Journal of Applied Research in Memory and Cognition 9 (2020) 242–253

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

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

Future Planning May Promote Prospective False Memories

Anna-Lisa Cohen[a,*], Michael J. Silverstein[a], Daniel G. Derksen[b], Zachariah I. Hamzagic[b], Daniel M. Bernstein[b], and D. Stephen Lindsay[c]

[a] Yeshiva University, United States
[b] Kwantlen Polytechnic University, Canada
[c] University of Victoria, Canada

Prospective memory (PM) involves remembering to execute future intentions. Pairs played a word game (Taboo) with an embedded PM task. In Taboo, one player (clue giver) must get their partner (clue guesser) to say aloud a target word (e.g., ROOF) by offering clues such as "home" without saying certain taboo words (e.g., fiddler, house). The PM task required clue givers to remember to say specific clue words if any predesignated PM targets appeared during the game (e.g., "If ROOF is a target, use 'home' as a clue"). Before playing Taboo, participants learned that half the PM targets did not have to be executed (cancelled intention) and half did (active intention). One day after playing, participants rated how clearly they remembered executing PM task and targets that had never appeared in the Taboo game. Memory ratings were higher for words from active intentions relative to cancelled intentions, evidencing false prospective memory.

General Audience Summary

Remembering to execute a future action (e.g., remembering to take medication) is known as *prospective memory* (PM). We hypothesized that forming an intent to execute a future task increases thoughts about that action, which makes one more likely to later misattribute memories of thoughts of that action as confirmation that the task had been performed. Participants played a charades-like word game known as *Taboo* with an additional prospective memory task that required them to remember to execute a future action. In Taboo, a player (the clue giver) must try to get their partner (the clue guesser) to say aloud a target word (e.g., TREE) by offering clues such as "found in a forest" without saying certain taboo words (e.g., green, branch, bark). The embedded PM task required clue givers to remember to say a *specific* clue word (e.g., tall) if the predesignated Taboo target (TREE) appeared during the game. Each participant took turns being the clue giver and clue guesser. Participants learned 10 of these PM intentions to a criterion; they then learned that a portion of the PM intentions had to be executed

Anna-Lisa Cohen, Department of Psychology, Yeshiva University; Michael J. Silverstein, Department of Psychology, Yeshiva University; Daniel G. Derksen, Department of Psychology, Kwantlen Polytechnic University, Zachariah I, Hamzagic, Department of Psychology, Kwantlen Polytechnic University; D. Stephen Lindsay, Department of Psychology, University of Victoria Daniel M. Bernstein, Department of Psychology, Kwantlen Polytechnic University.

Michael J. Silverstein is now at Department of Psychology, Drexel University. Daniel G. Derksen is now at Department of Psychology, Simon Fraser University.

We thank the Canada Research Chairs Program (950-232078) for funding this work. A-LC and MJS conceived of the initial experiment during an undergraduate research methods class. These ideas were further developed in an undergraduate honors thesis by MJS supervised by A-LC. Authors DMB and DSL made substantial contributions to study conception and design. A-LC and DMB were the principle investigators for the initial experiment and the replication, respectively. MJS, DGD, and ZIH oversaw data collection. MJS, A-LC, DGD, and ZIH analyzed the data, and DMB and DSL verified analytic methods and assisted with results interpretation. The manuscript was written by several individuals: MJS wrote the Method and a portion of the Results for Experiment 1; DGD and ZIH wrote the Method and a portion of the Results for Experiment 2; and A-LC wrote the remaining sections with input from all other authors. All authors reviewed and revised the manuscript, and approved the final version. The authors would like to thank Natan Weissman, Jack Chen, Aaron Kong, and Joey Chhina for assistance with data collection and entry.

* Correspondence concerning this article should be addressed to Anna-Lisa Cohen, Department of Psychology, Room C05H, Belfer Hall, Yeshiva University. 2495 Amsterdam Ave, New York, NY, 10033, USA. Contact: acohen11@yu.edu.

E-mail address: acohen11@yu.edu (A.-L. Cohen).

(active intention), while the rest did not have to be executed (cancelled intention). One day later, participants returned to complete a post-experiment questionnaire. Clue givers rated how clearly they remembered successfully executing their PM intention for PM targets that never appeared in the Taboo game the day before (see video of procedure: https://osf.io/8mfp3/). Results showed participants were more likely to (falsely) rate words from the active intention as having higher memory clarity relative to cancelled intention PM words. We theorize that in forming a future intention, we create a representation of action performance. It is this representation that we may later mistake as a memory of executing that action (cf. Grèzes and Decety, 2001).

Keywords: Prospective memory, False memory, Future intentions

In the popular Broadway musical *Hamilton*, the title character sings: *"I imagine death so much it feels more like a memory."* The idea is that when a person engages in enough thought about a future event, the imagined event may feel like an event that already occurred. Indeed, Einstein, McDaniel, Smith, and Shaw (1998) speculated that thoughts about a future action can later be mistaken as memory of the completed action (see also Johnson & Sherman, 1990).

Remembering to execute future actions at an appropriate future time (e.g., remembering to turn off the stove or take medication) is called *prospective memory* (PM; Einstein & McDaniel, 1990; see Cohen & Hicks, 2017; Kliegel et al., 2008, for reviews). In typical laboratory PM studies, participants perform an ongoing task (e.g., a lexical decision task) and, in PM conditions, participants receive an additional PM instruction to respond differently if a pre-specified target event happens in the future (e.g., "Press the spacebar button if an animal-name word is presented"). These paradigms are sensitive to participants' ability to encode an intended future action and then to act on that intention at the appropriate future time. Research suggests that remembering to perform an action in the future (such as turning off the oven) leads one to create a set of mental images of performing that action. That is, there can be an intentionality to imagery (see Harman, 1998). As Grèzes and Decety (2001) stated, "the subject imagines himself performing a given action" through "mental simulation of action, defined as mental rehearsal of a motor act" (p. 2).

Forgetting to turn off a stove or take medication can have devastating consequences, which has motivated systematic research that has produced a nuanced understanding of prospective memory. Researchers categorize errors as either commission or omission errors. Commission errors occur when one erroneously executes the PM response despite having already completed the task, such as mistakenly taking a second dose of medicine (Pink & Dodson, 2013; Scullin, Bugg, McDaniel, & Einstein, 2011; Scullin, Bugg, & McDaniel, 2012; Walser, Fischer, & Goschke, 2012). Omission errors are the most common type of PM error and occur when there is a failure to execute a PM response when a PM cue is encountered (e.g., forgetting to take one's medication at the appointed time). One possible reason for failing to execute a PM response may be that one mistakenly believes that it was already executed. That is, I fail to take my pill because I assume that I already took it. Perhaps the more easily an intended future action pops into mind, the more likely one will erroneously infer that it has already been successfully completed.

