Journal of Applied Research in Memory and Cognition 5 (2016) 159–167

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Metacognitive Effects of Initial Question Difficulty on Subsequent Eyewitness Memory Performance☆

Shiri Portnoy, Ainat Pansky*

Department of Psychology, University of Haifa, Israel

In this study, we examined the influence of initial question difficulty on the confidence in the answers to subsequent questions, affecting the tendency of eyewitnesses to report them. Target questions of intermediate difficulty about event details were preceded by either difficult or easy questions. In contrast to forced-report performance, free-report performance was affected by initial question difficulty: When preceded by difficult questions, more answers to the target questions were confidently held and hence were more likely to be reported, yielding a larger quantity of correct reported answers. These findings demonstrate how changes in subjective experience, as a result of initial question difficulty, can influence metacognitive monitoring and control, thereby affecting free-report eyewitness memory performance. From an applied perspective, our findings suggest that preceding questions about a witnessed event by relatively difficult as opposed to relatively easy questions can yield more event information from eyewitnesses, resulting in more complete eyewitness reports.

Keywords: Relative fluency, Metacognition, Monitoring and control, Confidence, Eyewitness memory, Free-report memory performance

When eyewitnesses are questioned, they typically have the freedom of deciding which questions to answer and which to refrain from answering. Under such free-report conditions, two measures of memory performance can be distinguished: *Memory quantity* and *memory accuracy* (see Koriat & Goldsmith, 1994, 1996). Memory quantity is *input-bound*, reflecting the likelihood that an event detail will be remembered correctly (or the completeness of an eyewitness account). In contrast, memory accuracy is *output-bound*, reflecting the likelihood that a reported event detail is correct (or the reliability of the reported information).

Previous studies have shown that certain aspects of eyewitness questioning can affect the quantity and/or the accuracy of the reported information (for a review, see Pansky, Koriat, & Goldsmith, 2005). For example, merely using a definite article when questioning about an object that was not part of the original event (e.g., "Did you see *the* broken headlight?") rather than an indefinite article ("Did you see *a* broken headlight?") has been shown to bias eyewitnesses into falsely remembering the specified object, thereby reducing memory accuracy (Loftus & Zanni, 1975). In contrast, questioning eyewitnesses using the Cognitive Interview—a technique that assists eyewitnesses in recollecting event details by utilizing psychological principles (e.g., tailoring questions so they are compatible with the witness's unique mental representation of the crime rather than asking all witnesses questions in a standardized format)—was found to increase memory quantity without reducing memory accuracy (see Fisher, Milne, & Bull, 2011). These examples demonstrate cognitively-mediated effects on eyewitness reports in the sense that they influence memory retrieval per se. Other studies have demonstrated metacognitively-mediated effects on eyewitness memory performance. For example, it was shown that inducing higher accuracy motivation using implicit or explicit payoffs leads rememberers to set a more strict criterion for reporting, resulting in more accurate memory reports (e.g., Koriat & Goldsmith, 1994, 1996). However, the increase in accuracy usually comes at a reduction in the quantity of correct reported information, in what is known as the *quantity-accuracy tradeoff*.

Author Note

This paper is based on the master's thesis submitted by the first author to the Department of Psychology at the University of Haifa.

The research was supported by THE ISRAEL SCIENCE FOUNDATION (grant 819/13) awarded to Ainat Pansky.

☆ This paper was handled by the former editorial team of JARMAC.

* Correspondence concerning this article should be addressed to Ainat Pansky, Department of Psychology, University of Haifa, 199 Aba Khoushy Ave., Haifa 3498838, Israel. Tel.: +972 4 8249437; fax: +972 4 8249431.

Contact: pansky@research.haifa.ac.il.

In the present study, we focused on another potential metacognitively-mediated effect of eyewitness questioning that may affect the quantity and/or accuracy of eyewitness reports: initial question difficulty. Specifically, we examined the effect of initial question difficulty on free-report eyewitness memory performance via the metacognitive processes of *monitoring* and *control*. With regard to monitoring, several studies have highlighted the importance of *retrieval fluency*—the ease with which information comes to mind when retrieving it—as a metacognitive cue that is used to evaluate the accuracy of the retrieved information by heuristically providing the rememberer with a sense of familiarity (e.g., Benjamin, Bjork, & Schwartz, 1998; for reviews, see Benjamin & Bjork, 1996; Kelley & Rhodes, 2002). For example, Shaw and McClure (1996) found higher confidence in the accuracy of items that were repeatedly questioned about than of those that were not (with no difference in accuracy). Similarly, Bonham and González-Vallejo (2009) found higher confidence in answers to postevent questions after reading a related (as opposed to a general) narrative, presumably as a result of the retrieval fluency emanating from this additional exposure.

