



Effects on Memory of Early Testing and Accuracy Assessment for Central and Contextual Content

Jessica S Wasserman, Cody W Polack, Crystal Casado, Maïté Brunel, Mohamad El Haj, Ralph R Miller

► To cite this version:

Jessica S Wasserman, Cody W Polack, Crystal Casado, Maïté Brunel, Mohamad El Haj, et al.. Effects on Memory of Early Testing and Accuracy Assessment for Central and Contextual Content. *Journal of Cognitive Psychology*, 2020, 32 (7), pp.598-614. 10.1080/20445911.2020.1809433 . hal-03662388

HAL Id: hal-03662388

<https://hal.science/hal-03662388>

Submitted on 28 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Published in final edited form as:

J Cogn Psychol (Hove). 2020 ; 32(7): 598–614. doi:10.1080/20445911.2020.1809433.

Effects on Memory of Early Testing and Accuracy Assessment for Central and Contextual Content

Jessica S. Wasserman¹, Cody W. Polack¹, Crystal Casado¹, Maïte Brune^{1,2}, Mohamad El Haj³, Ralph R. Miller¹

¹State University of New York at Binghamton (USA);

²Univ. Lille, EA 4072 – PSITEC – Psychologie: Interactions, Temps, Emotions, Cognition, F-59000 Lille (France);

³Nantes Université, Univ Angers, Laboratoire de Psychologie des Pays de la Loire (LPPL - EA 4638), F-44000 Nantes, France

Abstract

Memory for an event is influenced by many factors including retention interval, frequency of assessment, and type of information assessed concerning the event. We examined the usefulness of observer memory for contextual information in assessing accuracy of memory for central information. Participants viewed a video of a purse being stolen and were asked questions concerning the perpetrator and surrounding context of the event, including where and when the event occurred and who else was present. Participants tested immediately after seeing the video exhibited better memory than those tested for the first time 48-hour after the event. Additionally, testing immediately after viewing the video reduced forgetting over the 48-hour delay (i.e., early testing attenuated subsequent forgetting). Moreover, memory for the context of the event correlated positively with memory of the central information (i.e., perpetrator), and memory concerning other people at the event tended to have the highest correlation with perpetrator memory.

Keywords

context; preventing forgetting; central information; peripheral information; repeated testing

Accurately remembering past events can be a major concern in everyday life. This includes more trivial events, such as remembering where we put our keys, and more significant events, such as remembering the details of a criminal event. Research has examined many factors that influence the accuracy of such episodic memories. One such factor is the time

Corresponding author: Ralph R. Miller, Department of Psychology, SUNY – Binghamton, Binghamton, NY 13902-6000 USA, rmiller@binghamton.edu, Phone: (607) 777-2291.

Disclosure of Interest

The authors report no conflicts of interest.

Data Availability Statement

The video, raw data, data analyses, and computer program (in eprime3 titled 'EW Real Perception Experiment 1.esb2' and 'EW Real Perception Experiment 1.es2') are available at doi.org/10.22191/orb/rrmiller/lab/1 and upon request from the first (JW) and last (RM) authors.

that passes between when one observes an event and when one is asked about that event. Research has demonstrated that, in most instances, participants remember significantly less precise information after a delay (e.g., Jones & Pipe, 2002; Ornstein et al., 2006; Paterson, Eijkemans, & Kemp, 2015; Paz-Alonso & Goodman, 2008; Tuckey & Brewer, 2003). In contrast, Oberauer and Lewandowsky (2008) demonstrated that a longer delay may not always worsen memory; in select circumstances participants can exhibit superior memory following a longer delay compared to a shorter delay (Hicks, Marsh, & Russell, 2000). Specifically, Hicks et al. (2000) found that event-based prospective memory performance was enhanced over longer delays of 15 minutes, as compared to shorter delays of 2.5 or 5 minutes, and that performance improved when participants completed a greater number of intervening tasks during the retention interval. Hicks et al. (2000) suggest that a possible explanation for their results is that lengthening the retention interval and increasing the number of intervening tasks provides opportunities for participants to engage in more self-reminders of the original instructions. Having more time before the final task, and more down-time between intervening tasks, may allow for extended reflection on the original instructions.

Another factor that can play a role in memory accuracy is the number of times that a respondent is asked to retrieve information regarding a certain event. Conducting repeated tests or interviews may protect the memory of the event from significant forgetting, assuming that the tests or interviews do not introduce erroneous information. Research has demonstrated that being asked to remember information repeatedly can prevent some forgetting, especially when one is asked to remember that information soon after first learning information or observing an event, as compared to being asked about that information after a delay (Abel & Roediger, 2017; Gates, 1917; Pipe, Sutherland, Webster, Jones, & La Rooy, 2004; Poole & White, 1991; Wheeler & Roediger, 1992). Studies within the eyewitness literature have examined the influences of conducting multiple interviews after a delay between when an event occurred and when one is asked about the event. Flin, Boon, Knox, and Bull (1992) had participants view a staged presentation and questioned some of the participants about the event both on the day after the event and 5 months later. The other participants were only asked about the event 5 months later. All participants experienced some degree of forgetting during the delay. Yet, participants who were tested twice demonstrated a higher degree of recall accuracy on the test given 5 months later, as compared to participants who were tested for the first time after 5 months. Therefore, it appears that repeated tests can prevent some forgetting (Pipe et al., 2004; Poole & White, 1991; Quas et al., 2007; Scrivner & Safer, 1988; Turtle & Yuille, 1994). These results are in line with the testing effect, which is the phenomenon that one exhibits superior memory on subsequent tests when one is tested early after learning information or observing an event, as compared to not being tested early (Abel & Roediger, 2017; Gates, 1917; Wheeler & Roediger, 1992). Furthermore, being tested early often results in less forgetting as compared to just restudying information before a final test (Bäuml & Kliegl, 2017; Roediger & Butler, 2011; Rowland, 2014).

Initial testing, however, can sometimes result in a lower recall rate on later tests than restudying information, a phenomenon known as the negative testing effect (Mulligan & Peterson, 2015; Peterson & Mulligan, 2013). Other researchers have observed a negative

testing effect in a misinformation paradigm, in which participants are given misleading information after observing an event. These studies found that taking an early test before a later test can sometimes produce an increase in distortion in the later test. This result is referred to as 'retrieval enhanced suggestibility' (Chan & Langley, 2011; Chan, Manley, & Lang, 2017).

In addition to the influences of the repetition and timing of tests on memory, the degree to which people remember different types of information about an event can vary. This includes remembering information about the appearance of people and remembering information regarding actions or objects to different extents. Prior studies have evaluated forgetting of different types of information that participants report after observing an event, and found that participants typically demonstrate superior memory for details concerning objects and actions relative to the appearance of people (Akehurst, Milne, & Kohnken, 2003; Memon, Wark, Bull, & Kohnken, 1997). In contrast, other research has demonstrated that memory of social information about people and their interactions with one another is often superior to memory of non-social information (Mesoudi, Whiten, & Dunbar, 2006). To discern which types of information are best remembered after observing an event, many researchers have categorized information for an event as 'central' or 'peripheral'. Research has found that people tend to better remember 'central' information as compared to 'peripheral' information, for both emotional events (Burke, Heuer, & Reisberg, 1992; Christianson & Loftus, 1987) and non-emotional events (Ibabe & Sporer, 2004; Heath & Erickson, 1998). Although there are varying definitions of 'central' and 'peripheral' information, we will provide an example of the definitions used by Libkuman, Stabler, and Otani (2004): "Central detail was defined as the detail that was associated with the central characters, whereas background detail was defined as the detail that was not associated with the central characters" (p. 241).

Within this framework, the question can be asked whether accurate memory for the context of the event, that is, 'peripheral' information, can predict accuracy of memory for the 'central' information. Juries often use such measures to assess the credibility of witnesses with respect to the critical, or 'central', information (Bell & Loftus, 1988). Wells and Leippe (1981) examined the relationship between memory accuracy of 'peripheral' details and accuracy of the identification of a thief ('central' details). The authors observed that witnesses who accurately identified the thief were less accurate in reporting 'peripheral' details than witnesses who identified an innocent person. The authors suggest that participants who accurately identified the thief may have been devoting more attention to the thief, and therefore less attention to 'peripheral' details. This is in line with Easterbrook's (1959) attentional narrowing hypothesis, which suggests that people often devote more attentional resources to 'central' information, leaving fewer resources to attend to 'peripheral' information. Additionally, other participants who played the role of jurors were asked to state which group of witnesses they felt were more credible (e.g., those who accurately identified the thief but were less accurate in remembering 'peripheral' details, or those who were better able to remember 'peripheral' details, but did not accurately identify the thief). The participant-jurors were more likely to discredit the participants who were more accurate in identifying the thief and were less accurate regarding 'peripheral' details. The authors concluded that many of the participant-jurors assumed that there was a positive

correlation between remembering ‘central’ and ‘peripheral’ details, such that those who were better able to remember peripheral details should be better able to accurately remember ‘central’ details. Considering these results and Easterbrook’s attentional narrowing hypothesis, this assumption warrants further examination.