Research on the *intention superiority effect* shows that information related to future intentions is more easily retrieved (e.g., faster reaction times and higher accuracy) compared to information that is not future oriented (Goschke & Kuhl, 1993; Marsh, Hicks, & Bink, 1998; Marsh, Hicks, & Bryan, 1999). After an intension to act has been carried out, memories of that intention may pop to mind, Scullin and Bugg (2013) showed that PM commission errors (carrying out a PM action when it no longer needs to be completed) can result from spontaneous retrieval of the intention in an inappropriate context and failed cognitive control. Despite the obvious value of studying output monitoring (i.e., memory for one's previously completed actions) in the context of prospective memory, there is still relatively little research on the topic (Ball, Pitães, & Brewer, 2018).

Einstein et al. (1998) described the specific challenges of managing prospective memory tasks due to the inherent metacognitive challenges. They noted that when a task is habitual and performed repeatedly, it increases the likelihood of internal source monitoring errors such that there is difficulty distinguishing between thoughts of a prospective action and memories of having performed that action previously. Thus, the more easily a postponed intention spontaneously comes to mind, the more likely the person will mistakenly think that they have completed the postponed intention. We explored this type of PM error in the current study.

We wondered whether young healthy participants can be led to believe that they executed a prospective memory action that they had not executed. This is akin to the type of false memory implantation studies described in the retrospective memory literature (see Bernstein, Scoboria, Desjarlais, & Soucie, 2018; Gallo, 2010; Laney & Loftus, 2013; Schacter, Guerin, & St. Jacques, 2011). Experiment 1 examined this research question by using a false memory implantation administered after a 24-h delay. Experiment 2 was a pre-registered replication of Experiment 1 (Open Science Framework; https://osf.io/4h8gx/)

Experiment 1

Participants played a charades-like word game known as *Taboo* (TABOO® & ©2019 Hasbro, Inc. Used with permission; https://products.hasbro.com/) that included a prospective memory task requiring them to remember to execute a future action. In Taboo, players (e.g., the clue giver) try to get their partner (e.g., the clue guesser) to say aloud a target word (e.g., TREE) by offering clues such as "found in a forest" while avoiding saying "taboo" words (e.g., green, branch, bark). The embedded prospective memory task required clue givers to remember to say a *specific* clue word (e.g., tall) if the predesignated Taboo target (TREE) appeared during the game. Each participant took turns being the clue giver and clue guesser. Participants learned 10 of these PM intentions to a criterion. They then learned that when they were the clue giver, 5 of these PM intentions were relevant and had to be executed (active intention) while the other 5 were not relevant and did not have to be executed (cancelled intention). Twenty-four hours later, participants returned to complete a post-experiment questionnaire. The independent variables were Appearance (whether PM targets appeared or did not appear during the Taboo game) and Relevance (whether the PM target had to be executed or it was cancelled). The dependent variable was participants' self-reported memory clarity ratings 24 h later for having said the associated PM clue word (e.g., "above") when the PM target word (e.g., ROOF) appeared during the Taboo game (1 = Not at All; 7 = Very Clearly). We hypothesized that participants would rate relevant items (active intentions) as having higher memory clarity compared to not relevant items (cancelled intentions), even for items that never appeared during the PM task. A video of the entire procedure can be downloaded here: https://osf.io/8mfp3/.

Method

Participants. Forty-eight Yeshiva University undergraduate students (47 males and one female) were tested. Eight participants were dropped from the analysis; Three failed to return for the follow-up questionnaire, three participants' data were excluded due to experimenter error, one participant stated during debriefing that he haphazardly circled answers on the questionnaire, and one participant was unable to memorize all 10 word pairs. The participants received either research participation credit for their psychology class or $10 compensation. The final sample was N = 40.

Design and materials. Figure 1 shows how the overall experiment unfolded over time. Figure 2 depicts the actual game of Taboo. The experiment was a 2 Relevance (relevant, not relevant) × 2 Appearance (appear, not appear) repeated measures design with both factors as within-subject factors. "Relevant" word pairs meant that they were a to-be-performed or active intention, and "not relevant" word pairs were a cancelled intention meaning that they did not have to be performed during the Taboo game. "Appeared" word pairs meant that they occurred during the Taboo game, and "not appeared" word pairs meant that they never occurred during the Taboo game. The self-reported memory clarity ratings were conducted 24 h later (plus or minus 1 h) during the completion of a post-experiment questionnaire.

We used 118 different "Taboo" cards (see Appendix A). Participants received 11 different cards during each Taboo round. Taboo cards are double sided with four colors. On one side, the card is blue and yellow; on the other side, it is purple and orange. We only used words from the blue and purple sections of the cards; therefore, we refer to the cards as blue or purple. There is no difference between the blue or purple cards, except that the two sides do not contain any of the same target words. Participants received a sheet of paper containing 10 randomly ordered word pairs (see Appendix B). Using the University of South Florida Free Association Norms database, we chose word pairs that had forward cue-to-target strength association ranging from .070 to .180, and backward cue-to-target strength association ranging from .012 to .096 (Nelson, McEvoy, & Schreiber, 1998). Twenty-four hours later, participants returned to the lab to complete a post-experiment questionnaire (See Appendix C). As a manipulation check and to evaluate memory for instructions during the experiment, participants answered several questions about their memory of the Taboo game and critical questions about their memory clarity for having said the appropriate cue word when a target word occurred during the Taboo game.

There were four possible word pair types: (a) relevant-appeared items were relevant words that appeared during the Taboo game (i.e., part of the prospective memory task); (b) not relevant-appeared items were not relevant word pairs that appeared during the Taboo game; (c) relevant-not appeared items were relevant word pairs that did not appear during the game; and finally, (d) not relevant-not appeared items were not relevant word pairs that did not appear during the Taboo game. Of the 10 possible word pairs, only eight appeared in the follow-up questionnaire. All three relevant-appeared items appeared in the questionnaire, as did the two relevant-not appeared items and the two not relevant-not appeared items. However, the questionnaire asked participants to rate their memory clarity for only one of the three not relevant-appeared items. Although there were three not relevant-appeared items that appeared in the Taboo game, in the follow-up questionnaire, we randomly asked about memory clarity for only one item. We did this because we thought it would make it less likely that participants would guess the hypothesis. If we had participants report their memory clarity for all 10 word pairs, then we reasoned that participants might be more likely to realize that we were asking about their target-word performance for items that never appeared during the game. Importantly, the critical comparison between relevant and not relevant items that never appeared in the Taboo task had an equal number of items.

Procedure. Participants played Taboo in pairs. One player (clue giver) had a Taboo card with a target word on it (e.g., FLASHLIGHT) that they described to the other player without saying "FLASHLIGHT" and without using any "taboo" words (e.g., hand, shine, dark). The other player (clue guesser) tried to guess the target word based on the clues as quickly as possible (see Appendix A). Each participant took turns serving as both the clue giver and clue guesser.

Day 1. On Day 1, participants met their game partner and learned how to play the Taboo game. Participants also studied and memorized a list of words that served as the prospective memory instructions.