In this study, we examined the role of relative rather than absolute retrieval fluency. Several studies that have manipulated fluency have shown that it is the discrepancy between experienced and expected levels of fluency (i.e., relative fluency), rather than its absolute level, that is critical in influencing memory judgments and other cognitive evaluations (e.g., Whittlesea & Leboe, 2003; Whittlesea & Williams, 1998; see also Hansen & Wänke, 2008; McCabe & Balota, 2007). The significant impact of relative fluency was attributed to the high salience of such a deviation from what was expected, and to its role in signaling about environmental changes that call for a new appraisal of the situation and may require a change of strategy (Wänke & Hansen, 2015). The contribution of relative retrieval fluency to subjective confidence gains further importance when considering the potential influence of subjective confidence on one's controlled behavior (e.g., Alter & Oppenheimer, 2009; Nelson & Narens, 1990). In eyewitness questioning, such influences may be observed because people heavily base their controlled volunteering decisions on their subjective confidence (Goldsmith & Koriat, 2008). Thus, when people are confident that they correctly know the answer to a question, they will generally provide it; otherwise, they will usually respond "don't know" (e.g., Koriat & Goldsmith, 1996).

Koriat and Goldsmith (1996) and Goldsmith and Koriat (2008) have developed a framework for studying the metacognitive processes that mediate between the retrieval of information and actual free-report performance: Upon the retrieval of a "best-candidate" answer, a monitoring process operates in order to subjectively assess its correctness (i.e., confidence). Based on the monitoring output, a control mechanism determines whether this answer should be reported or withheld by setting a *report criterion* to which the confidence in the answer is compared. The "best-candidate" answer is volunteered if the assessed probability correct passes the criterion; otherwise, it is withheld.

Consistent with Koriat and Goldsmith's (1996) framework, an effect on controlled behavior via an effect on monitoring was demonstrated by Hanczakowski, Pasek, Zawadzka, and Mazzoni (2013), who manipulated cue familiarity as a factor known to affect metacognitive monitoring at retrieval. They found that presenting primed rather than unprimed cues on a recognition test increased the participants' tendency to believe that they knew the correct answer, increased confidence for both correct and incorrect candidate responses, and consequently, increased the tendency to venture an answer instead of withholding it.

In the present study, we applied Koriat and Goldsmith's (1996) framework to an eyewitness situation, in which the rememberer is typically free to choose which information to report from memory (Goldsmith, Pansky, & Koriat, 2014). According to the framework, given that eyewitnesses try to provide as much accurate information as possible, they must subjectively evaluate the probability that the information that comes to mind is in fact correct—an evaluation which will eventually affect their decision of whether or not to report it. This evaluation might be biased by the relative retrieval fluency heuristic described above, via a *metacognitive contrast effect* (cf. Hansen & Wänke, 2008), by which the experienced ease (or difficulty) of retrieving the answer to a current question is based on an implicit comparison to the ease of answering the preceding questions. Namely, if answering the current question is experienced as relatively difficult compared to the experienced ease of answering the preceding questions, relative retrieval fluency should be quite low, resulting in fairly low confidence in the answer, making it less likely to be volunteered under free-report conditions. Conversely, if answering the current question is experienced as relatively easy compared to the experienced difficulty of answering the preceding questions, relative retrieval fluency should be higher, resulting in higher confidence in the answer. As mentioned, higher subjective confidence is more likely to lead to volunteering an answer, ultimately affecting eyewitness free-report memory performance.

Findings obtained by Bodner and Richardson-Champion (2007) seem to be consistent with an interpretation of a metacognitive contrast effect. They found higher recognition rates and more "remember" (rather than "know") judgments for medium-difficulty details from a crime film following a block of difficult-to-retrieve details than following a block of easy-to-retrieve details. However, because the data they report are combined for both old and new items, they do not speak to the effect of initial difficulty on actual memory performance in terms of quantity and/or accuracy. Furthermore, as the authors correctly note (p. 725), it is not possible to determine on the basis of their data whether initial difficulty affected the participants' recollection of the subsequent items (i.e., memory discrimination) or their decision-making process (i.e., response bias).

In a recent study, Pansky and Goldsmith (2014) examined the effects of initial question difficulty on both subjective experience and actual free-report memory (quantity and accuracy) performance on a multiple-choice general-knowledge test. Using a variation of Koriat and Goldsmith's (1996) research methodology, they demonstrated effects of initial question difficulty on free-report performance via its effects on metacognitive monitoring and control, alongside comparable forced-report performance. Specifically, the participants in an initially-difficult

group were more confident in their answers to the subsequent target questions than the participants in an initially-easy group. The effect of initial question difficulty on subjective confidence was translated into an overt effect on controlled behavior, with a higher tendency to volunteer an intermediate-difficulty answer after initially answering difficult questions than after initially answering easy questions. The effect of initial difficulty on confidence, and consequently, on volunteering rate, resulted in higher free-report memory quantity in the initially-difficult group, such that a larger quantity of correct answers was freely reported after answering initially-difficult questions than after answering initially-easy questions. Free-report memory accuracy, or the proportion of correct answers among those that were volunteered, was not affected by initial difficulty. Finally, Pansky and Goldsmith showed that the estimated report criterion was not affected by initial difficulty, supporting the interpretation that the obtained differences in volunteering rate and free-report quantity (of correct answers) stemmed primarily from an effect of initial question difficulty on control, mediated by monitoring (i.e., subjective confidence)—an embodiment of a metacognitive contrast effect—rather than from a direct effect on the placement of the report criterion.

