.io/tv9pg/?view_only=002448b94461447b9425ddeb4f5e8be9) and materials can be accessed on request from the corresponding author.

ORCID

Kimberley A. Wade  <http://orcid.org/0000-0002-7134-9600>
Emily R. Spearing  <http://orcid.org/0000-0002-7304-8828>

References

- Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting—with and without satisfying knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1224–1245. <https://doi.org/10.1037/a0012938>
- Boccaccini, M. T. (2002). What do we really know about witness preparation? *Behavioral Sciences & the Law*, 20(1–2), 161–189. <https://doi.org/10.1002/bsl.472>
- Boccaccini, M. T., Gordon, T., & Brodsky, S. L. (2004). Effects of witness preparation on witness confidence and nervousness. *Journal of Forensic Psychology Practice*, 3(4), 39–51. https://doi.org/10.1300/J158v03n04_03
- Charman, S. D., Carlucci, M., Vallano, J., & Gregory, A. H. (2010). The selective cue integration framework: A theory of postidentification witness confidence assessment. *Journal of Experimental Psychology: Applied*, 16(2), 204–218. <https://doi.org/10.1037/a0019495>
- Defence-Barrister. (n.d.). Defence-Barrister: Conducting a cross-examination. Retrieved 19 March 2022, from <https://www.defence-barrister.co.uk/cross-examination#Conducting-a-cross-examination>
- Ellison, L. (2001). The mosaic effect?: Cross-examination and the vulnerable witness. *Legal Studies*, 21(3), 353–375. <https://doi.org/10.1111/j.1748-121X.2001.tb00172.x>
- Flowe, H. D., Humphries, J. E., Takarangi, M. K., Zelek, K., Karoğlu, N., Gabbert, F., & Hope, L. (2019). An experimental examination of the effects of alcohol consumption and exposure to misleading postevent information on remembering a hypothetical rape scenario. *Applied Cognitive Psychology*, 33(3), 393–413. <https://doi.org/10.1002/acp.3531>
- Gabbert, F., & Hope, L. (2018). Suggestibility in the courtroom: How memory can be distorted during the investigative and legal process. In H. Otgaar (Ed.), *Finding the truth in the courtroom: Dealing with deception, lies, and memories*, Vol. 277 (pp. 31–57). Oxford University Press, x.
- Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Applied Cognitive Psychology*, 26(4), 568–575. <https://doi.org/10.1002/acp.2828>
- Greene, C. M., Bradshaw, R., Huston, C., & Murphy, G. (2021). The medium and the message: Comparing the effectiveness of six methods of misinformation delivery in an eyewitness memory paradigm. *Journal of Experimental Psychology: Applied*. Advance online publication.
- Henderson, E. (2015). Bigger fish to fry: Should the reform of cross-examination be expanded beyond vulnerable witnesses? *The International Journal of Evidence & Proof*, 19(2), 83–99. <https://doi.org/10.1177/1365712714568072>
- Henkel, L. A. (2017). Inconsistencies across repeated eyewitness interviews: Supportive negative feedback can make witnesses change their memory reports. *Psychology, Crime & Law*, 23(2), 97–117. <https://doi.org/10.1080/1068316X.2016.1225051>
- Hickey, L. (1993). Presupposition under cross-examination. *International Journal for the Semiotics of Law*, 6(1), 89–109. <https://doi.org/10.1007/BF01458741>
- Holst, V. F., & Pezdek, K. (1992). Scripts for typical crimes and their effects on memory for eyewitness testimony. *Applied Cognitive Psychology*, 6(7), 573–587. <https://doi.org/10.1002/acp.2350060702>
- Hope, L., Anakwah, N., Antfolk, J., Brubacher, S. P., Flowe, H., Gabbert, F., Giebels, E., Kanja, W., Korkman, J., Kyo, A., Naka, M., Otgaar, H., Powell, M. B., Selim, H., Skrifvars, J., Sorkpah, I. K., Sowatey, E. A., Steele, L. C., Stevens, L., & Anonymous. (2022). Urgent issues and prospects at the intersection of culture, memory, and witness interviews: Exploring the challenges for research and practice. *Legal and Criminological Psychology*, 27(1), 1–31. <https://doi.org/10.1111/lcrp.12202>