Learning to play Taboo. Participants sat on opposite sides of a table. We explained that the clue giver would draw a card with a target word printed at the top (e.g., TREE). The objective

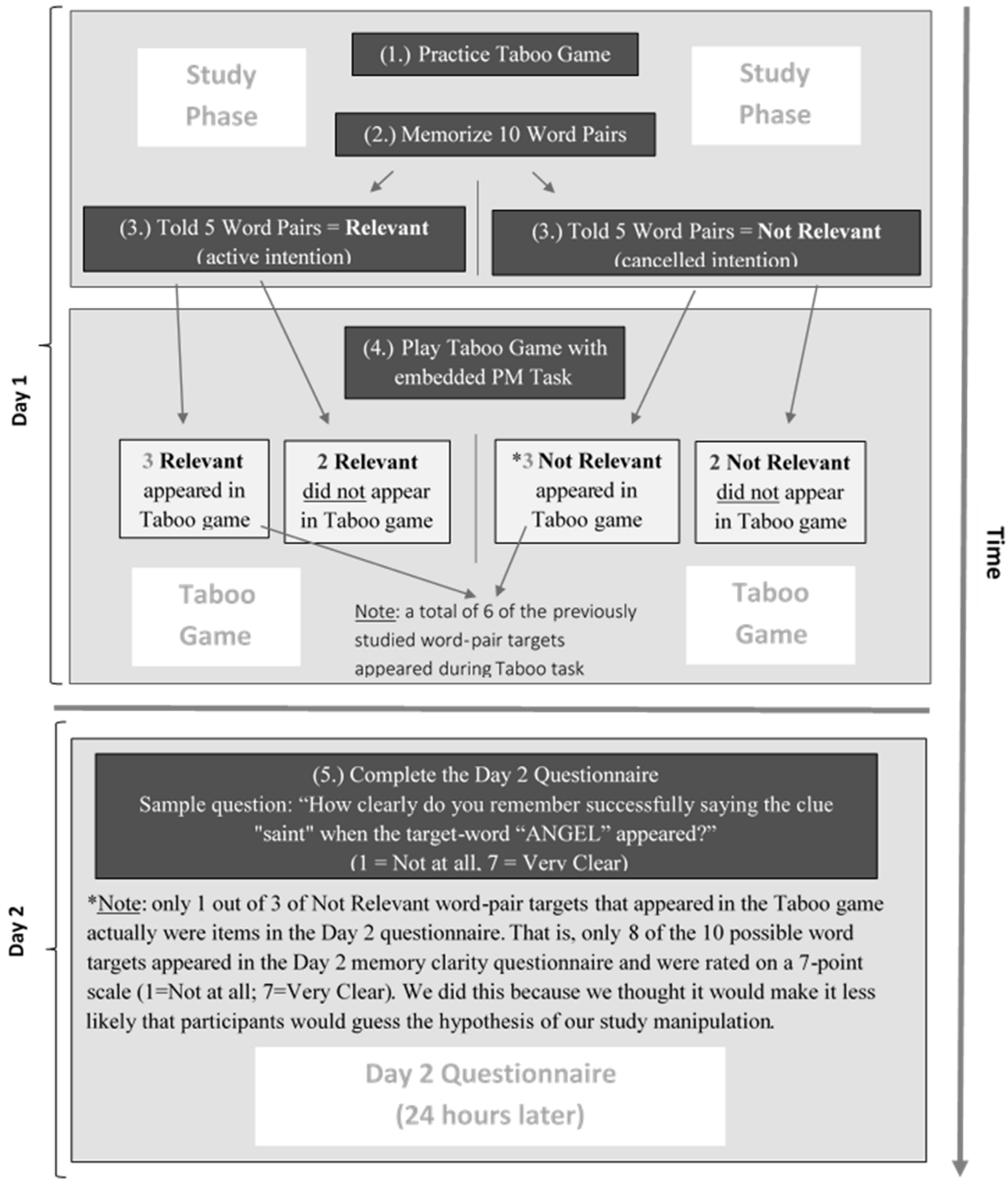

Figure 1. Schematic of experimental procedure.

of the game was to get the clue guesser to say the target word (e.g., TREE) by giving clues (e.g., "found in the forest") without saying taboo words (e.g., branch, trunk, leaves, green) which were listed on the playing card. We told participants that they would take turns being the clue-giver and clue-guesser. Each round lasted 70 s. Participants could spend a maximum of 12 s per card before moving to the next card. We chose this rate to increase difficulty while exposing participants to multiple cards during each round. We told participants that there was no penalty for incorrect guesses and that the clue-giver could decide to skip a card to proceed to the next one.

We showed participants a sample Taboo card to demonstrate examples of target words and taboo words (see Appendix A). As mentioned, Taboo cards are double-sided. One participant learned that they would use the blue side of the cards, and the other participant learned that they would use the purple side. (Recall that there was no difference between the blue and purple sides, aside from the fact that there were different target words