The Present Study

In the present study, we examined the extent to which the findings of Pansky and Goldsmith (2014) would generalize to an episodic-eyewitness memory situation entailing open-ended questioning. More specifically, we tested the effect of initial question difficulty on eyewitnesses' free-report memory reporting of event details via the operation of monitoring and control processes. Toward this aim, we used Pansky and Goldsmith's (2014) adaptation of Koriat and Goldsmith's (1996) research methodology.

One week after watching a crime film (i.e., the target event), the participants were tested on event details using 30 open-ended questions: Twenty items of intermediate difficulty served as the target items, and ten preceding difficult- or easy-to-retrieve items were used to manipulate the initial level of retrieval fluency. Thus, one group of participants first answered relatively difficult questions before answering the target questions, whereas the other group first answered relatively easy questions.

After answering a question, the participants provided a confidence judgment assessing the likelihood that their answer was correct, followed by a decision whether or not they chose to volunteer it (for a monetary bonus if correct and a penalty if wrong). We expected that the initial-difficulty manipulation would not affect the participants' forced-report memory performance on the intermediate-level target questions, as there was no reason to expect that the difficulty of previous questions should affect the participants' ability to retrieve the correct answer to a given question (see also Pansky & Goldsmith, 2014). However, by a way of a metacognitive contrast effect, we expected that answering intermediate-difficulty questions would be experienced as relatively more fluent after initially answering the difficult than the easy questions, consequently leading to higher confidence in the answers, and finally, to a higher tendency to volunteer them. This, in turn, was expected to affect free-report memory performance.

To verify that the explicit reporting of confidence judgments did not bias any obtained effects of initial question difficulty, only half of the participants in each group were requested to report confidence judgments, and we examined whether this report condition (i.e., whether or not confidence judgments were explicitly solicited) interacted with initial question difficulty in affecting any of the dependent measures.

Methods

Participants

Eighty undergraduates from the University of Haifa participated for course credit or a payment of 60 shekels (approximately 15$US). They were randomly and equally assigned to each of the four initial difficulty/report condition groups.

Materials

The target event was a film depicting a robbery of a gas station.[1] For the compilation of the memory tests, 65 open-ended cued-recall questions (in Hebrew) were prepared about various details from the film. Then, following Koriat (1995), question difficulty was assessed based on the proportion of participants who freely provided an answer to that question in a preliminary test ($n = 10$): The two sets of 10 questions with the lowest and highest tendency to be freely answered (mean volunteering rate = .03 and .81, respectively) were selected for the initially-difficult and initially-easy condition, respectively, and the 20 questions in the center of the distribution (mean volunteering rate = .41) were selected as the target questions of intermediate difficulty. Consequently, two versions of the memory test were compiled, each containing the same 20 intermediate-difficulty target questions, preceded by either the 10 difficult or the 10 easy questions (see Appendix), for a total of 30 consecutive questions in each version. The order within each set of questions was randomized for each participant.

Procedure

In the first of two sessions,[2] the participants viewed the film, followed by a 10-min non-verbal distraction task. A week later, in the second session, the participants answered either the initially-difficult or the initially-easy test version. Koriat and Goldsmith's (1996) paradigm was applied to the memory test on an item-by-item basis for each question. In the initial forced-report phase, the participants were required to provide the most accurate and precise answer that they could, based on their memory for the event they had viewed. They were asked to provide

[1] This 12-minute film, called "OCTANE-96", was prepared by students in the "*Sam Spiegel Film & Television School, Jerusalem*", and permission to use the film for research purposes was granted by the school.

[2] Both sessions were run by a computer program developed using the E-Prime experiment-generating software.

a substantial (non-blank) answer to each and every question. In the second phase, they were required to provide a confidence judgment between 0% and 100%, reflecting the likelihood that the answer they had provided was correct. Finally, in the free-report phase, the participants decided whether or not to volunteer their answer. Volunteering accurate responses was induced by a moderate-incentive payoff schedule of a 1-point bonus for each correct volunteered answer and a 1-point penalty for each incorrect volunteered answer. The participants were informed that they would not be penalized (but neither would they receive any bonus) for withheld responses.[3] This three-phase procedure was repeated for each question. Within each initial-difficulty group, half of the participants performed this procedure without providing confidence judgments.

Results

Two independent judges determined the correctness of each reported answer. An answer to a question was judged as correct only if it matched the answer/s provided by three research assistants while they were viewing the film. The classifications made by these two judges were identical in 97% of the cases; a third judge determined the correctness of the controversial 3% of the responses. To facilitate the comparison between the proportion correct and the assessed probability correct (i.e., subjective confidence), we divided the confidence ratings by 100, converting them to a scale between 0 and 1.