Many of the studies mentioned above aimed at assessing the effect of testing people who have observed an event using open-ended questions (Quas et al., 2007; Scrivner & Safer, 1988; Turtle & Yuille, 1994). Studies in this area rarely use forced-choice questions. Avoiding forced-choice questions is wholly appropriate given that best-practice witness interviewing protocols call for the use of open questions (e.g., cognitive interview, Fisher & Geiselman, 1992; NICHD protocol, Hershkowitz, Lamb, & Katz, 2014). Fisher, Geiselman, and Raymond (1987) examined police interviewing techniques used in the field, and reported the type of questions that police interviewers typically employ. Open questions were categorized as questions that gave witnesses an opportunity to provide many pieces of information, as compared to direct, short-answer questions that asked for a particular piece of information. For example, an open question might ask for a description of the suspect’s clothing, whereas a direct, short-answer question might ask for the color of the suspect’s shirt. Although using open questions is recommended, analysis of police officer interviewing techniques showed frequent use of leading questions, direct, short-answer questions, and closed questions (Clifford & George, 1996; Fisher et al., 1987; Ginet & Py, 2001), all of which can have deleterious consequences. Using direct, short-answer questions to ask for a particular piece of information typically limits a witness’ response to only the information that was specifically asked for. If an interviewer fails to directly inquire about other relevant information, information that might be important to the event may not be revealed (Fisher et al., 1987).

In addition to differentiating between open and direct, short-answer questions, it is useful to distinguish between forced-choice and recall procedures. Forced-choice procedures provide respondents specific answers to recognize and choose from, which is similar to the format of direct, short-answer questions in that respondents must provide a specific answer that is limited by the wording of the questions and allowable answers. In comparison, recall procedures ask respondents to supply recollections from memory without sharply focusing on a specific piece of information, which gives respondents the ability to provide multiple pieces of information. This is similar to the format of open questions. Research on witness memory often uses recall procedures (Quas et al., 2007; Scrivner & Safer, 1988; Turtle & Yuille, 1994; Yuille & Cutshall, 1986). However, despite the fact that police officers are generally aware of the detrimental effect of closed or forced-choice questions since the development of best-practice interviewing guidelines (e.g., the cognitive interview, Fisher & Geiselman, 1992; NICHD protocol, Hershkowitz et al., 2014), witnesses are sometimes interrogated by people other than professionals. Witnesses can be interviewed by other witnesses, by relatives, or by others in proximity (e.g., friends, coworkers) who are not necessarily conscious of the effects of different types of questions on memory. Moreover and unfortunately, witnesses can still be interviewed by police officers who use deleterious questioning style perhaps because of a lack of training, or because they simply do not use open-questions despite being instructed to do so. Indeed, training in non-directive

interviewing skill is complex, and difficult to maintain over time (Lamb, Sternberg, Orbach, Esplin, & Mitchell, 2002; Smith, Powell, & Lum 2009).

Considering all of the literature previously mentioned, we were interested in understanding the effects of early testing on memory for ‘central’ and ‘peripheral’ information in the less desirable testing format of forced-choice. Memory assessment measures that facilitated quantitative comparisons of performance across memory categories and across the delay favored the present forced-choice testing procedure, which differs substantially from the free recall favored in eyewitness studies. Thus, we have avoided using the term ‘eyewitness’ in regards to the present experiment. Additionally, interviewers may employ unfavorable question types that are similar to the format of forced-choice questions. Hence, it is important to examine the effects of testing people who have observed an event in less desirable conditions, such as forced-choice.

The first hypothesis assessed in the present experiment was whether observer forgetting can be reduced by early testing using a forced-choice procedure. Based on previous literature, we hypothesized that participants who were tested immediately after observing an event would show higher recognition scores on the early test than participants who were tested only after a delay. We predicted that those participants who were tested twice would benefit from the early testing, and would exhibit superior memory relative to participants who were tested only after a delay. That is, early testing was expected to retard forgetting. Although we had strong reason to expect a testing effect based on evidence from previous research, we deemed it important to examine whether we would actually observe this phenomena within the present design.

The second and more novel goal of the current experiment was to examine observer correlations of memory for ‘central’ information (e.g., regarding the perpetrator) with the various aspects of contextual (i.e., ‘peripheral’) memory, as accuracy of information about the context of an event is often used to assess the likely accuracy of memory for focal information (Bell & Loftus, 1988; Wells & Leippe, 1981). The aspects of contextual memory examined here included where and when the event took place, and who else was present at the event. As aforementioned, many researchers have compared accuracy of memory of ‘central’ and ‘peripheral’ information, but these studies did not differentiate among various types of ‘peripheral’ information. Our experimental design is unique in that it targeted specific contextual details, and allowed for a deeper examination of which types of contextual information are more likely to be forgotten and to be correlated with information about the perpetrator. This knowledge may be useful in witness testimony, as it may help jurors assess the likely accuracy of ‘central’ information when witnesses provide different types of ‘peripheral’ information, which are often more readily subject to confirmation.

To our knowledge, the levels of accuracy obtained in a forced-choice paradigm for memory of different types of non-emotional, contextual information remains unknown. We postulated that the perpetrator would likely be viewed as the central character, due to the perpetrator committing what was presumably the most salient act in the video. This is in line with Flowe, Takarangi, Humphries, and Wright’s (2016) designation of ‘central’ information as information related to the perpetrator. In contrast, other research that has examined memory

of ‘central’ and ‘peripheral’ information did not use this approach, and did not limit ‘central’ information to just the perpetrator as we did (Burke et al., 1992; Heath & Erickson, 1998; Libkuman et al., 2004). Examining ‘central’ details often includes seeking information about the victim of an event, as the victim can be considered a central part of the event. However, our focus here was to examine the typically sought after, or central, information regarding the perpetrator; interviewers often know information about the victim, but may lack crucial information regarding the perpetrator. Hence, we here refer to the identity and details about the perpetrator as the ‘central’ information. Of note, we did not define or differentiate between ‘central’ or ‘peripheral’ information for participants. That is, participants watched the film without any instruction as to what was ‘central’ information.

We hypothesized that forgetting in the different categories of the contextual information would correlate positively with forgetting of the perpetrator information. Conversely, participants could possibly devote the most attention to perpetrator information at the cost of contextual information, which could result in a negative correlation between contextual information and perpetrator information (Easterbrook, 1959). We take a novel approach in examining memory for different contextual aspects of an event in a forced-choice paradigm, in order to evaluate whether certain ‘peripheral’ contextual information correlate more than others with the ‘central’ information.

Method

Participants

We recruited 114 SUNY-Binghamton undergraduates to participate in this experiment ($M = 19$ years of age, $SD = 1.01$). Sixty-nine participants were female and 45 participants were male. Participants were randomly assigned to either an ‘Immediate’ condition ($n = 54$), or a ‘Delay’ condition ($n = 60$). The difference in group sizes reflects the random procedure used in assigning participants to the two groups. We initially collected data from 189 participants, but the data of any participant who failed to fill out the Scantron correctly (i.e., failing to answer a question), or of any participant who failed to return after the 48-hour delay were eliminated from the analysis. Thus, the data for 75 participants were eliminated. Consequently, the analyzed data in this experiment reflect 114 participants. The total number of participants in each group was based on sample sizes of 45–48 being appropriate for detecting differences between two groups based on a small to moderate effect size, Cohen’s $d = 0.30$ (Cohen, 1988). The protocol for this study was approved by the SUNY-Binghamton Institutional Review Board and all participants gave prior written informed consent in accordance with the Declaration of Helsinki.

Materials and Design

Participants initially signed up for both parts of the experiment, such that they completed the first part of the experiment on the first day, and returned for the second part 48 hours. All participants watched a video of a purse theft. The Immediate Group was given a test immediately after watching the video (the Initial Test), and returned 48 hours later for a second test (the Final Test). The Delay Group was tested only after a 48-hour delay. We use the term ‘Day One’ to refer to the first part of the experiment, during which all participants

watched the video, but only the Immediate Group took the Initial Test. We use the term ‘Day Three’ to refer to the second part of the experiment that took place after a 48-hour delay, upon which the Immediate group took their second test (the Final Test) and the Delay Group took the test for the first time (the Final Test). The Immediate Group was given the same questions, in a different order, on their Day One and Day Three tests.

We used a forced-choice procedure for the memory test because it facilitated the quantitative comparison of recall across different contextual aspects of the target event and reflected the direct, short-answer formats often employed in interviews. We assessed participants’ recognition accuracy of information about the perpetrator and aspects of contextual memory, including where and when the event took place, and who else was present at the event.

Test questions.