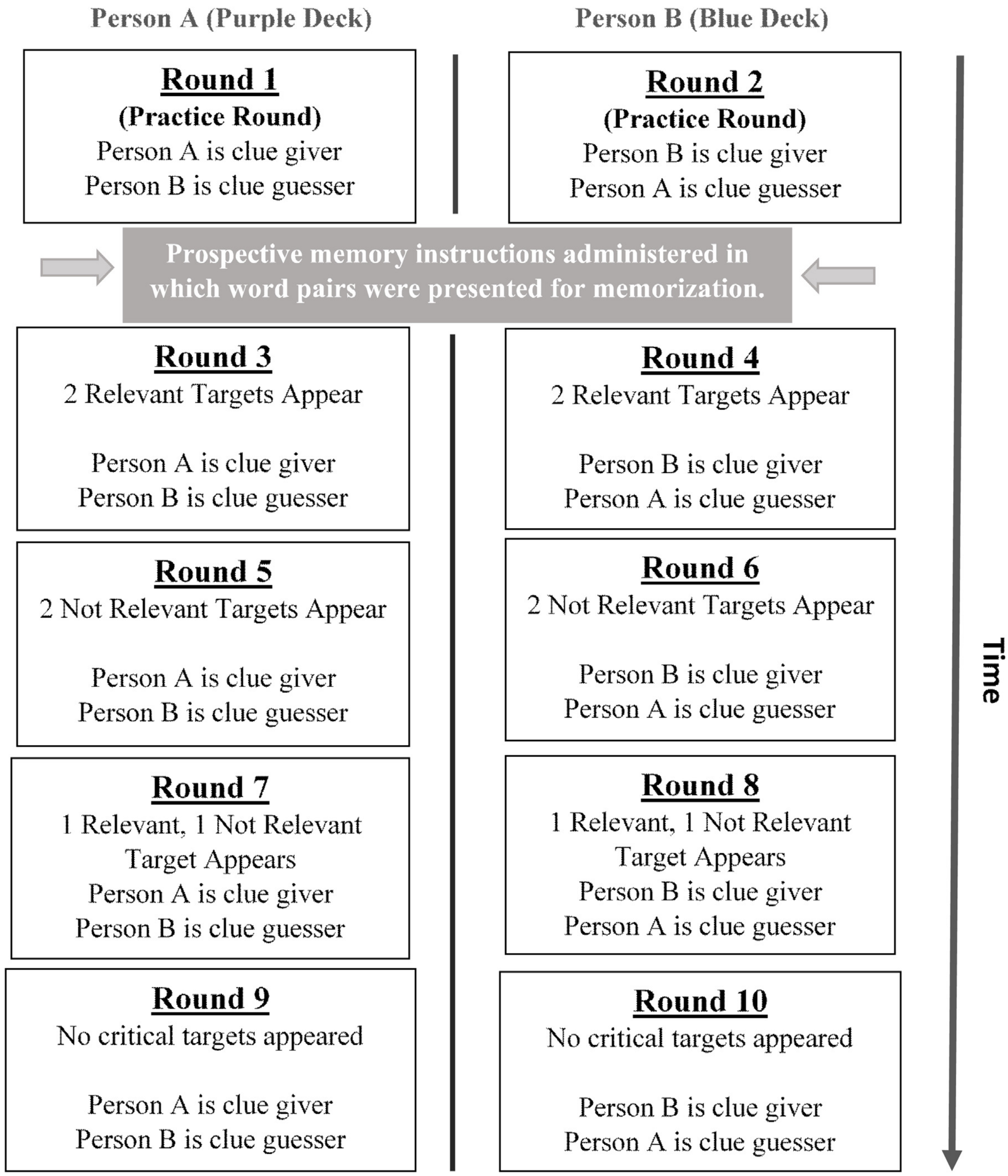

Figure 2. Schematic representing the Taboo game playing phase. There were 11 cards in each round and the critical items (e.g., relevant or not relevant targets) always appeared in the first 5 cards. There was a maximum of 12 s allowed for each Taboo card and 1 min and 10 s for each round.

on all the cards.) After participants learned the Taboo rules, the experimenter clarified any rules that were unclear. Next, participants played two practice Taboo rounds in which each participant practiced being clue-giver and clue-guesser. If, at any point during the Taboo rounds, the clue giver said one of the taboo words (i.e., one of the five words at the bottom of the card that the clue-giver cannot say), we had the clue-giver skip to the next card.

Learning the 10 word pairs. After the practice rounds, participants separated and had 4 min to memorize a list of 10 word pairs that would later serve as prospective memory targets and clue words. If after that time they were unable to correctly write down every word pair, they received 2 additional min to review the word pairs. Most participants were able to learn the word pairs in the 4 min of allotted time. The few participants who had trouble generally forgot 1 or 2 word pairs, and successfully wrote down all 10 word pairs after receiving two additional min.

Learning relevant (PM) and not relevant (cancelled) word pairs. After participants learned the 10 word pairs, they learned that five of the word pairs were relevant and had to be executed (active intention) while the other five were not relevant and did not have to be executed (cancelled intention). For relevant word pairs (e.g., BEACH-coast), the first word of each pair (e.g., BEACH) was a PM target word that could be one of the targets on a Taboo card. We then instructed participants that if any of the five relevant PM targets appeared as a Taboo target during the Taboo game, then they should remember to state the second word of the word pair (e.g., coast) as a clue to help their partner guess the target word. This relevant item instruction closely mirrored the typical instructions in standard PM laboratory tasks in which a PM cue (in our case BEACH) must trigger the delayed intention of remembering to say the word "coast." Participants learned that they could say other clues for the relevant target item, but they should make sure to say the associated word as one of their clues. Participants did not have to say the associated cue word for not-relevant items.

Participants then used a computerized slide-show presentation to rehearse and learn which word pairs were relevant. The word pairs appeared for three s sequentially, with a plus sign "+" separating each pair. Participants learned that the relevant word pairs would appear in green and the not relevant word pairs would appear in red. They also learned that under each word pair a Taboo card for the target-word would appear. Participants watched the presentation of the word pairs twice. They then received a list of the word pairs and had to indicate which word pairs were relevant. If participants were unable to accurately report which word pairs were relevant, they watched the presentation of the word pairs twice more. All participants were able to learn which word pairs were relevant after one or two attempts of watching the presentation. Although there was no filled delay between learning the prospective memory instructions and resuming the Taboo game, there was a built-in natural delay because the experiment occurred in pairs where both participants had to re-enter the testing room to resume playing the Taboo game. At this time, participants were asked not to speak to each other; experimenters offered no further reminders about the PM task.

Playing the remaining rounds of taboo with the embedded PM task. After participants completed the practice rounds, they played eight additional rounds of Taboo. In total, participants played 10 rounds of Taboo. Each round involved a deck of 11 Taboo cards. To ensure that participants saw every critical relevant PM or not relevant PM card, critical cards always appeared in the first six cards of the deck. See Figure 2 for a schematic of the Taboo game with the embedded PM task. Throughout the game, we tracked performance and noted if the clue-giver remembered to successfully say the associated word for prospective memory items.

Day 2. Approximately 24 h later, participants returned to complete the follow-up questionnaire that included the key questions probing their memory for the critical word pairs (Appendix C). They were instructed to answer questions as best they could. As a manipulation check and to evaluate memory for instructions during the experiment, participants answered several questions regarding their memory of the Taboo game. They also answered critical questions about their memory clarity for having said the appropriate clue word when a target word occurred during the Taboo game. A sample of a critical question was the following: How clearly do you remember successfully saying the clue "saint" when the target-word "ANGEL" appeared? Ratings on these questions were evaluated using a 7-point Likert scale (1 = Not at All, 7 = Very Clearly). Critically, participants were asked to rate their memory clarity for both relevant (active intentions) and not relevant (cancelled intentions) word pairs that never appeared the previous day in the Taboo game. Asking participants to rate their memory for items that never appeared was our false memory implantation, because it gave participants the impression that all items on the questionnaire were there because those items had occurred the previous day during the Taboo game. After completing the questionnaire, participants received debriefing and compensation.

Results

All analyses used an alpha level of .05. Although participants were tested in pairs, for all analyses we treated individual subjects' data as the unit of analysis. As described above, we had two groups based on the two decks that participants received (i.e., blue or purple). We collapsed across this variable because performance did not differ significantly between them (all $ps > .62$).

To confirm that participants appropriately encoded the instructions for relevant and not relevant word pairs, we analyzed prospective memory performance by examining the number of times that participants correctly followed instructions by (a) saying the associated clue word when they detected the relevant PM target word and (b) not saying the associated clue word for not relevant targets. Because three out of the five relevant-appeared and three out of the five not relevant-appeared items appeared during the Taboo game, prospective memory performance scores for relevant word pairs and not relevant word pairs could each range from 0-3. A t test for paired samples revealed that there was a significant difference between relevant-appeared prospective memory performance ($M = 2.50$, $SD = 0.88$) and not relevant-appeared prospective memory performance ($M = 2.08$, $SD = 1.00$); $t(39) = 3.08$, $p = .004$, $d = 0.40$). This indicates that participants, on average, were more likely to correctly say the clue word for the Relevant word pairs than they were to say the not relevant word pairs, suggesting that participants accurately understood the instructions for relevant and not relevant word pairs.