To estimate each participant's report criterion, we adopted the computational procedure developed by Koriat and Goldsmith (1996). Each confidence level (between 1% and 100%) was evaluated as a potential report criterion by summing the number of reported answers with an equal or higher confidence rating and the number of withheld answers with a lower confidence rating. The potential report criterion with the highest proportion of such (correctly predicted) report decisions was chosen as the report criterion estimate. If a range of values yielded an equally good fit, the average of these estimates was chosen as the estimated report criterion.

As a manipulation check, the difficulty of the two sets of initial questions was compared. As expected, the proportion of correct responses was much lower for the questions assigned to the initially-difficult condition (.18) than for those assigned to the initially-easy condition (.72), $t(78) = 15.10$, $p < .001$, $d = 3.42$. The initially-difficult questions were also subjectively experienced as more difficult than the initially-easy questions, with substantially lower confidence judgments associated with the answers to the former (.31) than with the answers to the latter (.66), $t(38) = 7.13$, $p < .001$, $d = 2.31$. Finally, answers to the initially-difficult questions were less likely to be freely volunteered (.25) than answers to the initially-easy questions (.66), $t(78) = 8.35$, $p < .001$, $d = 1.89$.

Next, we examined the response time of providing an answer as an independent measure of fluency. Because this response time reflected both the time required to retrieve the answer and the time required to type it in the response window, we first compared the mean response length (in characters) between the two sets of initial questions to see whether this factor would have to be controlled for. Indeed, the responses to the initially-easy questions were lengthier (11.92 characters), on average, than the responses to the initially-difficult questions (8.04 characters), $t(78) = 4.49$, $p < .001$, $d = 1.02$. When controlling for this difference by entering response length as a covariate, shorter response times were found for the initially-easy questions (21.73 s) than for the initially-difficult questions (26.65 s), $F(1,75) = 4.52$, $p = .037$, $\eta_p^2 = .057$, confirming that the former were processed more fluently, as could be expected. Regarding the intermediate-difficulty target questions, although the response times were somewhat shorter in the initially-difficult group (14.05 s) than in the initially-easy group (15.79 s), the difference in absolute fluency was not significant, $F(1,76) = 2.05$, $p = .157$, $\eta_p^2 = .026$. However, the difference in response times between the target questions and the initial questions (again, with initial response length serving as a covariate) was significantly larger in the initially-difficult group (−11.61 s) than in the initially-easy group (−6.94 s), $F(1,75) = 4.52$, $p = .037$, $\eta_p^2 = .057$, consistent with our prediction of higher relative fluency in answering intermediate-difficulty questions after answering initially-difficult questions than after answering initially-easy questions.[4]

To test the effect of initial question difficulty on subsequent responding to the intermediate-difficulty target questions, an ANOVA was conducted on each of the dependent measures: (1) forced-report proportion correct, (2) confidence, (3) volunteering rate (irrespective of accuracy), (4) free-report memory quantity—the proportion of target items that were correctly and freely reported, and (5) free-report memory accuracy—the proportion of correct answers among those that were freely reported. Initial question difficulty and report condition (whether or not confidence judgments were solicited) served as independent variables. As report condition did not interact with initial difficulty in any of the analyses, the data for the two report conditions were pooled. Despite our instruction to provide substantial answers to all of the questions in the forced-report phase, 7.6% of the responses to the target questions in the initially-difficult group and 8.1% of those in the initially-easy group were "don't remember" responses, $t(78) = .16$, $p = .875$, $d = 0.04$. As the same pattern of results was found whether or not these "don't remember" responses were included in the analyses, the reported results are based on the entire set of responses. The means and standard deviations of the dependent measures, calculated for the target questions in each initial-difficulty group, are presented in Table 1.

First, as expected, initial question difficulty did not affect the ability to answer the intermediate-difficulty target questions,

[3] Upon completion of the entire experiment, the points were calculated and translated to monetary bonuses, with each bonus point worth one shekel (approximately .25$US).

[4] One should note that these differences between the response times to the target questions and the response times to the initial questions were negative in both groups due to a general reduction in response times from trial to trial, probably reflecting the participants' increasing familiarity with the task of retrieving answers and typing them.

Table 1
Performance Measures and Metacognitive Indices for the Intermediate-Difficulty Target Questions, Following an Initial Set of Easy Versus Difficult Questions. Standard Deviations Appear in Parentheses

Initial difficulty	Forced-report proportion correct	Confidence	Volunteering rate	Free-report memory quantity	Free-report memory accuracy	Report criterion
Easy	.39 (.11)	.44 (.14)	.37 (.20)	.21 (.10)	.60 (.27)	.62 (.25)
Difficult	.41 (.11)	.51 (.21)	.51 (.26)	.28 (.13)	.60 (.20)	.57 (.22)

with comparable forced-report proportion correct for the participants who had initially answered difficult questions (.41) and for those who had initially answered easy questions (.39), $F(1,76) = 0.45$, $p = .503$, $\eta_p^2 = .006$.