The test consisted of 34 forced-choice questions pertaining to the video, with two additional questions (35 and 36) at the end asking the participants whether they had watched the video today and whether they had previously answered questions about the video. The 34 questions focused on four main categories: ‘Perpetrator,’ ‘When’ (temporal information), ‘Where’ (spatial information), and ‘Who’ (who else was at the event). ‘Perpetrator’ questions emphasized the perpetrator’s physical appearance and clothing. ‘When’ questions asked about temporal information, such as the duration of actions, order of actions, and the time of year in which the event occurred, based on the date appearing on a whiteboard menu, trees outside a window, and the clothing of the people in the video. ‘Where’ questions alluded to the relative spatial locations of objects and people, such as where the cash register was located, the seating orientation of the victim relative to the victim’s friend, and features of objects, such as their color. ‘Who’ questions asked about features of the other people at the event, such as the victim’s hair and shirt color, and what the other patrons were doing at the event. For all but the last two questions, the answer choices consisted of one correct answer, two plausible foils (incorrect answer choices), a “None of the above” choice, and an “I do not know” choice. Both foil options, “I do not know” and “None of the above,” were coded as incorrect answers relative to the correct answer reflecting what actually occurred in the video. The last two questions (35 and 36) could only be answered with “yes” or “no”, and their purpose was simply to provide evidence that we had correctly matched the Day One and Day Three answer sheets (Scantrons) from each participant. See Appendix for representative test questions.

Prior to the present experiment, we had run a pilot study designed to identify a set of questions from each content area (‘Perpetrator’, ‘When’, ‘Where’, and ‘Who’) that on average would be equally apt to be answered correctly and that demonstrated a relatively high degree of inter-item reliability within each of the designated content areas (see Supplementary Materials).

The selected 34 questions for the present experiment included nine questions in the three context categories, which allowed us to maintain comparable sensitivity. The ‘Perpetrator’ category contained only seven questions due to the lack of further testable content concerned solely with the perpetrator.

Procedure

All participants completed the task in individual cubicles devoted to computer-based psychology experiments. On Day One, participants were asked to read and sign the Informed Consent form. Then, they were asked to follow the instructions on the computer screen, and when prompted to do so, follow the directions in the packet in a folder next to the computer.

All participants viewed the following instructions upon sitting down at their computers: “Thank you for participating in our study. The experiment depends on your participation both today and two days from now. You will be shown a video shortly. Pay close attention to the video. Press [SPACEBAR] to start the video.” After watching the video, participants were asked to turn to the folder next to their computers. Participants in the Immediate Group received printed instructions that they would be taking a test (the Initial Test), and would need to answer all questions. At the end of the test booklet, these participants were informed that this part of the experiment was over, and that they should return in exactly 48 hours. The Delay Group participants were informed that this part of the experiment was over, and that they should return in exactly 48 hours. Upon returning 48 hours later, all participants saw the following instructions: “Thank you for returning today. Now please open the folder next to the computer and carefully follow the directions provided.” All participants were informed that they would be taking a test (the Final Test) on the video they had previously viewed, and that the experiment was complete when they finished the test.

Methods are described in further detail in the Supplementary Materials.

Statistical Analysis

Participant accuracy was determined by calculating the mean number of questions correct per content category. First, a 2×4 mixed-design analysis of variance (ANOVA) was performed to assess forgetting over a 48-hour delay with Immediate Day One and Delay Day Three as a between-subjects factor, and content area as within-subjects factors. This was followed by planned contrasts to examine the change in accuracy across the delay for each content area. Second, a 2×4 mixed-design ANOVA with Immediate Day Three and Delay Day Three as a between-subjects factor, and content area as within-subjects factors was used to examine the effect of the Initial Test on test performance 48 hours later. Subsequent planned contrasts were conducted to examine the effect of immediate testing on later accuracy across the different content areas. Third, a 2×4 fully within-subject ANOVA was conducted to compare performance on the first test (the Initial Test) of Group Immediate with performance on the second test (the Final Test) of Group Immediate to assess the effects of early testing on later testing. Fourth, exploratory Pearson correlations were performed within content areas to determine whether for Group Immediate there was a relationship between performances on Day One and Day Three. Fifth, a composite score combining contextual ‘Who’, ‘Where’ and ‘When’ information was created to assess the overall relationship of all contextual information relative to ‘Perpetrator’ information by performing Pearson correlations. Additionally, Pearson correlations were used to examine whether memory of one context area was correlated with memory of the other context areas. We also examined how many participants in each group (Group Immediate on both Day One

and Day Three and Group Delay) responded “I do not know”. We calculated the number of incorrect answers each group provided, and then calculated the percentage of participants who responded “I do not know” out of the total number of incorrect answers per group. Lastly, we calculated the number of incorrect answers per group omitting the “I do not know” response as an incorrect answer. Results were considered significant when $p < .025$. We used a decision axis of $p < .025$ rather than the conventional $p < .05$ because the various ANOVAs conducted collectively used each data set twice. Hence, the more stringent alpha value of .025 corrected for this, thereby reducing the chances of a Type I error.

Results

Initial and Final Tests

In order to assess forgetting over the 48-hour delay, recognition proportions of Group Delay on Day Three were compared with those of Group Immediate on Day One (i.e., the first test for each group). A 2 (Immediate Day One vs. Delay Day Three; between-subjects) \times 4 (Content Area; within-subjects) ANOVA detected a main effect of the 48-hour delay, $F(1, 112) = 17.26$, $p < .001$, Cohen's $f = 0.39$, 95% CI [0.20, 0.58], with those participants who were first tested after the delay period demonstrating less accuracy on the questionnaire in all four categories of questions (see Figure 1). A main effect of content area on accuracy was also observed, $F(3, 336) = 18.86$, $p < .001$, Cohen's $f = 0.69$, 95% CI [0.50, 0.88]. Additionally, there was a significant interaction between delay and content area, $F(3, 336) = 3.35$, $p = .019$, Cohen's $f = 0.26$, 95% CI [0.04, 0.46].

As we were interested in potential differences in the degree of forgetting over the 48-hour delay among the different content areas, planned contrasts were conducted (using the error term from the overall ANOVA) to test for changes in performance across the delay for each content area (i.e., Group Delay on Day Three vs. Group Immediate on Day One). Group Delay exhibited markedly lower performance on the ‘Perpetrator’ questions, $F(1, 112) = 8.21$, $p < .005$, Cohen's $f = 0.27$, 95% CI [0.08, 0.45], as well as the ‘Where,’ $F(1, 112) = 20.23$, $p < .001$, Cohen's $f = 0.42$, 95% CI [0.23, 0.61], and ‘When’ content areas, $F(1, 112) = 6.45$, $p = .012$, Cohen's $f = 0.24$, 95% CI [0.05, 0.42]. However, the nominal decrease in performance on the ‘Who’ context area questions was not reliable, $F < 1$. Thus, we observed appreciable forgetting for ‘Perpetrator’, ‘Where’, and ‘When’ questions, but not for ‘Who’ questions.

To determine whether early testing protected against forgetting, the recognition proportions of Group Immediate on Day Three were compared with Group Delay on Day Three. Thus, recognition proportions on the second test for participants who were tested twice were compared to those tested only once (see Figure 2). A 2 (Immediate Day Three vs. Delay Day Three; between-subjects) \times 4 (Content Area; within-subjects) mixed-design ANOVA detected a main effect of the 48-hour delay, $F(1, 112) = 15.65$, $p < .001$, Cohen's $f = 0.37$, 95% CI [0.18, 0.56], with those participants tested only after the 48-hour delay displaying less accuracy for all content areas. We also observed a main effect of content area on accuracy, $F(3, 336) = 19.67$, $p < .001$, Cohen's $f = 0.71$, 95% CI [0.51, 0.90]. Additionally, there was a significant interaction between delay and content area, $F(3, 336) = 4.47$, $p < .005$, Cohen's $f = 0.32$, 95% CI [0.11, 0.51]. This interaction makes interpretation of

preceding main effects tenuous. Therefore, we conducted planned contrasts to examine the change in accuracy across the delay for each content area. Group Delay participants displayed lower performance on the ‘Perpetrator’ questions, $F(1, 112) = 9.95, p < .005$, Cohen’s $f = 0.29$, 95% CI [0.11, 0.48], the ‘Where’ questions, $F(1, 112) = 22.62, p < .001$, Cohen’s $f = 0.44$, 95% CI [0.25, 0.64], and the ‘When’ questions, $F(1, 112) = 5.91, p = .017$, Cohen’s $f = 0.23$, 95% CI [0.04, 0.41]. However, the nominally lower performance on the ‘Who’ questions was not reliable, $F < 1$. Thus, the Initial Test on Day One resulted in better performance on the Final Test on Day Three relative to no testing on Day One for ‘Perpetrator,’ ‘Where,’ and ‘When’ questions, but not for ‘Who’ questions.

In addition, the Initial Test of Group Immediate was compared to the Final Test of Group Immediate to determine whether there was a significant difference; that is, despite early testing clearly providing protection against forgetting, we were interested in assessing whether there was still appreciable forgetting over 48 hours following immediate testing. A 2 (Immediate Day One vs. Immediate Day Three; within-subjects) \times 4 (Content Area; within-subjects) ANOVA detected no main effect of the 48-hour interval between the first and second tests, $F < 1$. There was a main effect of content area on accuracy, $F(3, 159) = 10.31, p < .001$, Cohen’s $f = 0.52$, 95% CI [0.31, 0.72], but no significant interaction between the 48-hour interval between tests and content area, $F(3, 159) < 1$. See Figure 3.