Next, we compared participants' memory clarity on Day 2 for having said the associated clue words when the PM target words appeared during the Taboo game. Participants rated how clearly (1 = Not at All and 7 = Very Clearly) they remembered stating the associated clue word (e.g., "saint") when the target word (e.g., "ANGEL") appeared. Memory clarity ratings were analyzed by conducting a 2 (Relevance: relevant, not relevant) × 2 (Appearance: appear, not appear) repeated measures ANOVA with both

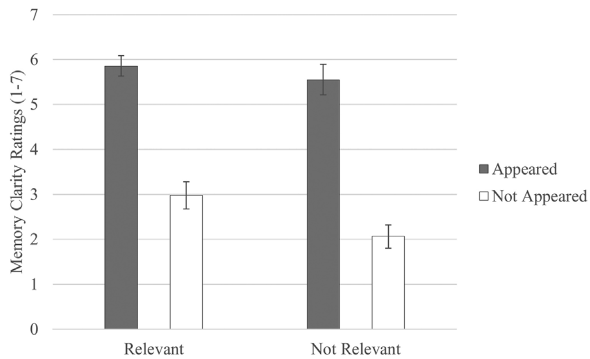

Figure 3. Experiment 1: Mean memory clarity ratings as a function of word pair type. Error bars represent standard errors of the mean.

Relevance and Appearance as within-subject factors. Results revealed a significant main effect of Relevance, $F(1,39) = 7.97$, $p = .007$, $\eta_p^2 = .17$, with memory clarity ratings being higher for relevant word pairs ($M = 4.42$, $SE = .221$) compared to not relevant word pairs ($M = 3.81$, $SE = .240$). There was a main effect of appearance, $F(1,39) = 116.03$, $p < .001$, $\eta_p^2 = .75$, with memory clarity ratings being higher for word pairs that appeared in the Taboo game ($M = 5.71$, $SE = .250$) compared to word pairs that did not appear ($M = 2.52$, $SE = .254$). The interaction between relevance and appearance was not statistically significant, $F(1,39) = 3.38$, $p = .07$, $\eta_p^2 = .08$. Although the interaction was not significant, we conducted simple effect tests as planned. Most critically for our hypothesis, Bonferroni-adjusted comparisons indicated that participants rated their memory clarity for relevant items that never appeared ($M = 2.98$, $SD = 1.90$) as 0.91 points higher compared to not relevant items that never appeared ($M = 2.06$, $SD = 1.65$; $p < .001$, 95% CI [.420, 1.405]). Conversely, memory clarity ratings of relevant and not relevant items that appeared in the Taboo game did not differ significantly ($p = .31$). Participants reported higher memory clarity for items that never appeared in the Taboo game when they had been told that they were relevant (active intention) compared to not relevant items (cancelled PM intention) see Figure 3). These two results indicated that participants showed prospective false memory as they rated memory clarity for having said a PM word as higher for active intentions compared to cancelled intentions that never appeared the previous day.

In the next analysis, we examined if memory clarity ratings differed as a function of whether or not participants remembered to successfully say the prospective memory clue response when a PM target word appeared during the Taboo game. A majority of participants said all of the associated relevant clue words correctly and only 10 participants met criteria for this analysis (i.e., said 1 or 2 of the 3 clue words when the relevant clue words appeared during the game). Therefore, we conducted a paired-samples t-test that showed a significant difference, $t(9) = 3.95$, $p = .003$; $d = 1.25$ between memory clarity ratings for those items in which participants correctly stated the PM clue when the relevant/appeared target appeared ($M = 6.50$, $SE = 0.34$) compared to memory clarity ratings for those who failed to state the correct clue word ($M = 2.90$, $SE = 0.74$).

Discussion

The current experiment examined the effects of encoding future intentions on false memory clarity of task performance. The results supported our hypothesis revealing that participants reported significantly higher memory clarity for the relevant prospective memory task word pairs compared to the not relevant word pairs. Therefore, word pairs that participants believed that they would have to execute in a subsequent task led to an enhanced status in memory which may have made it more likely that they falsely rated these words as having higher memory clarity when questioned 24 h later. We interpret this finding to mean that when participants encode information that they intend to enact in the future, this results in a richer and deeper encoding such that this to-be-enacted memory trace can be more easily mistaken as an already-enacted memory trace leading to higher memory clarity ratings. Thus, when participants were making retrospective evaluations of their performance 24 h later, they may have misattributed thoughts of the action to be memories of actual action performance, and therefore were more susceptible to developing a false memory of action performance.

To our knowledge, this study was the first to directly explore the effects of encoding prospective memory instructions on later false memory of task performance. Our results suggest that encoding a future intention may lead to a higher likelihood that the future event will be later falsely evaluated as having been completed.

Experiment 2

In Experiment 2, we conducted a pre-registered replication of Experiment 1 (https://osf.io/4h8gx/) with the following changes: First, during the follow-up questionnaire administered 24 h later, all participants wrote down all the word pairs that they could remember from the previous day as a check of their memory for the 10 word pairs. To examine whether this additional step influenced memory clarity of falsely saying the associated clue word, half the participants answered these questions *before* being asked their memory clarity of having said the associated clue words (i.e., the critical questions), and half answered these questions *after* they were asked the critical questions. This mixed-measures design allowed us to examine if being asked about the word pairs (before or after) had any effect on false memory of saying the clue words. Second, we included a measure of how well participants knew their Taboo partner in case it affected performance. Third, we replaced one potentially confusing Taboo card.

Method

Participants. We recruited 216 undergraduate students for the study. We subsequently excluded 93 participants: 25 due to attrition (i.e., failing to appear for Day 2); 51 due to memory errors (either failing to successfully memorize all 10 word pairs or failing to memorize which word pairs were relevant), five due to major language issues (inability to understand instructions), or other special needs which made completing the task impossible, 11 due to experimenter error, and two due to missing data. Our final sample was 122 participants (98 female; 13 male; 11 unreported).

A power analysis using G-*Power* 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) was pre-registered prior to data collection. The analysis indicated that a total sample of 98 participants was needed to detect a small to medium interaction effect in a 2 × 2 within-subject ANOVA ($f = .15$; Power = .95; $\alpha = .05$). There was a minor error with the pre-registered power analysis (we entered 4 measurements instead of 2). Therefore, we conducted a post-hoc power analysis to test whether our design in Experiment 2 had enough power to detect a small to medium interaction effect in a 2 × 2 within-subject ANOVA with a sample size of 122 participants. The analysis revealed that the power to detect an interaction effect of this size was 0.91 ($F = .15$; $\alpha = .05$). Thus, we had sufficient power to detect the interaction.