Second, as predicted, the subjective confidence associated with the answers was higher for the initially-difficult group (.51) than for the initially-easy group (.44), but this difference was not statistically significant, $t(38) = 1.27$, $p = .213$, $d = 0.41$. Nonetheless, the volunteering rate for the target questions was substantially higher after initially answering difficult questions (.51) than after initially answering easy questions (.37), $F(1,76) = 7.02$, $p = .010$, $\eta_p^2 = .085$. Furthermore, we observed no significant difference in the estimated report criterion between the initially-difficult group (.57) and the initially-easy group (.62), $t(38) = 0.64$, $p = .527$, $d = 0.21$ (fit rate: 94%), suggesting that the effect of initial difficulty on volunteering rate was not mediated by report criterion. Following Pansky and Goldsmith (2014), we examined whether the effect of initial question difficulty on confidence might have been expressed in terms of the proportion of answers accompanied by medium-to-high confidence judgments, presumably constituting those answers that the participants actually considered as plausible candidates for volunteering. We defined medium-to-high confidence answers as those with an assessed probability correct higher than one standard deviation below the mean estimated report criterion (assessed probability correct >.36). Consistent with the predicted effect of initial question difficulty on subjective confidence, and, in turn, with its observed effect on volunteering rate, the proportion of medium-to-high confidence answers was substantially higher in the initially-difficult group (.66) than in the initially-easy group (.51), $t(38) = 1.96$, $p = .029$, $d = 0.63$ (one-tailed).

Third, we examined the extent to which the effect of initial question difficulty on confidence, and consequently on volunteering rate, also affected free-report memory performance for the target questions. Indeed, initial question difficulty affected the quantity of freely-reported correct information: The proportion of correct volunteered answers (out of the total number of questions) was higher for the initially-difficult group (.28) than for the initially-easy group (.21), $F(1,76) = 7.41$, $p = .008$, $\eta_p^2 = .089$. However, free-report accuracy was not affected by initial question difficulty, with identical proportions of correct answers out of the total number of volunteered answers in the two groups (.60), $F(1,76) = 0.00$, $p = .966$, $\eta_p^2 = .000$.

How can we account for the finding that the lower free-report quantity in the initially-easy group was not accompanied by higher free-report accuracy, as would be expected according to the Koriat and Goldsmith (1996) model? One possibility could be that this was a result of an (unexpected) poorer ability to discriminate between correct and incorrect answers in the initially-easy group than in the initially-difficult group (i.e., lower monitoring resolution in the former group). However, this was not the case, as subjective confidence was significantly higher for correct answers (0.61) than for incorrect answers (0.38) across both groups, $F(1,38) = 62.00$, $p < .001$, $\eta_p^2 = .620$, and this difference in confidence (i.e., discrimination index; Schraw, 2009) was comparable for the initially-difficult group (0.23) and the initially-easy group (0.22), $F(1,38) = .06$, $p = .803$, $\eta_p^2 = .002$, as would be expected. Another possibility is that the lack of gain in accuracy in the face of the decreased memory quantity of the initially-easy group was due to deficient monitoring in the medium-confidence range, around the report criterion, constituting those answers that the participants considered as plausible (but not definite) candidates for volunteering. We defined medium-confidence answers as those with confidence ranging between .50 and .80, based on the data indicating that only a small minority (10%) of the answers with confidence lower than .50 and nearly all (95%) of the answers with confidence higher than .80 were volunteered by the participants in both groups [the group differences were non-significant, $t(37) = 0.49$, $p = .628$, $d = 0.16$, and $t(37) = 0.46$, $p = .648$, $d = 0.15$, respectively]. An examination of the monitoring resolution in the medium-confidence range revealed that the confidence judgments in this range did not significantly discriminate between correct answers (0.63) and incorrect answers (0.60), $F(1,33) = 2.08$, $p = .158$, $\eta_p^2 = .059$, with a non-significant interaction with group, $F(1,33) = .651$, $p = .425$, $\eta_p^2 = .019$. According to the Koriat and Goldsmith framework, when "resolution is completely lacking, the exercise of control could simply reduce quantity performance with no gain in accuracy at all" (Koriat & Goldsmith, 1996, p. 508). It is possible that due to this deficient monitoring, the volunteering of fewer answers by the participants in the initially-easy group than by those in the initially-difficult group merely reduced memory quantity without increasing accuracy.