As previously stated, participants in Group Immediate had highly similar recognition proportions on Day One and Day Three. However, we wanted to further examine whether recognition proportions for any specific content area on Day One had any relationship to recognition proportions for that same content area on Day Three. Therefore, autocorrelations within content areas between Day One memory and Day Three memory for Group Immediate participants were calculated, which found ‘Perpetrator’: $r = 0.89, r^2 = 0.79, p < .001$; ‘Where’: $r = 0.80, r^2 = 0.64, p < .001$; ‘When’: $r = 0.78, r^2 = 0.61, p = .001$; and ‘Who’: $r = 0.74, r^2 = 0.55, p < .001$. Thus, participants in the Group Immediate were significantly self-consistent across the delay for all content categories.

Memory of Context Areas

To determine whether performance on one or another type of context question could predict performance on ‘Perpetrator’ questions, Pearson correlations were calculated. When all three tests were considered, contextual questions about the other people at the event (the ‘Who’ questions) tended to be the most predictive of accuracy on ‘Perpetrator’ questions. That is, participants’ performance on the ‘Who’ questions was best correlated with performance on ‘Perpetrator’ questions (see Table 1). In addition, we ran Pearson correlations to examine whether memory of one content area can predict memory of the other content areas (see Table 2). ‘When’ questions correlated significantly with performance on ‘Where’ and ‘Who’ questions for Group Delay participants, and performance on ‘Where’ questions correlated significantly with performance on ‘Who’ questions for Group Immediate participants on Day Three.

“I do not know” Responses

To assess whether differences in confidence might have masked intact memories of the video, we compared the frequency of “I do not know” responses between Group Delay and Group Immediate on both Day One and Day Three. We first calculated the total number of incorrect responses for each condition, and then calculated the percentage of participants who responded “I do not know” from the total number of incorrect responses. See Table 3. We used this method because participants in Group Delay did worse than those in Group Immediate Day One and Group Immediate Day Three, so there were apt to be more “I do not know” responses observed for Group Delay merely because of the difference in the base rate of incorrect responses. The percentages of participants who responded “I do not know” out of the total number of incorrect responses for each group were: Delay, 39.8%, Immediate Day One, 40.2%, Immediate Day Three, 38.6%. To further examine these differences, we conducted Mann-Whitey tests. When we compared Group Delay to Group Immediate Day One, we found that $z = 0.91$, $p = .363$. When we compared Group Delay to Group Immediate Day Three, we found that $z = 1.53$, $p = .126$. Thus, the Mann-Whitney tests were non-significant. We then computed Bayes factors, and used the conventional definition of 3.00 for moderate support for the hypothesis being tested. When we compared Group Delay to Group Immediate Day One using a scale of $r = 1$, we found that the Scaled JZS Bayes Factor (BF_{01}) = 4.68. Thus, when Group Delay was compared to Group Immediate Day One, the Bayes factor lends support to the null hypothesis, which indicates that the hypothesis of a difference in the frequency of “I do not know” responses between these groups was not supported. When we compared Group Delay to Group Immediate Day Three, we found that the Scaled JZS Bayes Factor (BF_{01}) = 2.31. Thus, the Bayes factor fails to lend support to the absence of a difference in the frequency of “I do not know” responses between Group Delay and Group Immediate Day Three.

Incorrect Responses Omitting “I do not know” Responses in all Conditions

We were interested in examining whether omitting the “I do not know” responses from participants’ incorrect responses would maintain significant differences regarding accuracy between groups. Therefore, only foil answers and “None of the above” were coded as incorrect for these analyses. We calculated the percentages of incorrect responses and removed the percentage of “I do not know” for each group. To do so, we subtracted the percentage of “I do not know” responses from the percentage of incorrect responses, and divided that total by the percentage of participants who answered “I do not know” subtracted from the percent of the total incorrect responses (100%). See Table 4. To compare these differences, we ran Mann-Whitey tests. When we compared Group Delay to Group Immediate Day One, we found that $z = 2.91$, $p = .004$. When we compared Group Delay to Group Immediate Day Three, we found that $z = 2.54$, $p = .011$. Thus, both Mann-Whitney tests were significant. We then computed Bayes Ratios, and when we compared Group Delay to Group Immediate Day One, we found that, using a scale of $r = 1$, the Scaled JZS Bayes Factor (BF_{10}) = 7.12. Thus, the Bayes Factor provided substantial evidence for the alternative, suggesting that the difference between the number of incorrect responses between Group Delay and Group Immediate Day One, even when omitting “I do not know” responses, was supported. When we compared Group Delay to Group Immediate Day Three, we found that Scaled JZS Bayes Factor (BF_{10}) = 2.86. Thus, the Bayes Factor was weakly in

support of the alternative. Although this Bayes Ratio falls just short of the 3.00 value that is commonly looked at as a criterion for rejection of the null hypothesis, that is based on comparing the null against a two tailed alternative. Here we have grounds for a one-tailed alternative, making the Bayes factor appropriate for concluding that it supports rejecting the null hypothesis.

Discussion

The present results confirm that recognition accuracy declined over a delay; that is, forgetting occurred over the 48-hour delay. Specifically, we found a main effect of the 48-hour delay, such that participants who were tested for the first time after a 48-hour delay exhibited less accurate memory than participants who had been tested immediately after viewing the video. Additionally, early testing slowed forgetting. Lastly, participants' overall accuracy in identifying contextual components of the event was a good indicator of accuracy in remembering features of the perpetrator.

That forgetting occurred over the 48-hour delay is in line with previous findings that a delay between viewing an event and being asked questions about it often leads to a decrease in memory accuracy (Jones & Pipe, 2002; Ornstein et al., 2006; Poole & White, 1991). Moreover, that initial testing provided some protection against forgetting as assessed on a later test is consistent with the testing effect (Abel & Roediger, 2017; Gates, 1917; Wheeler & Roediger, 1992). Notably, this protection against forgetting was observed despite the absence of feedback on the Initial Test. Although recognition tests have been shown to sometimes be less sensitive to the testing effect than recall tests (Darley & Murdock, 1971), we nevertheless saw a testing effect with the present recognition measure. Despite having observed a testing effect, we must clarify that the present experiment was not centrally an examination of the testing effect per se. Rather, we were interested in the benefit for a later test of an earlier test as opposed to no earlier test. Although we expected to observe a testing effect based on previous literature, it was important to actually test this hypothesis in the present design.

Initial testing of Group Immediate seems to have provided some protection against subsequent forgetting; however, the data do not suggest that performance improved at the Final Test relative to the Initial Test. Similar results have been observed in other experiments that used a forced-choice procedure (Dunning & Stern, 1992; Shaw & McClure, 1996). Moreover, we are unable to conclude whether taking the Initial Test on Day One strengthened the memory of the responses, or whether participants simply remembered the responses they had provided on the Initial Test when taking the Final Test on Day Three. Although we are unable to ascertain which is the case, taking the Initial Test clearly prevented forgetting on the Final Test in Group Immediate relative to the absence of an Initial Test in Group Delay.

In regards to our secondary hypothesis, participants' memory of the contextual categories was positively correlated with memory of the perpetrator. Therefore, assessing observer accuracy for memory of the surrounding context of an event appears to be informative in assessing observer accuracy for memory of the perpetrator. Our results confirm the widely

assumed, but infrequently tested, positive correlation between memory for ‘central’ and ‘peripheral’ information. Our approach was novel in that we examined specific types of contextual information, which were separated into categories of where and when events took place, and who else was at the scene. This is in contrast to previous research that evaluated memory for peripheral information, but did not differentiate among various types of ‘peripheral’ details, and thus do not provide insight as to which types of ‘peripheral’ information are most likely to be forgotten and which types are correlated with perpetrator information. Understanding which specific types of ‘peripheral’ details are more likely to be forgotten, and the degree of correlation between ‘peripheral’ information and information about the perpetrator, potentially can be useful, particularly in witness testimony, in order to help jurors assess the likely accuracy of ‘central’ information when witnesses provide various types of ‘peripheral’ information (Bell & Loftus, 1988; Wells & Leippe, 1981). Additionally, this information may guide interviewers in forming questions that focus on aspects of ‘peripheral’ information that have a greater degree of correlation with information about the perpetrator.

Regarding correlations between contextual categories and information about the perpetrator, answers to questions concerning other people at the event (‘Who’ questions) tended to be the most predictive information concerning accuracy for questions about the perpetrator. Additionally, memory of each context area tended to be positively correlated with memory of the other context areas. Despite there being no significant difference in overall performance between the Initial Test and the Final Test for Group Immediate, there was a significant correlation of performance on the Initial Test with performance on the Final Test for all content categories. This suggests that participants who did well on each content category on the Initial Test also did well on the same content category questions on the Final Test. Taken together, this demonstrates stability of performance on all categories over the delay.

