2019

EW Wasserman et al Supplementary Material Bystander Reporting (Supplementary material for Wasserman et al.)

Ralph Miller
rmiller@binghamton.edu

Supplementary Material for:

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

This file contains:

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). Participants were given partial credit of a course requirement for taking part in each half of the experiment. 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 later. 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).

Stimulus video. The experiment started with on-screen instructions prompting participants to attend to a short video on the computer monitor. The video began with two patrons sitting and a cashier standing towards the back of a café. There was a white board towards the center right side of the screen that listed prices of goods. A few seconds into the video, a woman and male friend walk into the café and sit down at a table in the foreground of the scene. After sitting for a few moments, the couple walks towards the back of the café to place an order, leaving their belongings at the table. About 35 seconds into the video, a man who was sitting towards the front right corner of the café walks through the foreground of the scene, snatches the woman's unattended purse, and leaves the café. A few moments later, the victim and friend notice the purse is gone and begin to search for it near their table. Some seconds later, they too exit the café, seemingly in search of the purse. The video was in color, was silent, lasted 1 minute, 6 seconds, and took up 85% of the height of the screen and 85% of the width of the screen. Participants sat with their eyes approximately 0.45 m from the center of the screen. The screen's dimensions were 54 x 30 cm and the resolution was 2250 x 2450 pixels. Instructions appeared in black, Courier font style, text size 40, on a gray background. Next to the computer were a #2 pencil and a folder containing an answer sheet (Scantron) and a packet of instructions. The packet informed participants about how to anonymously fill out the Scantron in order for the experimenters to later match the Scantrons from Day One to the Scantrons from Day Three, so

that the participants' responses in each session could be linked. We used a forced-choice procedure for the memory test because it facilitated 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 concerning 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 crime scene).'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 (i.e., drinking coffee or texting). 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 Scantrons from each participant. See Appendix for representative test questions.

Pilot study. 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. We tested 72 participants, half of whom were assigned to an Immediate condition (tested immediately after watching the video of a purse being stolen and the same participants tested again 48 hours later) and half of whom were assigned to a Delay condition tested for the first time 48 hours after watching the video. The set of questions consisted of 49 target questions from four main content areas (‘Perpetrator’, ‘When’, ‘Where’ and ‘Who’). There was only one order of questions, and one question from each of the four content areas appeared on a page at least until questions concerning a given content were exhausted. Informed by analysis of the data from only participants who were tested immediately after watching the purse-snatching video, we eliminated potential questions for the experiment reported here in which 10% or fewer participants answered the question correctly, or 90% or more of participants answered the question correctly. We additionally removed questions that had very low point bi-serial correlations (i.e., less than -0.10) with other questions within a content area, such that performance on one question in a content area did not accurately predict performance on other questions in that content area. We then matched the three content areas for the quantity of items by removing questions with the lowest point bi-serial correlations, which we recalculated having

removed items using the previous criteria. In total, we removed three items from the ‘Perpetrator’ category and four items from each of the ‘Who’, ‘What’ and ‘When’ categories.

Order of test questions and participant instructions. The selected 34 questions for the present experiment included nine questions in each context category, 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. The first page of the question packet provided instructions on how to fill out the Scantron in order for us to pair Scantrons from the two sessions representing the same participant. For the present experiment, on Day One, the following pages of the packet contained questions regarding the video for only the Immediate Group. Specifically, each page of the question packet contained one question from each content area, except for two pages that did not include a ‘Perpetrator’ question due to there being fewer perpetrator questions. The order of the four types of questions from each content area on each page was randomized. No questions from the same content area appeared in immediate succession (e.g., if the last question on a page was a ‘Perpetrator’ question, then the first question on the next page was from any content area except ‘Perpetrator’). On Day One, after the first page, the Delay Group was presented with a page informing them that this part of the experiment was complete. We had six different versions of the test that contained different pseudorandomized orders of questions. The Immediate group was given the same questions, in a different order, on their Day One and Day Three tests. On Day Three, the first page of the packet was identical to the one that both groups had seen 48 hours earlier. The subsequent pages contained the 36 questions regarding the video. The last page of the packet thanked participants for taking part in the experiment, informed them that we were studying memory of different types of information, and asked them not to discuss the experiment with anyone else. Participants

were randomly assigned to cubicles constrained by counterbalancing between groups, and were seated in the same cubicle for both parts of the experiment.

Procedure

All participants completed the task in individual cubicles devoted to computer-based psychology experiments. Upon arriving, the experimenter reminded participants that this was the first part of the study, and that they were to return in two days. 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. In addition, participants were asked not to use cell phones during the experiment, nor to discuss the experiment with anyone else during or after the experiment.

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.

Statistical Analysis

Participant accuracy was determined by calculating the mean number of questions correct for each content category. First, a 2 x 4 mixed-design analysis of variances (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 x 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 x 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 differences in 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 for each 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.