As in Experiment 1, the task demanded that participants memorize 10 word pairs, and whether these words should be said during the Taboo game. However, the sample in Experiment 2 was obtained from an open-registration University in which many students were multilingual and spoke English as a second language. Attrition was much higher than in Experiment 1. Additionally, a significant portion of the sample failed to successfully remember the word pairs and other vital information. Due to these considerations, we continued to run participants until we met our power requirements. As this was a replication, we did not anticipate the higher rate of attrition, difficulty remembering instructions and/or the word pairs, and language difficulties. Therefore, we added a self-report language proficiency questionnaire administered during the collection of demographic information before completing the study. We did not pre-register this prior to running the experiment, but later deemed it necessary given the unexpectedly high level of difficulty our participants exhibited while playing and understanding the Taboo game. Further demographic information on the participants can be found in the supplementary materials in Appendix D.

Design and materials. As in Experiment 1, we used a 2 (Relevance: relevant, not relevant) × 2 (Appearance: appear, not appear) repeated measures design. The primary dependent variable was self-reported memory clarity ratings of having said an associated PM clue word when the associated PM target appeared during the Taboo game (1 = Not at All to 7 = Very Clearly).

The materials used in the replication resembled those in Experiment 1, with a few exceptions. Several additions were made to the memory clarity survey. Namely, we included a free recall task in which participants wrote all the word pairs that they remembered from Day 1. Participants completed the free recall task either before or after the memory clarity survey, and the order of presentation was randomized across participants. We added the free recall task to see if relevant words were better recalled than not relevant words on Day 2. Finally, we added a friendship question so that participants could rate how well they knew the other participant; we do not discuss this measure further, because there were not enough friends who participated together (see Appendix D for further information).

Procedure. The procedure was identical to Experiment 1, except for the addition of the free-recall task on Day 2. First, participants completed informed consent, and then completed the demographic and language proficiency questionnaire. As in Experiment 1, participants learned that they would be playing the verbal charades game “Taboo” with a partner. On occasions when only one participant signed up for the study, or if one of the two participants failed to show up, an experimenter played the role of the participant’s partner during the Taboo game. As in Experiment 1, after completing the Taboo practice rounds, participants memorized the 10 word pairs, learned which were relevant or not relevant, and played the Taboo game with embedded PM task. Finally, participants returned 24 h later (plus or minus one hour) to complete the post-experiment questionnaire that included the memory clarity questions.

Results and Discussion

An alpha level of .05 was used for all analyses unless otherwise noted. We collapsed across deck type (blue vs. purple) as the pattern of results did not differ based on type of deck (all $ps > .22$). See the supplementary materials in Appendix D for more details on this analysis.

We analyzed prospective memory performance by examining the number of times that participants correctly said the appropriate clue word when a relevant PM target appeared compared to the likelihood of saying the not relevant word pair. As discussed prior, prospective memory performance scores for relevant and not relevant word pairs could each range from 0-3. A paired samples t-test revealed a significant difference between relevant-appeared targets ($M = 1.96$, $SD = 1.11$) and not relevant-appeared targets ($M = 1.58$, $SD = 1.21$); $t(120) = 3.41$, $p = .001$; $d = 0.31$, indicating that participants said the appropriate clue word significantly more often when the relevant-appeared items appeared. As in Experiment 1, this finding indicates participants understood the instructions for relevant and not relevant word pairs.

The dependent variable for all subsequent analyses, unless otherwise stated, was memory clarity ratings (1 = Not at All to 7 = Very Clearly). A 2 Relevance (relevant, not relevant) × 2 Appearance (appear, not appear) repeated measures ANOVA revealed a significant main effect of Relevance, $F(1,121) = 11.85$, $p = .001$, $\eta_p^2 = .09$, with memory clarity ratings being higher for relevant word pairs ($M = 5.04$, $SE = .15$) compared to not relevant word pairs ($M = 4.62$, $SE = .18$). There was a main effect of Appearance, $F(1,121) = 62.00$, $p < .001$, $\eta_p^2 = .34$, with memory clarity ratings being higher for appear word pairs ($M = 5.54$, $SE = .18$) compared to not appear word

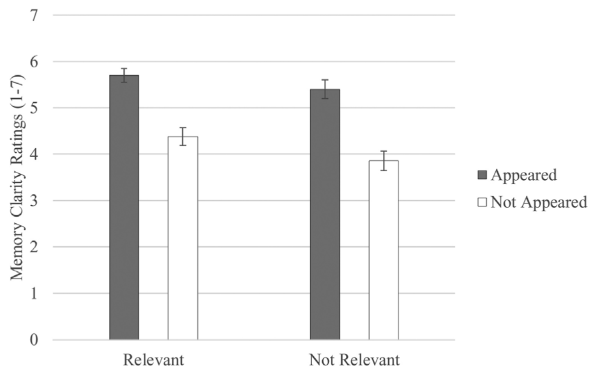

Figure 4. Experiment 2: Mean memory clarity ratings as a function of word pair type. Error bars represent standard errors of the mean.

pairs ($M = 4.12$, $SE = .19$). The interaction was not significant, $F(1, 121) = 1.12$, $p = .29$, $\eta_p{}^2 = .01$).

To better compare results from Experiment 2 with Experiment 1, we report the pairwise comparisons between relevant and not relevant word pairs for items that appeared and for relevant and not relevant items that did not appear in the Taboo game. Bonferroni-adjusted comparisons indicated that participants rated their memory clarity for relevant items that never appeared ($M = 4.38$, $SE = .19$) as 0.52 memory clarity points higher compared to not relevant items that never appeared ($M = 3.86$, $SE = .21$; p = .001, 95% CI [.226, .815]). In contrast to Experiment 1, participants rated their memory clarity .32 memory clarity points higher for Relevant items that Appeared ($M = 5.70$, $SE = .15$) compared to not relevant items that appeared in the Taboo game ($p = .04$, 95% CI [.012, .638]). The pattern of results can be seen in Figure 4.

As in Experiment 1, we conducted a paired-samples t-test to compare memory clarity ratings for those participants who remembered to state the PM clue word to those who failed to state the word. Results revealed a nonsignificant difference, $t(55) = .50$, $p = .62$, $d = 0.07$ between memory clarity ratings for those items in which participants correctly stated the PM clue when the relevant/appeared target appeared ($M = 5.05$, $SD = 2.78$) compared to memory clarity ratings for those who failed to state the correct clue word ($M = 4.84$, $SD = 2.07$). It would appear, based on this nonsignificant result, that performing the PM action (or not) did not significantly influence memory clarity ratings. This finding implies that participants are considering several sources of information to inform their memory clarity ratings above and beyond memory of actual performance 24 h earlier.

We took the subset of participants who completed the memory test on Day 2 before completing the questionnaire ($n = 59$) and compared the number of relevant words recalled with the number of not relevant words recalled. We made this division because many of the words that participants were asked to recall during the memory test were present on the questionnaire. As such, meaningful results could only be collected from those who completed the memory test before completing the questionnaire. Participants did not report significantly more relevant than not relevant words on the memory test, $t(58) = 1.04$, $p = .30$. As well, the order of the survey presentation (i.e., before or after the free recall task) did not influence memory clarity ratings (all ps $> .54$)

In order to explore more closely the effects at the participant level, we calculated memory clarity difference scores for items that never appeared in the Taboo game. We subtracted memory clarity ratings for Not relevant/not appeared items from relevant/not appeared items. Therefore, positive scores would indicate that participants rated their memory clarity higher for relevant items (active intentions) compared to not relevant items (cancelled intentions) that never appeared. Correspondingly, a negative score would indicate that they rated not relevant items as having higher memory clarity than relevant. We graphed the frequencies of these difference scores; inspection of these frequencies shows a larger number of positive values than negative (see Figure S1 in supplemental materials and results).