As stated above, accuracy on ‘Who’ questions tended to be best correlated with accuracy on ‘Perpetrator’ questions. Possibly, this reflects the perpetrator and the other people in the video all being people, and consequently participants who were prone to remembering information about people being able to do well on both ‘Perpetrator’ and ‘Who’ questions. Additionally, participants who were strongly predisposed to attend to people might be expected to form more elaborate associative networks with respect to attributes of people, such as having hair, clothing, and facial features, and consequently performed better on questions about both the perpetrator and other people at the event (Chi & Koeske, 1983; Hockley, 2008). Thus, remembering a specific detail of an event when being asked a relevant question may allow attributes remembered at that time to serve as additional cues for subsequently remembering related details of the event. This suggests that a means of obtaining more accurate observer reports is to ask about verifiable details that are similar in nature to the ‘central’ content to not only assess the likely accuracy of the observer, but also to enhance accuracy.

We note that it is possible that the stable memory across the delay for ‘Who’ information may have in part been due to the ‘Who’ characters having been present during much of the video. Specifically, the victim and the victim’s friend were in the foreground for the majority

of the video. Although the perpetrator was present for the duration of the video, he was highly inconspicuous, as one could only see the top of his head in the right bottom corner of the screen, until he stood up to steal the purse towards the end of the video. Perhaps after a 48-hour delay, the initial memory for even the most salient event (including details about the perpetrator) decayed to a nearly equivalent level of memory to that for the other people present at the event. As the other characters likely held the viewers' attention for the majority of the film, perhaps the memory for the 'Who' information remained more stable across the 48-hour delay. Nevertheless, it appears that perpetrator and 'Who' information received greater amounts of attentional resources, which resulted in less forgetting for these aspects of the event.

For Group Delay participants, memory for both the 'Where' and 'When' information were less strongly correlated than 'Who' information with memory for 'Perpetrator' information. As previously mentioned, Easterbrook's (1959) attentional narrowing hypothesis suggests that people often devote more attentional resources to 'central' information, leaving fewer resources to attend to 'peripheral' information. Based on this, the presumably less salient nature of temporal and spatial information may have led to the temporal and spatial information being processed as 'peripheral' information. This could have resulted in participants devoting less attention to these types of information, and consequently allowing the larger amount of forgetting observed in Group Delay participants concerning 'When' and 'Where' information compared to 'Perpetrator' and 'Who' information. This is consistent with previous research that demonstrated that participants are more likely to maintain a superior level of accuracy when asked to remember 'central' information as compared to 'peripheral' information (Ibabe & Sporer, 2004). However, we did not specifically examine participants' attention or encoding strategies; rather, our results only suggest that varying amounts of attentional resources were devoted to different types of information, which resulted in different amounts of observed forgetting.

Under the assumptions of a monitoring and control framework, it is possible that participants in Group Delay may have refrained from providing an answer because they were feeling unsure, rather than because they did not remember the information. Research has been conducted to examine metacognition, defined by Ackerman and Thompson (2017) as "processes that monitor our ongoing thought processes and control the allocation of mental resources" (p. 607). Koriat, Ma'ayan, and Nussinson (2006) distinguish between metacognitive monitoring and metacognitive control: "metacognitive monitoring refers to the subjective assessment of one's own cognitive processes and knowledge, whereas control refers to the processes that regulate cognitive processes and behavior" (p. 38). Koriat et al. (2006) cite evidence suggesting that metacognitive monitoring and metacognitive control influence one another, such that metacognitive monitoring (such as metacognitive judgments) can affect information processes, and information processes provide feedback for metacognitive monitoring. A related phenomenon is 'feeling of knowing', which acts as an internal monitor to signal to a person whether an item is stored in memory (Hart, 1965). If this signal suggests that an item is not in memory, Hart (1965) posited that it would not be beneficial to continue using resources to search for the item. Research has examined feeling of knowing judgments, and has found that people spend more time looking for an item in memory when they feel that the item is accessible in memory, compared to when they feel

that it is not accessible (Costermans, Lories, & Ansay, 1992). Additionally, Koriatic and Goldsmith (1996) observed that participants were inclined to provide or refrain from giving information regarding a witnessed past event depending on their subjective confidence that the information is correct.

To assess the potential role of metacognition, we examined whether participants in Group Delay responded “I do not know” more frequently than the participants in Group Immediate on Day One and Day Three. We found that there was not a significant difference between Group Delay and Group Immediate Day One for the number of participants who, out of the participants who responded with an incorrect answer, specifically responded “I do not know”. This suggests that the fewer questions correct in Group Delay, as compared to Group Immediate Day One, does not appear to have resulted from lower confidence (as opposed to weaker memory) by the participants in Group Delay. However, the significantly larger number of “I do not know” responses by Group Delay relative to Group Immediate Day Three may indicate that participants in Group Delay had lower levels of confidence than those in Group Immediate Day Three. If this is the case, perhaps participants in Group Immediate on Day Three were more confident in their responses due to the fact that they were more familiar with the test questions because they had previously seen them. We did not measure confidence levels, metacognition, or feeling of knowing in the current experiment, and cannot speak with any degree of certainty to the underlying processes that occurred, but we note that these are phenomena that may have played a role in our observed results.

Although in some eyewitness analyses “I do not know” is not counted as an incorrect choice, the present research was a study of general observer memory, not eyewitness accuracy. However, to assess our data from the viewpoint that “I do not know” was not an incorrect answer, we reanalyzed the data omitting the “I do not know” responses. We found that there was a significant difference in the number of incorrect responses between Group Delay and Group Immediate Day One, and between Group Delay and Group Immediate Day Three, even when omitting “I do not know” responses. Thus, our results demonstrate that the Initial Test on Day One for Group Immediate participants provided some protection from forgetting that would have otherwise occurred.

Limitations

The present study has several limitations that merit comment. First, the questions concerning the different content areas were not perfectly matched with respect to difficulty. Due to the inherently different natures of the different types of information, it was not clear to us how to create questions of equal difficulty, despite our attempt to do so through the pilot study. Second, our usage of a forced-choice procedure is not a recommended questioning technique. Ideally, people who observed an event should be asked to provide detailed free-recall responses, with minimal direction from an interviewer. However, in practice, interviewers are often imperfect when it comes to avoiding leading questions and direct, short-answer questions (Clifford & George, 1996; Fisher et al., 1987; Ginet & Py, 2001), thereby making it necessary to explore the potential consequences of not only the best practices, but also the less desirable practices. These include forced-choice questions and the

similarly formatted direct, short-answer questions. Additionally, although many previous studies have tested witnesses using a recall procedure (Quas et al., 2007; Scrivner & Safer, 1988; Turtle & Yuille, 1994; Yuille & Cutshall, 1986), it is important to continue examining the effects of forced-choice procedures and other practices that reflect how observers may be interviewed, particularly when they are interviewed by people who use deleterious questioning styles, or by other people who ask questions about the event in non-professional settings. Despite the obvious drawbacks of our use of forced-choice questions, the procedure facilitated quantitative comparison of the different memory categories. Although participants were unable to reveal details outside the scope of the forced-choice answers, we did provide “None of the above” and “I do not know” answers as options to imitate real-life interviews in which people may say that they do not know the answer. In addition to these options and the correct answer, we provided two concrete foils to verify that participants were able to recognize the foils as false. Admittedly, our providing foil choices close to the time of acquisition may have led to consequences such as distorted memory. But our data and preparation were not intended to speak to our participants’ susceptibility to being misled, nor do the present Group Immediate data suggest that Day One testing impaired Day Three recognition.

As previously stated, we observed that early testing slowed future forgetting. Notably, our conclusion here is not to say that early testing is better than late testing in respect to a final test. Rather, we examined the result of testing half of the participants immediately after watching the video, and re-testing them after a delay, and not testing the other half of the participants until after a delay. We observed greater forgetting by the participants who were tested for the first time after a delay, as compared to the participants who had been tested immediately after watching the video and again after a delay. Thus, we highlight the importance of early testing when we compare the effect of taking an early test as compared to not taking an early test. Whether the benefit seen in Group Immediate on Day Three relative to Group Delay is actually due to the Initial Test received by Group Immediate being on Day One or merely to Group Immediate having had a prior test by the Day Three test is unclear based on the present design. That is, we are unable to conclude how great of an impact of the Initial Test being soon after observing the video compared to simply the absence of a prior test in the Delay Group participants’ forgetting of information. Future research should examine the consequences of having a Delay Group taking two tests after a 48-hour delay to determine whether taking a test for the first time on Day Three would provide an observed benefit for a second test also taken on Day Three.

Given our experimental approach that used only one video as our stimulus, it is possible that memory for the different contextual categories may have been driven by the specific video stimulus. Future research should replicate the present study’s design with other video stimuli to examine whether our findings generalize to other target material. Nevertheless, the present results suggest that it is beneficial to ask observers about the context of an event in addition to asking about the ‘central’ information. This is consistent with some witness interviewing protocols, such as the cognitive interview, which asks witnesses about the context of an event (Fisher & Geiselman, 1992). However, the techniques employed in a cognitive interview refer to the context in a different manner than the present experiment. One technique used in a cognitive interview encourages witnesses to create a mental

reinstatement of the event (Memon et al., 1997). This includes asking witnesses to remember perceptual details of the event, including smells and sounds, how they felt emotionally, and what they were doing at the time of the event. Our findings suggest asking observers about the context of an event in regards to where and when the event occurred and who else was at the event may also be helpful in obtaining comprehensive information about an event.