- Hope, L., & Gabbert, F. (2019). Interviewing witnesses and victims. In N. Brewer, & A. B. Douglass (Eds.), *Psychological science and the law* (pp. 130–156). The Guildford Press.
- Iida, R., Itsukusima, Y., & Mah, E. Y. (2020). How do we judge our confidence? Differential effects of meta-memory feedback on eyewitness accuracy and confidence. *Applied Cognitive Psychology*, 34, 397–408. <https://doi.org/10.1002/acp.3625>
- Ito, H., Barzykowski, K., Grzesik, M., Gülgöz, S., Gürdere, C., Janssen, S. M. J., Khor, J., Rowthorn, H., Wade, K. A., Luna, K., Albuquerque, P. B., Kumar, D., Singh, A. D., Cecconello, W. W., Cadavid, S., Laird, N. C., Baldassari, M. J., Lindsay, D. S., & Mori, K. (2019). Eyewitness memory distortion following co-witness discussion: A replication of Garry, French, Kinzett, and Mori (2008) in Ten countries. *Journal of Applied Research in Memory and Cognition*, 8(1), 68–77. <https://doi.org/10.1037/h0101833>
- Jack, F., & Zajac, R. (2014). The effect of age and reminders on witnesses' responses to cross-examination-style questioning. *Journal of Applied Research in Memory and Cognition*, 3(1), 1–6. <https://doi.org/10.1016/j.jarmac.2013.12.001>
- Kapardis, A. (2010). *Psychology and law: A critical introduction* (3rd ed.). Cambridge University Press.
- Keibell, M., Deprez, S., & Wagstaff, G. (2003). The direct and cross-examination of complainants and defendants in rape trials: A quantitative analysis of question type. *Psychology, Crime & Law*, 9(1), 49–59. <https://doi.org/10.1080/10683160308139>
- Keibell, M., Evans, L., & Johnson, S. D. (2010). The influence of lawyers' questions on witness accuracy, confidence, and reaction times and on mock jurors' interpretation of witness accuracy. *Journal of Investigative Psychology and Offender Profiling*, 7(3), 262–272. <https://doi.org/10.1002/jip.125>
- Keibell, M., & Giles, D. (2000). Some experimental influences of lawyers' complicated questions on eyewitness confidence and accuracy. *The Journal of Psychology*, 134(2), 129–139. <https://doi.org/10.1080/00223980009600855>
- Keibell, M., Hatton, C., & Johnson, S. D. (2004). Witnesses with intellectual disabilities in court: What questions are asked and what influence do they have? *Legal and Criminological Psychology*, 9(1), 23–35. <https://doi.org/10.1348/135532504322776834>
- Keibell, M., & Johnson, S. D. (2000). Lawyers' questioning: The effect of confusing questions on witness confidence and accuracy. *Law and Human Behavior*, 24(6), 629–641. <https://doi.org/10.1023/A:1005548102819>
- Laver, N. (n.d.). Cross-examination in criminal cases. *In brief*. Retrieved 19 March 2022, from <https://www.inbrief.co.uk/court-proceedings/cross-examination/>
- Levy, E. (1991). *Examination of witnesses in criminal cases* (2nd ed.). Thompson Professional Publishing.
- O'Neill, S., & Zajac, R. (2013). The role of repeated interviewing in children's responses to cross-examination-style questioning: Cross-examination and repeated interviewing. *British Journal of Psychology*, 104(1), 14–38. <https://doi.org/10.1111/j.2044-8295.2011.02096.x>
- Paulo, R. M., Albuquerque, P. B., & Bull, R. (2019). Witnesses' verbal evaluation of certainty and uncertainty during investigative interviews: Relationship with report accuracy. *Journal of Police and Criminal Psychology*, 34(4), 341–350. <https://doi.org/10.1007/s11896-019-09333-6>
- Plotnikoff, J., & Woolfson, R. (2009). *Measuring up? Evaluating implementation of government commitments to young witnesses in criminal proceedings*. NSPCC and Nuffield Foundation.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rasor, S. A., Spearing, E., King-Nyberg, B., Mah, E. Y., Lindsay, D. S., & Wade, K. A. (July, 2021). *Online co-witness discussions can lead to memory conformity*. Co-witness conformity in an online paradigm. Poster presented at the virtual SARMAC conference 2021.