As shown in Figure 1, the option of free-report allowed both initial-difficulty groups to achieve comparable and significant gains in accuracy (of .20, on average) relative to the forced-report condition, $F(1,78) = 73.52$, $p < .001$, $\eta_p^2 = .485$, with a non-significant interaction between report option (forced, free) and initial difficulty, $F(1,78) = 0.11$, $p = .738$, $\eta_p^2 = .001$. However, as also shown in Figure 1, the gain in accuracy with the exercise of free-report came at a cost in the quantity of correct information, $F(1,78) = 180.75$, $p < .001$, $\eta_p^2 = .699$, evidencing a quantity–accuracy tradeoff. This cost was smaller for the initially-difficult group (.13) than for the initially-easy group (.19), with a significant interaction between report option and

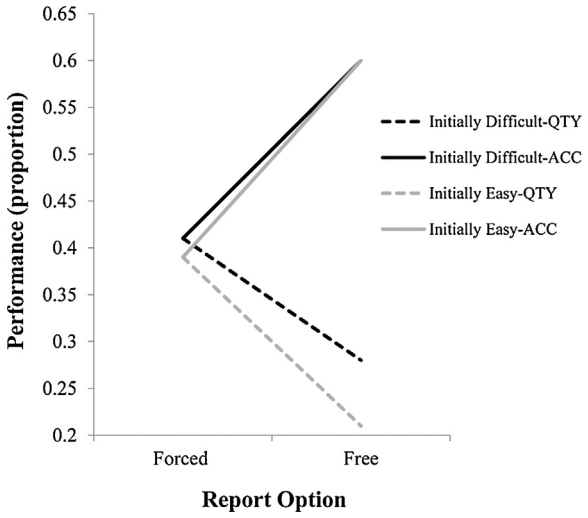

Figure 1. Memory quantity (QTY: the proportion of questions that were answered correctly) and memory accuracy (ACC: the proportion of reported answers that were correct) under forced versus free memory reporting, following an initial set of easy versus difficult questions. Note that quantity and accuracy measures can be distinguished operationally only under conditions of free report, whereas these measures are operationally equivalent under forced-report conditions, in which all questions must be answered.

initial difficulty, $F(1,78) = 5.06$, $p = .027$, $\eta_p^2 = .061$. That is, the initially-difficult group exhibited a milder quantity-accuracy tradeoff than did the initially-easy group.

Discussion

In the present study, we attempted to illuminate the manner in which relative retrieval fluency may be translated, via the operation of monitoring and control processes, into an explicit effect on controlled behavior—the decision of whether or not to provide an answer to a question, thereby affecting free-report memory performance. Toward this end, we manipulated the difficulty of the initial questions in an open-ended questionnaire about a crime film.

To summarize our major findings, the proportion of medium-to-high confidence answers (i.e., answers that were likely to be considered as plausible candidates for volunteering) was higher after initially answering difficult questions than after answering easy questions. Consequently, compared to the participants in the initially-easy group, the participants in the initially-difficult group volunteered a larger proportion of their answers to the target questions under free-report conditions. Finally, the effect of initial question difficulty on volunteering rate, via its effect on confidence, also affected free-report memory quantity, with a larger quantity of correct freely-reported answers in the initially-difficult than in the initially-easy group.

Effects of relative retrieval fluency on subjective experience and on behavior

In contrast to previous studies that have highlighted the effects of absolute retrieval fluency on subjective confidence (e.g., Bonham & González-Vallejo, 2009; Shaw & McClure, 1996), the present study was focused on the effects of relative retrieval fluency on subjective confidence, and, consequently, on behavior. First, we validated our assumption that answering the medium-difficulty questions would be relatively more fluent after initially answering difficult questions than after answering easy questions. Using response times as a measure of fluency, we confirmed that, although absolute fluency in answering the target questions was comparable regardless of the difficulty of the initial questions, relative fluency (estimated as the difference between the response times to the target questions and those to the initial questions) was higher when answering the target questions after initially answering difficult questions than after initially answering easy questions. Second, replicating the findings of Pansky and Goldsmith (2014), we found that initial question difficulty, via its effect on relative fluency, affected subjective confidence: More answers to the target questions were accompanied by medium-to-high confidence judgments in the initially-difficult group than in the initially-easy group. These findings are consistent with studies that have shown that the discrepancy between experienced and expected levels of fluency (i.e., relative fluency) is critical in influencing memory judgments and other cognitive evaluations (e.g., Hansen & Wänke, 2008; McCabe & Balota, 2007; Whittlesea & Leboe, 2003; Whittlesea & Williams, 1998).

Finally, the effect of initial difficulty on relative fluency and confidence also translated into an effect on one's controlled behavior, in terms of a larger tendency to freely report answers after initially answering difficult questions, as in Pansky and Goldsmith's (2014) study. This finding is consistent with several theoretical approaches (e.g., Alter & Oppenheimer, 2009; Nelson & Narens, 1990), and, importantly, with Koriat and Goldsmith's (1996) and Goldsmith and Koriat (2008) model of the strategic regulation of memory accuracy, and the empirical findings supporting it. A similar effect on controlled behavior via an effect on monitoring was also demonstrated by Hanczakowski et al. (2013), who found a stronger feeling of knowing the correct answers to test questions when the presented cues were previously primed, eventually resulting in a higher tendency to freely report such answers.