We note that the observed differences in Group Immediate participants' Initial Tests across the different types of information were potentially failures either in encoding or retrieval, as we have no way of knowing based on our design. Similarly, the performance decrement that we observed across the 48-hour delay could have been due to either an irrevocable loss or a retrieval failure (i.e., a lapse), but not an encoding failure in that Group Immediate did better on their Initial Tests, which testifies to the strength of initial encoding.

Although we have briefly discussed underlying mechanisms potentially responsible for the effects that we observed, they were not central to the focus of the present research, which was empirical rather than theoretical. Assessment of the present hypotheses is potentially important to application independent of the bases of the phenomena. The present results suggest which specific types of 'peripheral' information are best correlated with information about the perpetrator, which may be useful in witness testimony. Moreover, it is important to understand which types of peripheral details are more likely to be forgotten, in contrast to simply stating that peripheral details in general have a more rapid rate of forgetting relative to central details (Sekeres et al., 2016). We found that peripheral information regarding where and when the event took place were more likely to be forgotten over a delay than information concerning who else was at the scene of the event. Understanding which types of peripheral information are more prone to being forgotten can potentially guide interviewers in asking more questions that pertain to certain contextual details (i.e., who else was at the event) that are less likely to be forgotten, and are highly correlated to information about the perpetrator. This may help activate associative networks and elicit memory of the 'central' information that is typically sought in witness testimony (i.e., information about the perpetrator). Additionally, understanding the correlation between the different types of 'peripheral' information and information about the perpetrator may assist jurors in assessing the likely accuracy of 'central' information when provided with 'peripheral' information.

Conclusions

In sum, we observed that information about the other people at the event was best correlated with information about the perpetrator. The present data support the widely held but rarely tested supposition that accurately answering questions about contextual details about the event is indicative of the observers' accuracy in correctly answering questions about the perpetrator. Yet, memory for temporal and spatial information, which may be interpreted as 'peripheral' information, may be less accurate, at the expense of greater attentional resources having been devoted to the 'central' information. Nevertheless, asking questions about the perpetrator and the context may better activate the associative network concerning the entire event than asking questions only about the perpetrator. Thus, we suggest that investigators ask questions about the surrounding context as well as about the perpetrator in order to obtain information about the perpetrator, and to assess the likely accuracy of memory of the

perpetrator. This lends support to the lay belief of many jurors who assume that the accuracy of memory of ‘peripheral’ details of a crime is indicative of the accuracy of memory of the culprit (Bell & Loftus, 1988; Wells & Leippe, 1981). Thus, memory of contextual information is potentially a useful index of observer credibility outside of the laboratory when the authorities have detailed knowledge concerning the context of a crime (Berman, Narby, & Cutler, 1995). In addition, we observed that testing participants early after observing an event and then re-testing them protected target memories from partial forgetting that may have otherwise occurred. Therefore, we suggest that investigators interview people who have observed a specific event as early after the event as possible, not only to minimize forgetting at the initial interview, but also to reduce forgetting by the time of subsequent interviews.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by NIMH grant 33881. We would like to thank Alaina S. Berruti, Zekiel Z. Factor, Jeff J. Joseph, Audrey Li, and Patty Li for their comments on an earlier version of this paper.

Appendix:

Representative Test Questions

‘Perpetrator’ Questions:

1. The thief:
 - a. Had a long beard
 - b. Had a short beard
 - c. Was clean shaven
 - d. None of the above
 - e. I do not know
2. The thief was wearing:
 - a. A hat
 - b. A t-shirt
 - c. A leather jacket
 - d. None of the above
 - e. I do not know

‘Where’ Questions:

1. What was the menu displayed on?
 - a. A white board

- b.** A chalkboard
 - c.** A large computer monitor
 - d.** None of the above
 - e.** I do not know
- 2.** Where was the cash register located?
 - a.** At the left side of the room
 - b.** At the front of the room
 - c.** At the right side of the room
 - d.** None of the above
 - e.** I do not know

‘When’ Questions:

- 1.** Approximately how long was the victim seated before the victim got up to order?
 - a.** 5 seconds
 - b.** 10 seconds
 - c.** 20 seconds
 - d.** None of the above
 - e.** I do not know
- 2.** Which of the following events happened last?
 - a.** The server approached the bystanders
 - b.** The victim’s friend sat down
 - c.** The victim approached the server
 - d.** None of the above
 - e.** I do not know

‘Who’ Questions:

- 1.** The customers at the table nearby both had:
 - a.** Brown hair
 - b.** Blond hair
 - c.** Red hair
 - d.** None of the above
 - e.** I do not know
- 2.** In all, how many people did you see in the scene?
 - a.** 9

- b. 4
- c. 6
- d. None of the above
- e. I do not know

References

- Abel M, & Roediger HL (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition*, 45(1), 81–92. doi:10.3758/s13421-016-0641-8 [PubMed: 27464491]
- Ackerman R, & Thompson VA (2017). Meta-Reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607–617. doi:10.1016/j.tics.2017.05.004 [PubMed: 28625355]
- Akehurst L, Milne R, & Koehnken G (2003). The effects of children's age and delay on recall in a cognitive or structured interview. 0, 9(1), 97–107. doi:10.1080/10683160308140
- Bäuml KT, & Kliegl O (2017). Retrieval induced remembering and forgetting In Byrne JH (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., Vol. 2, pp. 27–51). Oxford: Academic Press.
- Bell BE, & Loftus EF (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology*, 18(14), 1171–1192. doi:10.1111/j.1559-1816.1988.tb01200.x
- Berman GL, Narby DJ, & Cutler BL (1995). Effects of inconsistent eyewitness statements on mock-jurors evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior*, 19(1), 79–88. doi:10.1007/bf01499074
- Burke A, Heuer F, & Reisberg D (1992). Remembering emotional events. *Memory & Cognition*, 20(3), 277–290. doi:10.3758/bf03199665
- Chan JCK, & Langley MM (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 248–255. doi:10.1037/a0021204
- Chan JC, Manley KD, & Lang K (2017). Retrieval-enhanced suggestibility: A retrospective and a new investigation. *Journal of Applied Research in Memory and Cognition*, 6(3), 213–229. doi:10.1016/j.jarmac.2017.07.003
- Chi MT, & Koeske RD (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, 19(1), 29–39. doi:10.1037//0012-1649.19.1.29
- Christianson SA, & Loftus EF (1987). Memory for traumatic events. *Applied Cognitive Psychology*, 1(4), 225–239. doi: 10.1002/acp.2350010402
- Clifford BR, & George R (1996). A field evaluation of training in three methods of witness/victim investigative interviewing. *Psychology, Crime & Law*, 2(3), 231–248. doi:10.1080/10683169608409780
- Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Costermans J, Lories G, & Ansary C (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(1), 142–150. doi:10.1037//0278-7393.18.1.142
- Darley CF, & Murdock BB (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, 91(1), 66–73. doi:10.1037/h0031836
- Dunning D, & Stern LB (1992). Examining the generality of eyewitness hypermnnesia: A close look at time delay and question type. *Applied Cognitive Psychology*, 6(7), 643–657. doi:10.1002/acp.2350060707
- Easterbrook JA (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66(3), 183–201. doi:10.1037/h0047707 [PubMed: 13658305]

- Fisher RP, & Geiselman RE (1992). Memory-enhancing techniques for investigative interviewing: The cognitive interview. Springfield, IL: Charles C. Thomas.
- Fisher RP, Geiselman RE, & Raymond DS (1987). Critical analysis of police interview techniques. *Journal of Political Science and Administration*, 15(3), 177–185.
- Flin R, Boon J, Knox A, & Bull R (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, 83(3), 323–336. doi:10.1111/j.2044-8295.1992.tb02444.x [PubMed: 1393364]
- Flowe HD, Takarangi MKT, Humphries JE, & Wright DS (2015). Alcohol and remembering a hypothetical sexual assault: Can people who were under the influence of alcohol during the event provide accurate testimony? *Memory*, 24(8), 1042–1061. doi: 10.1080/09658211.2015.1064536 [PubMed: 26278075]
- Gates AI (1917). *Recitation as a factor in memorizing*. New York: Science Press.
- Ginet M, & Py J (2001). A technique for enhancing memory in eye witness testimonies for use by police officers and judicial officials: The cognitive interview. *Le Travail Humain*, 64(2), 173–191. doi:10.3917/th.642.0173
- Hart JT (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56(4), 208–216. doi: 10.1037/h0022263 [PubMed: 5825050]
- Heath WP, & Erickson JR (1998). Memory for central and peripheral actions and props after varied post-event presentation. *Legal and Criminological Psychology*, 3(2), 321–346. doi:10.1111/j.2044-8333.1998.tb00369.x
- Hershkovitz I, Lamb ME, & Katz C (2014). Allegation rates in forensic child abuse investigations: Comparing the revised and standard NICHD protocols. *Psychology, Public Policy, and Law*, 20(3), 336–344. doi:10.1037/a0037391
- Hicks JL, Marsh RL, & Russell EJ (2000). The properties of retention intervals and their effect on retaining prospective memories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26(5), 1160–1169. doi:10.1037//U278393.26.5.1160
- Hockley WE (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34(6), 1412–1429. doi:10.1037/a0013016
- Ibabe I, & Sporer SL (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology*, 18(6), 711–726. doi :10.1002/acp.1025
- Jones CH, & Pipe M (2002). How quickly do children forget events? A systematic study of children's event reports as a function of delay. *Applied Cognitive Psychology*, 16(7), 755–768. doi:10.1002/acp.826
- Koriat A, & Goldsmith M (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103(3), 490–517. doi:10.1037//0033-295x.103.3.490 [PubMed: 8759045]
- Koriat A, Ma'ayan H, & Nussinson R (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, 135(1), 36–69. doi:10.1037/0096-3445.135.1.36 [PubMed: 16478315]
- Lamb ME, Sternberg KJ, Orbach Y, Esplin PW, & Mitchell S (2002). Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Applied Developmental Science*, 6(1), 35–41. doi:10.1207/s1532480xads0601_04
- Libkuman T, Stabler C, & Otani H (2004). Arousal, valence, and memory for detail. *Memory*, 12(2), 237–247. doi: 10.1080/09658210244000630 [PubMed: 15250188]
- Memon A, Wark L, Bull R, & Koehnken G (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, 88(2), 179–197. doi:10.1111/j.2044-8295.1997.tb02629.x
- Mesoudi A, Whiten A, & Dunbar R (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, 97(3), 405–423. doi:10.1348/000712605x85871 [PubMed: 16848951]