It is worth noting that across Experiments 1 and 2 (see Figures 3 and 4), memory clarity ratings for items that appeared in the Taboo game had fairly similar ratings; however, for items that *never* appeared in the Taboo game, participants in Experiment 2 rated those items substantially higher in memory clarity relative to Experiment 1. Many of the Experiment 2 participants were not native English language speakers; therefore, we suspected that participants may have been more reluctant to trust their impressions for the not appeared items and, in so doing, rated them higher compared to participants in Experiment 1. We conducted correlational analyses between the variables years of English and memory clarity ratings for items that never appeared in the Taboo task. Results showed that fewer years of speaking English correlated with increases in memory clarity ratings for word pairs that did not appear. See Figures S2 and S3 in supplemental materials and results for more detailed analysis.

General Discussion

We created a novel naturalistic prospective memory paradigm that is socially situated and that could provide a useful application for many basic research questions. In two experiments, we embedded a prospective memory task in the word game *Taboo*. The embedded task involved remembering to use certain clue words (those words deemed relevant in an earlier study episode) if a PM target appeared during the Taboo game. Twenty-four hours later, participants returned to complete a post-experiment questionnaire. Critically, participants rated their memory clarity of having said the PM clue word to PM targets that never appeared during the Taboo game. Results showed that those word pairs belonging to a relevant, to-be-executed intention were rated as having higher memory clarity even for items that never appeared during the Taboo game.

We predicted that the non-significant interaction observed in Experiment 1 between appearance and relevance would be significant in Experiment 2 as a result of the increased sample size and resulting statistical power. This prediction assumes that the effect size in Experiment 1 was an accurate reflection of the true effect size. The interaction in Experiment 2 was not significant, which may reflect the fact that the true effect is smaller than what we observed in Experiment 1. We believe that the interaction is not necessary to frame our results within the context of false PM. There was a significant difference between relevant/not relevant ratings for appeared items and also for not appeared items (hence the lack of an interaction). For the interaction to be significant, we would have needed a data pattern in which there was no difference between relevant/not relevant for appeared items but there was a difference for not appeared items. We argue that the interaction did not emerge because the "boost" from encoding items as relevant led to higher memory clarity ratings for *both* appeared items and for not appeared items. Encoding something as to-be-executed leads to a more vivid memory trace and enhanced accessibility (cf. the intention superiority effect); therefore, relevant items were rated higher regardless of whether they appeared or didn't appear. The fact that items that never appeared were rated as higher because they were to-be-executed is evidence of false prospective memory. The source monitoring framework and/or processing fluency may be helpful for understanding the mechanism underlying this effect.

Source Monitoring Framework

In the classic Deese, Roediger and McDermott (DRM) task, participants study lists of semantically-related words (e.g., nurse, hospital). After a delay, subjects try to recall or recognize these words. Typically, participants falsely recognize or recall the critical lure word that was never presented (e.g., "doctor") with high probability and confidence (Deese, 1959; Roediger & McDermott, 1995). Theoretical accounts of this false memory effect have included associative memory models stating that encoding highly associated words leads to spreading activation through an associative network to the non-presented associated lure (Roediger, McDermott, & Robinson, 1998). Other accounts attribute this effect to a source-monitoring error. The source monitoring framework can better account for the finding that participants often report that they remember presentation of the critical lure during encoding when in fact they are confusing thoughts of the critical lure with actual presentation of it (Johnson, Hashtroudi, & Lindsay, 1993). In the current study, higher memory clarity ratings for words that never appeared may occur because participants are misattributing the source of their memory. Given that relevant word pairs were encoded as to-be-executed, this special status may have led to a more vivid representation of these words. This may have led to later source misattributions such that the vividness was mistakenly attributed to being memory of an actual event.

Bugg, Scullin, and McDaniel (2013) stated that the strength of backward associations in the DRM paradigm accounts for the critical lure "doctor" coming to mind when highly associated words were studied. Bugg et al. showed that how one initially encodes a future intention influences the risk of commission errors (executing the PM intention when it was already executed). Specifically, participants encoded an implementation intention (Gollwitzer, 1999) augmented with instructions to imagine the event. Implementation intentions which have been shown to improve prospective memory (Cohen & Gollwitzer, 2008) increased the risk of committing PM commission errors. Similar to ideas expressed by Bugg et al. (2013), in the context of the current study, we suggest that encoding a future action involves making forward associations between the intended action and an anticipated or imagined future context. In this way, linking material (a Taboo target word and clue) with an intention to act in the future increases the likelihood that participants will falsely rate that material more clearly relative to an action that they did not have to execute.

Processing Fluency and Cue Discrepancy

As Rummel and Meiser (2016) state, due to their special meaning, PM cues are discrepant from their environment and therefore are sometimes spontaneously retrieved during the course of a prospective memory task. In their study, they presented participants with stimuli that were unrelated to the intention but discrepant from other stimuli. They manipulated processing fluency of stimuli for participants that had an active intention and found that stimuli whose fluency was increased led to a higher likelihood that participants would execute the intention despite its actual inappropriateness. That is, stimuli that felt discrepant or different from other ongoing stimuli were attributed to an active intention. Consistent with these ideas, word pairs in the current study that were to be executed had a special status relative to word pairs that belonged to a cancelled intention. When these stimuli were encountered 24 h later, they may have been processed more fluently and this discrepancy may have led to a higher memory clarity rating relative to word pairs belonging to a cancelled intention. Therefore, these misattributions of processing fluency (e.g., Jacoby, Kelley, & Dywan, 1989) may be interpreted as a sign of prior experience and therefore served as a cue for appearance in the Taboo game 24 h previously.

Our study had several limitations. First, our efforts to create a novel and ecologically-valid prospective memory paradigm prevented us from measuring some aspects of task performance such as reaction time. Second, we did not anticipate the considerable attrition of participants in Experiment 2. Finally, as mentioned previously, only one item from the not relevant/appeared category was included in the Day 2 questionnaire in an effort to reduce the chances that participants would realize the study's intent. Although it is highly unlikely, we cannot rule out the possibility that uneven representation of categories may have in some way influenced memory clarity ratings.

Conclusions

In two experiments, participants rated their memory clarity higher for items that they intended to execute relative to items that belonged to a cancelled intention. Although our data do not allow us to arbitrate among competing theoretical perspectives, the results are novel in that they are the first finding of young participants retrospectively rating their memory clarity of a prospective memory event that never occurred. As Whittlesea and Williams (2001) stated, people chronically evaluate their subjective processing experience. When the quality of processing is experienced as more fluent, then they engage in attributional processes to explain that experience of fluency. It may be that higher memory clarity ratings for relevant items 24 h later (for both appeared and not appeared items) reflected such attributional processes. In encoding a future action, we create a representation of action performance and it is this representation that we may later misattribute as memory of actual action performance. On Day 2, it may be that participants processed relevant items more fluently and attributed this fluency to a prior experience with this stimulus. Future studies would benefit by exploring the boundaries of this false prospective memory effect to better understand the specific underlying mechanisms that might account for this finding.