In terms of free-report memory performance, the higher volunteering rate in our initially-difficult group resulted in higher free-report memory quantity, compared to that of the initially-easy group. According to Koriat and Goldsmith's (1996) framework, given the comparable forced-report proportion correct, monitoring effectiveness, and report criterion in the two groups, this higher free-report quantity in the initially-difficult group would have been expected to come at a cost in accuracy. However, free-report accuracy was identical in the two groups. We attempted to account for this finding by examining the monitoring resolution in the medium-confidence range, around the report criterion, constituting those answers that the participants considered as plausible (but not definite) candidates for volunteering. Interestingly, monitoring resolution in this range was completely lacking, in which case, according to Koriat and Goldsmith's (1996) framework, the exercise of control would be expected to merely reduce quantity performance with no gain in accuracy. Thus, the deficient monitoring in the critical range of medium-confidence answers could have been

responsible for the obtained pattern of findings that the lower volunteering rate in the initially-easy group merely reduced free-report quantity, compared to the initially-difficult group, without improving accuracy.

The metacognitive-contrast account

In this study, we examined the generalizability of Pansky and Goldsmith's (2014) results to the domain of eyewitness memory entailing open-ended questioning. Their findings for general-knowledge questions supported the metacognitive-contrast interpretation of the effect of initial difficulty on confidence, volunteering rate, and free-report performance. According to their account of these effects, the experienced ease of retrieving the answers to the target questions was implicitly compared with the ease of answering the preceding questions: When attempts to answer the current questions were experienced as relatively easy compared to the experienced difficulty of answering the initial questions, higher relative fluency was experienced, resulting in higher confidence in the answers to the target questions. In accordance with Koriat and Goldsmith's (1996) model, these higher levels of subjective confidence in the retrieved answers (via retrieval fluency), in turn, increased the tendency to volunteer them, thereby increasing the quantity of freely-reported correct information.

Clearly, our findings for eyewitness memory resemble those obtained by Pansky and Goldsmith (2014) for semantic memory, and, in fact, completely parallel those found in their second experiment. We also replicated their finding that the report criterion was unaffected by initial question difficulty, supporting the interpretation by which the obtained differences between both difficulty groups in volunteering rate and free-report quantity (of correct answers) stemmed primarily from an effect of initial question difficulty on control that was mediated by monitoring (i.e., confidence). Thus, collectively, the results of the two studies provide support for a metacognitive contrast effect, such that when preceded by a set of difficult rather than easy questions, answering the intermediate-difficulty target questions was experienced as relatively more fluent, resulting in a higher proportion of medium-to-high confidence answers, and consequently, in a higher tendency to volunteer the answers to the target questions. Finally, such metacognitive monitoring and control processes affected free-report memory performance independently of forced-report performance.

Additional findings that may be interpreted as supporting a metacognitive contrast effect are those of Bodner and Richardson-Champion (2007), demonstrating that medium-difficulty details from a crime film were more likely to be classified as "remembered" (versus "known") when tested after difficult-to-retrieve than after easy-to-retrieve details. Actually, the findings obtained in the present study can be seen as an extension of theirs, with the additional examination of the effects of initial question difficulty on free-report quantity (of correct answers) and free-report accuracy.

Importantly, the higher volunteering rate and free-report quantity of correct answers that we found after the participants had first answered difficult questions were obtained whether or not confidence judgments were explicitly solicited. This enables us to rule out a demand characteristic interpretation that would attribute our results to an artificially high correlation between the reported confidence judgments and the volunteering decision. Nevertheless, one could claim that the higher volunteering rate obtained within the initially-difficult group was a way by which this group's participants could compensate for volunteering fewer answers in the initial stage. An alternative explanation is that, due to the greater difficulty of the initially-difficult questions, the participants in that group may have set a more liberal report criterion to begin with, which remained stable even after the questions became easier. Although the lack of a significant difference between the two initial-difficulty groups in the report criterion used for the intermediate-difficulty items seems to be inconsistent with these two alternative accounts of our findings, it is nonetheless possible that the confidence ratings used to estimate the report criterion might themselves have been biased by criterion placement (see Higham, Zawadzka, & Hanczakowski, 2016).

Effects of relative difficulty

In the present study, the relative difficulty of the questions was manipulated in a blocked design. An interesting question for future research is whether item-by-item changes in difficulty would also affect free-report memory performance via metacognitive processes of monitoring and control. Previous studies that examined the effect of item-by-item changes in difficulty on subjective ratings of performance have yielded a somewhat mixed pattern of results. For example, Bodner and Richardson-Champion (2007) found that intermediate-difficulty details were more likely to be classified as "remembered" when they were mixed with difficult-to-remember details than with easy-to-remember details, suggesting that the subjective experience of remembering can be influenced by item-by-item changes in difficulty. On the other hand, Weinstein and Roediger (2010, 2012) found that the order of questions on a general-knowledge test with regard to question difficulty—whether randomized, sorted from the easiest to the hardest, or vice versa—did not affect item-by-item confidence ratings. It is possible that changes in difficulty did not affect the item-by-item judgments in Weinstein and Roediger's (2010, 2012) studies due to the gradual nature of the changes in difficulty there, in contrast to the sharper changes in difficulty that were found to affect the item-by-item judgments in Bodner and Richardson-Champion's (2007) study. Based on these findings, item-by-item changes in difficulty may be expected to have an effect on free-report memory performance, via their effect on metacognitive monitoring and control processes (Koriat & Goldsmith, 1996), if these changes are sufficiently substantial to yield a contrast effect.