- Mulligan NW, & Peterson DJ (2015). Negative and positive testing effects in terms of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 41(3), 859–871. doi:10.1037/xlm0000056
- Oberauer K, & Lewandowsky S (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, 115(3), 544–576. doi:10.1037/0033-295X.115.3.544 [PubMed: 18729591]
- Ornstein PA, Baker-Ward L, Gordon BN, Pelphrey KA, Tyler CS, & Gramzow E (2006). The influence of prior knowledge and repeated questioning on children's long-term retention of the details of a pediatric examination. *Developmental Psychology*, 42(2), 332–344. doi:10.1037/0012-1649.42.2.332 [PubMed: 16569171]
- Paterson HM, Eijkemans H, & Kemp RI (2015). Investigating the impact of delayed administration on the efficacy of the self-administered interview. *Psychiatry, Psychology and Law*, 22(2), 307–317. doi:10.1080/13218719.2014.947670
- Paz-Alonso PM, & Goodman GS (2008). Trauma and memory: Effects of post-event misinformation, retrieval order, and retention interval. *Memory*, 16(1), 58–75. doi:10.1080/09658210701363146 [PubMed: 17852727]
- Peterson DJ, & Mulligan NW (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(4), 1287–1293. doi:10.1037/a0031337
- Pipe M, Sutherland R, Webster N, Jones C, & Rooy DL (2004). Do early interviews affect children's long-term event recall? *Applied Cognitive Psychology*, 18(7), 823–839. doi:10.1002/acp.1053
- Poole DA, & White LT (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology*, 27(6), 975–978. doi:10.1037//0012-1649.27.6.975
- Quas JA, Malloy LC, Melinder A, Goodman GS, D'mello M, & Schaaf J (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports. *Developmental Psychology*, 43(4), 823–837. doi:10.1037/0012-1649.43.4.823 [PubMed: 17605517]
- Roediger HL, & Butler AC (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. doi:10.1016/j.tics.2010.09.003 [PubMed: 20951630]
- Rowland CA (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. doi:10.1037/a0037559 [PubMed: 25150680]
- Sekeres MJ, Bonasia K, St-Laurent M, Pishdadian S, Winocur G, Grady C, & Moscovitch M (2016). Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learning & Memory*, 23(2), 72–82. doi:10.1101/lm.039057.115 [PubMed: 26773100]
- Scriver E, & Safer MA (1988). Eyewitnesses show hypermnesia for details about a violent event. *Journal of Applied Psychology*, 73(3), 371–377. doi:10.1037//0021-9010.73.3.371 [PubMed: 3182548]
- Shaw III JS, & McClure KA (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, 20(6), 629–653. doi:10.1007/bf01499235
- Smith RM, Powell MB, & Lum J (2009). The relationship between job status, interviewing experience, gender, and police officers adherence to open-ended questions. *Legal and Criminological Psychology*, 14(1), 51–63. doi:10.1348/135532507x262360
- Tuckey MR, & Brewer N (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, 9(2), 101–118. doi:10.1037/1076-898x.9.2.101 [PubMed: 12877270]
- Turtle JW, & Yuille JC (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology*, 79(2), 260–271. doi:10.1037//0021-9010.79.2.260 [PubMed: 8206816]
- Wells GL, & Leippe MR (1981). How do triers of fact infer the accuracy of eyewitness identifications? Using memory for peripheral detail can be misleading. *Journal of Applied Psychology*, 66(6), 682–687. doi:10.1037/0021-9010.66.6.682

- Wheeler MA, & Roediger HL (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3(4), 240–246. doi:10.1111/j.1467-9280.1992.tb00036.x
- Yuille JC, & Cutshall JL (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology*, 71(2), 291–301. doi:10.1037//0021-9010.71.2.291 [PubMed: 3722079]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

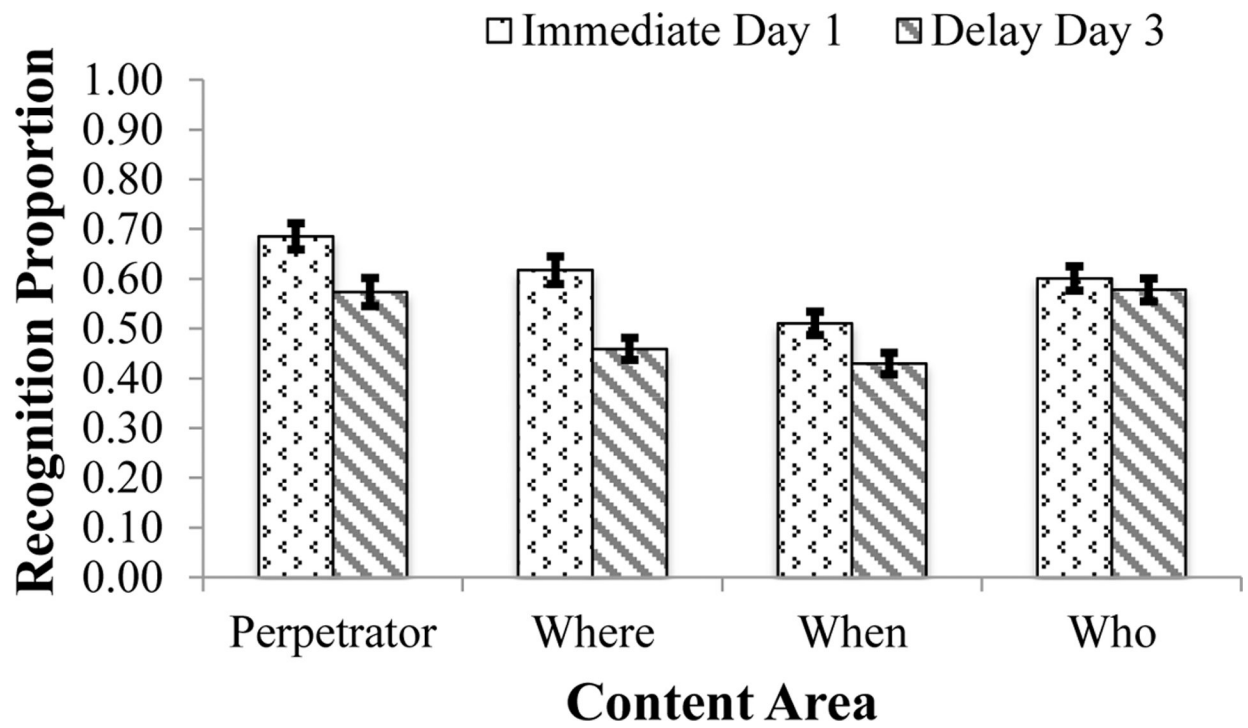


Figure 1.
Proportions of items correct for Immediate Day One vs.
Delay Day Three. Error bars denote standard error.