- Righarts, S., Jack, F., Zajac, R., & Hayne, H. (2015). Young children's responses to cross-examination style questioning: The effects of delay and subsequent questioning. *Psychology, Crime & Law*, 21(3), 274–296. <https://doi.org/10.1080/1068316X.2014.951650>
- Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied*, 8(1), 33–43. <https://doi.org/10.1037/1076-898X.8.1.33>
- Sauer, J., & Hope, L. (2016). The effects of divided attention at study and reporting procedure on regulation and monitoring for episodic recall. *Acta Psychologica*, 169, 143–156. <https://doi.org/10.1016/j.actpsy.2016.05.015>
- Shaw, J. S. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied*, 2(2), 126–146. <https://doi.org/10.1037/1076-898X.2.2.126>
- Shaw, J. S., & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, 20(Issue 6), 629–653. <https://doi.org/10.1007/BF01499235>
- Spearing, E. R., & Wade, K. A. (2021). Providing eyewitness confidence judgments during versus after eyewitness interviews does not affect the confidence-accuracy relationship. *Journal of Applied Research in Memory and Cognition*, 11(1), 54–65. <https://psycnet.apa.org/doi/10.1037/h0101868>
- Spearing, E. R., & Wade, K. A. (2022). Long retention intervals impair the confidence-accuracy relationship for eyewitness recall. *Journal of Applied Research in Memory and Cognition*. Advance online publication. <https://doi.org/10.1037/mac0000014>
- Spencer, J. R. (2012). Conclusions. In J. R. Spencer, & M. E. Lamb (Eds.), *Children and cross-examination: Time to change the rules* (pp. 171–202). Hart Publishing.
- Stark, C. E. L., Okado, Y., & Loftus, E. F. (2010). Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms. *Learning & Memory*, 17(10), 485–488. <https://doi.org/10.1101/lm.1845710>
- Stone, M. (1988). *Cross-examination in criminal trials*. Butterworths.
- Thorley, C., & Kumar, D. (2017). Eyewitness susceptibility to co-witness misinformation is influenced by co-witness confidence and own self-confidence. *Psychology, Crime & Law*, 23(4), 342–360. <https://doi.org/10.1080/1068316X.2016.1258471>
- Valentine, T., & Maras, K. (2011). The effect of cross-examination on the accuracy of adult eyewitness testimony. *Applied Cognitive Psychology*, 25(4), 554–561. <https://doi.org/10.1002/acp.1768>
- van Bergen, S., Horselenberg, R., Merckelbach, H., Jelicic, M., & Beckers, R. (2010). Memory distrust and acceptance of misinformation. *Applied Cognitive Psychology*, 24(6), 885–896. <https://doi.org/10.1002/acp.1595>
- Van Boeijen, I. M., & Saraiva, R. (2018). *Legalpsych: A tool for calculating calibration statistics in eyewitness research*. GitHub. <https://github.com/lngerMathilde/legalPsych>
- Vredevelde, A., & Sauer, J. D. (2015). Effects of eye-closure on confidence-accuracy relations in eyewitness testimony. *Journal of Applied Research in Memory and Cognition*, 4(1), 51–58. <https://doi.org/10.1016/j.jarmac.2014.12.006>
- Walker, A. G. (1998). *Linguistic analysis of two complex competency questions*. 27th international congress of applied psychology, San Francisco.
- Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied*, 14(1), 50–60. <https://doi.org/10.1037/1076-898X.14.1.50>
- Weinberg, H. I., Wadsworth, J., & Baron, R. S. (1983). Demand and the impact of leading questions on eyewitness testimony. *Memory & Cognition*, 11(1), 101–104. <https://doi.org/10.3758/BF03197667>
- Wheatcroft, J. M., Wagstaff, G. F., & Kebbell, M. R. (2004). The influence of courtroom questioning style on actual and perceived eyewitness confidence and accuracy. *Legal and Criminological Psychology*, 9(1), 83–101. <https://doi.org/10.1348/13553250432276870>
- Wheatcroft, J. M., & Woods, S. (2010). Effectiveness of witness preparation and cross-examination non-directive and directive leading question styles on witness accuracy and confidence. *The International Journal of Evidence & Proof*, 14(3), 187–207. <https://doi.org/10.1350/ijep.2010.14.3.353>
- Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science*, 13(3), 324–335. <https://doi.org/10.1177/1745691617734878>
- Zajac, R., Gross, J., & Hayne, H. (2003). Asked and answered: Questioning children in the courtroom. *Psychiatry, Psychology, and Law: An Interdisciplinary Journal of the Australian and New Zealand Association of Psychiatry, Psychology and Law*, 10(1), 199–209. <https://doi.org/10.1375/pplt.2003.10.1.199>
- Zajac, R., & Hayne, H. (2003). I don't think that's what really happened: The effect of cross-examination on the accuracy of children's reports. *Journal of Experimental Psychology: Applied*, 9(3), 187–195. <https://doi.org/10.1037/1076-898X.9.3.187>
- Zajac, R., & Hayne, H. (2006). The negative effect of cross-examination style questioning on children's accuracy: Older children are not immune. *Applied Cognitive Psychology*, 20(1), 3–16. <https://doi.org/10.1002/acp.1169>
- Zajac, R., Jury, E., & O'Neill, S. (2009). The role of psychosocial factors in young children's responses to cross-examination style questioning. *Applied Cognitive Psychology*, 23(7), 918–935. <https://doi.org/10.1002/acp.1536>
- Zaragoza, M. S., Hyman, I., & Chrobak, Q. M. (2019). False Memory. In N. Brewer & A. B. Douglass, (Eds.), *Psychological science and the law* (pp. 182–207). The Guildford Press.

Appendix

Overview of Leading-with-feedback Questions in Experiments 1 and 2.