Nonetheless, the order of questions on a memory test has been consistently shown to affect retrospective *global* evaluations of performance: Irrespective of their actual performance, test-takers believed they had answered more questions correctly when the questions were sorted from the easiest to the hardest than when randomized or sorted from the hardest to the easiest (e.g., Jackson & Greene, 2014; Weinstein & Roediger, 2010,

2012). These findings were taken to support the notion that initial impressions of difficulty anchor or predominate in overall perception of test difficulty. In a similar vein, answering an initial set of easy or difficult questions in the present study may have anchored the expected level of difficulty for the remainder of the test, an expectation that may have served as the basis for the observed contrast effects.

Practical Application

From an applied perspective, our results emphasize the importance of taking into account potential contrast effects when developing eyewitness questioning procedures. The present findings suggest that the amount and/or accuracy of information that is freely reported by an eyewitness in the course of an interview can be affected by the difficulty of the preceding questions. More specifically, preceding questions about a witnessed event by relatively difficult as opposed to relatively easy questions can be expected to yield more (correct) information from the eyewitness. In the present study (see also Pansky & Goldsmith, 2014), the increase in free-report memory quantity after answering difficult questions came at no cost in accuracy. However, more research is needed to establish to what extent, and under which conditions, this particular outcome is characteristic of eyewitness questioning situations.

More generally, the present findings contribute to the understanding of the many factors that can affect the completeness and reliability of eyewitness memory reports. In addition to various aspects of eyewitness interviewing that have been shown to influence free-report eyewitness memory performance via cognitively-mediated effects (e.g., Fisher et al., 2011; Loftus & Zanni, 1975), the present study demonstrates the metacognitively-mediated effects of initial question difficulty.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Appendix. Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jarmac.2016.04.007.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*, 219–235.

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

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

Bodner, G. E., & Richardson-Champion, D. D. L. (2007). Remembering is in the details: Effects of test-list context on memory for an event. *Memory*, *15*, 718–729.

Bonham, A. J., & González-Vallejo, C. (2009). Assessment of calibration for reconstructed eye-witness memories. *Acta Psychologica*, *131*, 34–52.

Fisher, R. P., Milne, R., & Bull, R. (2011). Interviewing cooperative witnesses. *Current Directions in Psychological Science*, *20*, 16–19.

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin, & B. H. Ross (Eds.), *The psychology of learning and motivation* (48) (pp. 1–60). London: Academic Press.

Goldsmith, M., Pansky, A., & Koriat, A. (2014). Metacognitive control of memory reporting. In T. J. Perfect, & D. S. Lindsay (Eds.), *Sage handbook of applied memory* (pp. 481–500). London, UK: Sage.

Hanczakowski, M., Pasek, T., Zawadzka, K., & Mazzoni, G. (2013). Cue familiarity and 'don't know' responding in episodic memory tasks. *Journal of Memory and Language*, *69*, 368–383.

Hansen, J., & Wänke, M. (2008). It's the difference that counts: Expectancy/experience discrepancy moderates the use of ease of retrieval in attitude judgments. *Social Cognition*, *26*, 447–468.

Higham, P., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. In J. Dunlosky, & S. Tauber (Eds.), *Oxford handbook of metamemory*. Oxford, UK: Oxford University Press.

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.

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. *Psychology of Learning and Motivation: Advances in Research and Theory*, *41*, 293–320.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*, 311–333.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297–315.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517.

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, *5*, 86–88.

McCabe, D. P., & Balota, D. A. (2007). Context effects on remembering and knowing: The expectancy heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 536–549.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (26) (pp. 125–173). San Diego, CA: Academic Press.

Pansky, A., & Goldsmith, M. (2014). Metacognitive effects of initial question difficulty on subsequent memory performance. *Psychonomic Bulletin & Review*, *21*, 1255–1262.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness recall and testimony. In K. D. Williams, & N. Brewer (Eds.), *Psychology and law: An empirical perspective* (pp. 93–150). New York: Guilford.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*, 33–45.

Shaw, J. S., & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, *20*, 629–653.

Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science*, *24*, 195–199.

Weinstein, Y., & 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.

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

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, *49*, 62–79.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica*, *98*, 141–165.

Received 13 December 2015;
received in revised form 13 April 2016;
accepted 15 April 2016
Available online 27 May 2016