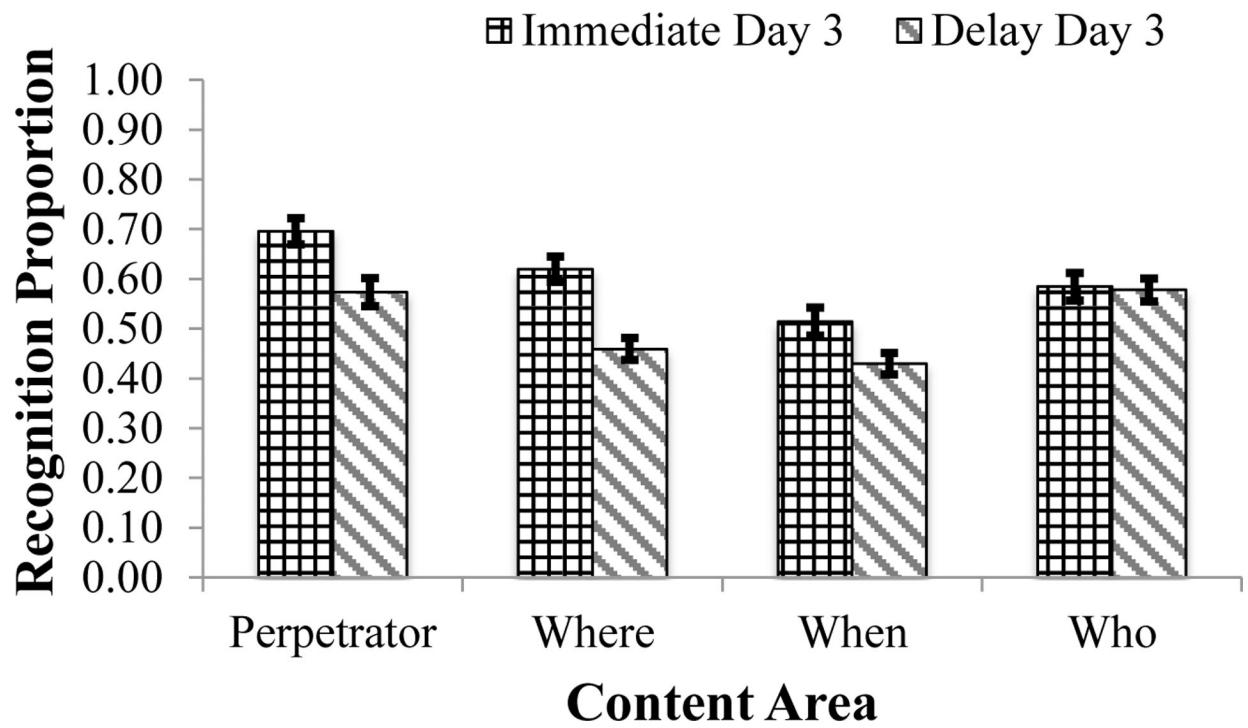


Figure 2.
Proportions of items correct for Immediate Day Three vs.
Delay Day Three. Error bars denote standard error.

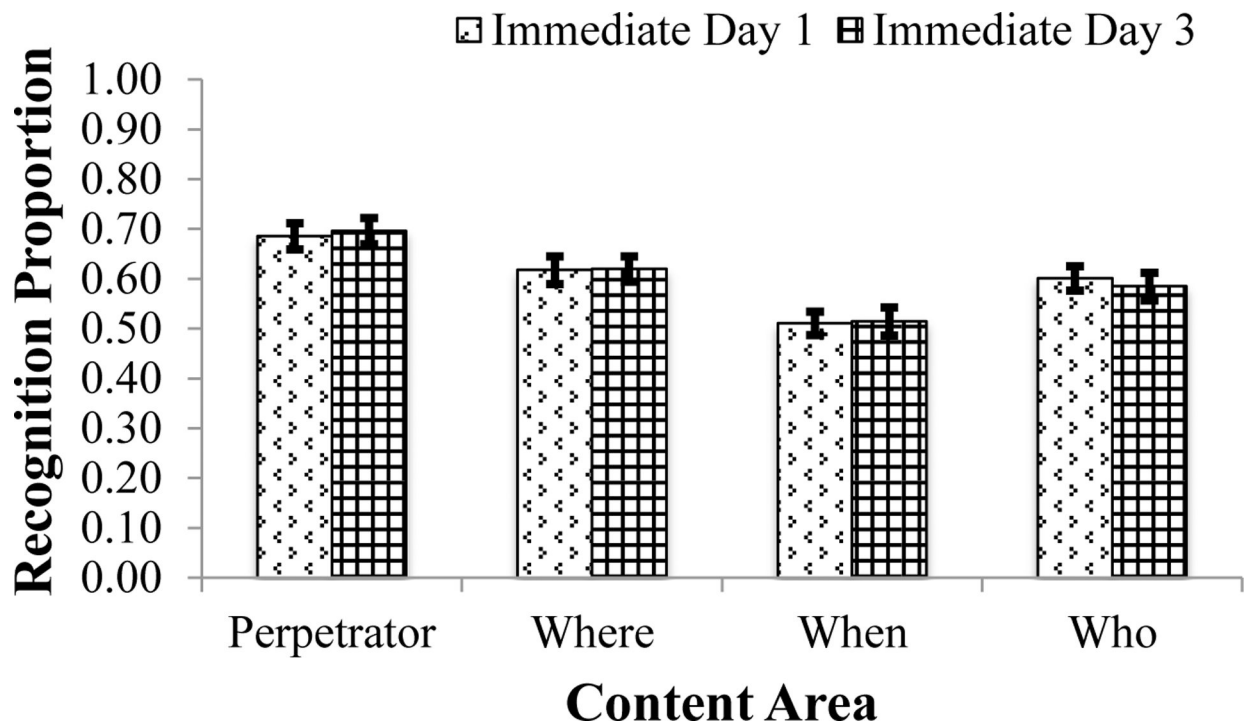


Figure 3. Proportions of items correct for Immediate Day One vs. Immediate Day Three. Error bars denote standard error. Note that this figure, unlike Figures 1 and 2, represents only within-subject data.

Table 1

Correlations between 'Perpetrator' and 'Where,' 'When,' and 'Who' Information

	Delay Day 3	Immediate Day 1	Immediate Day 3
'Perpetrator' – 'Where'	$r = 0.33$ $r^2 = 0.11$ $p = .009$	$r = -0.12$ $r^2 = 0.01$ $p = .385$	$r = -0.02$ $r^2 < 0.01$ $p = .858$
'Perpetrator' – 'When'	$r = 0.25$ $r^2 = 0.06$ $p = .059$	$r = 0.17$ $r^2 = 0.03$ $p = .232$	$r = 0.31$ $r^2 = 0.10$ $p = .021$
'Perpetrator' – 'Who'	$r = 0.39$ $r^2 = 0.15$ $p = .002$	$r = 0.38$ $r^2 = 0.15$ $p = .004$	$r = 0.31$ $r^2 = 0.10$ $p = .022$
'Perpetrator' – 'Where', 'When' and 'Who'	$r = 0.44$ $r^2 = 0.19$ $p < .001$	$r = 0.21$ $r^2 = 0.05$ $p = .122$	$r = 0.31$ $r^2 = 0.10$ $p = .020$

Note. r = correlation coefficient, r^2 = coefficient of determination, p = two-tailed probability

Table 2

Correlations between 'Where', 'When' and 'Who' Information

	Delay Day 3	Immediate Day 1	Immediate Day 3
'Where' – 'When'	$r = 0.38$	$r = -0.07$	$r = 0.08$
	$r^2 = 0.14$	$r^2 < 0.01$	$r^2 < 0.01$
	$p = .003$	$p = .604$	$p = .583$
'Where' – 'Who'	$r = 0.24$	$r = 0.18$	$r = 0.32$
	$r^2 = 0.06$	$r^2 = 0.03$	$r^2 = 0.11$
	$p = .061$	$p = .189$	$p = .017$
'When' – 'Who'	$r = 0.36$	$r = 0.13$	$r = 0.20$
	$r^2 = 0.13$	$r^2 = 0.02$	$r^2 = 0.04$
	$p = .005$	$p = .337$	$p = .156$

Note. r = correlation coefficient, r^2 = coefficient of determination, p = two-tailed probability

Table 3

Percentages Correct, Incorrect, I do not know, and None of the Above by Category

	Perpetrator	Where	When	Who
Delay				
% Correct	57.38	45.93	42.96	57.78
% Incorrect (all Incorrect Responses)	42.62	54.07	57.04	42.22
% "I do not know"	16.67	21.11	23.15	17.04
% "None of the above"	7.62	3.70	9.07	5.00
% Foil	18.33	29.26	24.81	20.19
Immediate Day 1				
% Correct	68.52	61.73	51.03	60.08
% Incorrect (all Incorrect Responses)	31.48	38.27	48.97	39.92
% "I do not know"	12.43	17.28	17.70	16.26
% "None of the above"	6.35	2.06	9.26	7.20
% Foil	12.70	18.93	22.02	16.46
Immediate Day 3				
% Correct	69.58	61.93	51.44	58.44
% Incorrect (all Incorrect Responses)	30.42	38.07	48.56	41.56
% "I do not know"	10.32	15.02	16.87	18.72
% "None of the above"	5.56	2.06	7.20	4.53
% Foil	14.55	20.99	24.49	18.31

Table 4

Percentages Incorrect Omitting “I do not know” Responses by Category

	Perpetrator	Where	When	Who
Delay				
% Incorrect Omitting % “IDK”	31.14	41.78	44.10	30.36
Immediate Day 1				
% Incorrect Omitting % “IDK”	21.75	25.37	38.0	28.26
Immediate Day 3				
% Incorrect Omitting % “IDK”	22.41	27.12	38.12	28.10

Note. % “IDK” indicates percentage of “I do not know” responses