Author Contributions

A-LC and MJS conceived of the initial experiment during an undergraduate research methods class. These ideas were further developed in an undergraduate honors thesis by MJS supervised by A-LC. Authors DMB and DSL made substantial contributions to study conception and design. A-LC and DMB were the principle investigators for the initial experiment and the replication, respectively. MJS, DGD, and ZIH oversaw data collection. MJS, A-LC, DGD, and ZIH analyzed the data, and DMB and DSL verified analytic methods and assisted with results interpretation. The manuscript was written by several individuals: MJS wrote the Method and a portion of the Results for Experiment 1; DGD and ZIH wrote the Method and a portion of the Results for Experiment 2; and A-LC wrote the remaining sections with input from all other authors. All authors reviewed and revised the manuscript, and approved the final version. The authors would like to thank Natan Weissman, Jack Chen, Aaron Kong, and Joey Chhina for assistance with data collection and entry.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgment

We thank the Canada Research Chairs Program granted to Daniel Bernstein (#950-232078) for funding this work.

Online Supplement

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jarmac.2020.02.004.

References

Ball, B. H., Pitães, M., & Brewer, G. A. (2018). Individual differences in episodic memory abilities predict successful prospective memory output monitoring. *Memory*, *26*, 1159–1168. http://dx.doi.org/10.1080/09658211.2018.1436180

Bernstein, D. M., Scoboria, A., Desjarlais, L., & Soucie, K. (2018). "False memory" is a linguistic convenience. Psychology of Consciousness. *Theory, Research, and Practice*, *5*, 161–179. http://dx.doi.org/10.1037/cns0000148

Bugg, J. M., Scullin, M. K., & McDaniel, M. A. (2013). Strengthening encoding via implementation intention formation increases prospective memory commission errors. *Psychonomic Bulletin & Review*, *20*, 522–527. http://dx.doi.org/10.3758/s13423-013-0378-3

Cohen, A.-L., & Gollwitzer, P. M. (2008). The cost of remembering to remember: Cognitive load and implementation intentions influence on-going task performance. In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.), *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives* (pp. 367–390). Mahwah, NJ: Erlbaum.

Cohen, A. L., & Hicks, J. L. (2017). *Prospective memory: Remembering to remember, remembering to forget*. Cham, Switzerland: Springer.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22. http://dx.doi.org/10.1037/h0046671

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *16*, 717–726. http://dx.doi.org/10.1037//0278-7393.16.4.717

Einstein, G. O., McDaniel, M. A., Smith, R. E., & Shaw, P. (1998). Habitual prospective memory and aging: Remembering intentions and forgetting actions. *Psychological Science*, *9*, 284–288. http://dx.doi.org/10.1111/1467-9280.00056

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. http://dx.doi.org/10.3758/bf03193146

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848. http://dx.doi.org/10.3758/MC.38.7.833

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *The American Psychologist*, *54*, 493–503. http://dx.doi.org/10.1037//0003-066x.54.7.493

Goschke, T., & Kuhl, J. (1993). Representation of intentions: Persisting activation in memory. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *19*, 1211–1226. http://dx.doi.org/10.1037//0278-7393.19.5.1211

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*, 1–9. http://dx.doi.org/10.1002/1097-0193(200101)12:1<1::aid-hbm10>3.0.co;2-v

Harman, G. (1998). Intentionality. In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science* (pp. 602–610). Oxford: Blackwell.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of endel tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins, & R. M. Sorrentino (Eds.), *Handbook of motivation and social cognition: Foundations of social behavior* (pp. 482–526). New York: Guilford Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–29. http://dx.doi.org/10.1037//0033-2909.114.1.3

Kliegel, M., McDaniel, M. A., & Einstein, G. O. (Eds.). (2008). *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives*. Mahwah: Erlbaum.

Laney, C., & Loftus, E. F. (2013). Recent advances in false memory research. *South African Journal of Psychology*, *43*, 137–146. http://dx.doi.org/10.1177/0081246313484236

Marsh, R. L., Hicks, J. L., & Bink, M. L. (1998). Activation of completed, uncompleted, and partially completed intentions. *Journal of*

Experimental Psychology Learning, Memory, and Cognition, *24*, 350–361. http://dx.doi.org/10.1037//0278-7393.24.2.350

Marsh, R. L., Hicks, J. L., & Bryan, E. S. (1999). The activation of unrelated and canceled intentions. *Memory & Cognition*, *27*, 320–327. http://dx.doi.org/10.3758/bf03211415

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods Instruments & Computers*, *36*, 402–407. http://dx.doi.org/10.3758/bf03195588

Pink, J. E., & Dodson, C. S. (2013). Negative prospective memory: Remembering not to perform an action. *Psychonomic Bulletin & Review*, *20*, 184–190. http://dx.doi.org/10.3758/s13423-012-0337-4

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *21*, 803–814. http://dx.doi.org/10.1037/e537272012-273

Roediger, H. L., McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in producing false remembering. In M. A. Conway, S. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–245). Hove, Sussex: Psychological Press.

Rummel, J., & Meiser, T. (2016). Spontaneous prospective-memory processing: Unexpected fluency experiences trigger erroneous intention executions. *Memory & Cognition*, *44*, 89–103. http://dx.doi.org/10.3758/s13421-015-0546-y

Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. S. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*, 467–474. http://dx.doi.org/10.1016/j.tics.2011.08.004

Scullin, M. K., & Bugg, J. M. (2013). Failing to forget: Prospective memory commission errors can result from spontaneous retrieval and impaired executive control. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *39*, 965–971. http://dx.doi.org/10.1037/a0029198

Scullin, M. K., Bugg, J. M., McDaniel, M. A., & Einstein, G. O. (2011). Prospective memory and aging: Preserved spontaneous retrieval, but impaired deactivation, in older adults. *Memory & Cognition*, *39*, 1232–1240. https://doi.org/10.3758/s13421-011-0106-z

Scullin, M. K., Bugg, J. M., & McDaniel, M. A. (2012). Whoops, I did it again: Commission errors in prospective memory. *Psychology and Aging*, *27*, 46–53. http://dx.doi.org/10.1037/a0026112

Walser, M., Fischer, R., & Goschke, T. (2012). The failure of deactivating intentions: Aftereffects of completed intentions in the repeated prospective memory cue paradigm. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *38*, 1030–1044. http://dx.doi.org/10.1037/e512592013-649

Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *27*(1), 3–13. http://dx.doi.org/10.1037//0278-7393.27.1.3

Received 24 August 2019;
received in revised form 18 February 2020;
accepted 18 February 2020
Available online 29 April 2020