Cognition 204 (2020) 104391

Contents lists available at ScienceDirect

Cognition

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

Eye movement analyses of strong and weak memories and goal-driven forgetting

Jonathon Whitlock[a], Yi-Pei Lo[a], Yi-Chieh Chiu[a], Lili Sahakyan[a,b,*]

[a] Department of Psychology, University of Illinois at Urbana-Champaign, United States of America
[b] Beckman Institute for Advanced Science and Technology, United States of America

ARTICLE INFO

Keywords:
Intentional forgetting
Eye-movements
Suppression
Preferential viewing
Memory strength

ABSTRACT

Research indicates that eye movements can reveal expressions of memory for previously studied relationships. Specifically, eye movements are disproportionately drawn to test items that were originally studied with the test scene, compared to other equally familiar items in the test display – an effect known as preferential viewing (e.g., Hannula, Ryan, Tranel, & Cohen, 2007). Across four studies we assessed how strength-based differences in memory are reflected in preferential viewing. Participants studied objects superimposed on background scenes and were tested with three-object displays superimposed on the scenes viewed previously. Eye movements were monitored at test. In Experiment 1 we employed an item-method directed forgetting (DF) procedure to manipulate memory strength. In Experiment 2, viewing patterns were examined across differences in memory strength assessed through subjective confidence ratings. In Experiment 3, we used spaced repetitions to objectively strengthen items, and Experiment 4 involved a list-method DF manipulation. Across all experiments, eye movements consistently differentiated the effect of DF from other strength-based differences in memory, producing different viewing patterns. They also differentiated between incidental and successful intentional forgetting. Finally, despite a null effect in recognition accuracy in list-method DF, viewing patterns revealed both common as well as critical differences between list-method DF and item-method DF. We discuss the eye movement findings from the perspective of theoretical accounts of DF and other strength-based differences in memory.

1. Introduction

To most people, remembering is adaptive, whereas forgetting represents a failure of memory. Contrary to lay beliefs, forgetting serves an adaptive function (Bjork, 2011). It is not always necessary to retain everything, and sometimes we need to forget information that is outdated, inaccurate, or painful. The process of how we intentionally or voluntarily forget information has been examined with laboratory methods like the directed forgetting (DF) procedure (Bjork et al., 1968), or the think-no-think procedure (Anderson & Green, 2001). We limit the discussion to DF because that is the procedure we used in our studies. In DF studies, participants are exposed to some information to learn for a later memory test, and are subsequently told to forget some portion of that material. Numerous studies demonstrate impaired memory for F-items compared to R-items, referred to as the DF effect (for recent reviews, see Sahakyan, in press; Sahakyan & Foster, 2013; see also MacLeod, 1998), and the critical question is how it is accomplished. The theoretical interpretations of the DF effect continue to evolve to keep pace with new methodologies available to investigate it. It is not an exaggeration to say that a paradigm shift has taken place in the 21st century DF research, where the traditional inhibitory interpretation of list-method DF has been challenged by the context-change mechanism, whereas the passive view of item-method DF involving selective encoding has been challenged by active inhibitory interpretations aimed at downregulating representations of F-items. We will expand the theoretical discussion of different DF methods.

Typical DF studies have relied on observable behavioral measures such as percent correct or accuracy to assess impairment of F-cued items. However, this measure limits investigation to items that were retained in memory despite the instruction to forget (i.e., a failure of DF). In other words, much research has examined "anti-forgetting" rather than successful DF. The critical question is what happens to the F-items when DF is successful. Addressing this question is challenging and it requires employing a concurrent measure that accompanies

* Corresponding author at: Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, United States of America.

E-mail address: LSAHAKY@illinois.edu (L. Sahakyan).

https://doi.org/10.1016/j.cognition.2020.104391
Received 17 May 2019; Received in revised form 23 June 2020; Accepted 26 June 2020

behavior so that *successful intentional forgetting* (F-items that are subsequently forgotten) could be compared with *incidental forgetting* (R-items that are subsequently forgotten). To this end, electrophysiological and imaging studies have made progress in differentiating between incidental and successful intentional forgetting (for a review, see Anderson & Hanslmayr, 2014). Many behavioral studies have also yielded crucial insights about the nature of intentional forgetting by combining DF procedures with other measures, such as attentional tasks (Fawcett & Taylor, 2008, 2012; Taylor, 2005), or color-judgments tasks (Fawcett et al., 2016). Accumulating evidence suggests that intentional forgetting is more effortful and demanding than previously thought.

1.1. Eye-tracking investigation of directed forgetting

In this investigation, we used a novel approach by combining DF with eye movement analyses. Established research indicates that eye movements are an extremely sensitive marker of memory (e.g., Hannula et al., 2007; Ryan et al., 2000; Ryan & Cohen, 2004; Ryan & Villate, 2009). Furthermore, the influence of retained memory on viewing patterns is observed even when the behavioral selection is inaccurate (Hannula et al., 2012; Hannula & Ranganath, 2009; Nickel et al., 2015; Ryan et al., 2000; Ryan & Cohen, 2004). We saw this as an opportunity to study eye movements in conjunction with DF because doing so could provide insight into forgetting processes by revealing potential differences between incidental forgetting and successful intentional forgetting. We adapted a paradigm from the neurocognitive literature that has been well-integrated with eye movement monitoring as well as imaging techniques to allow relating memory performance to the underlying mechanisms. Before introducing the paradigm, we briefly summarize the theoretical mechanisms of DF.

1.2. Theoretical accounts of directed forgetting

In item-method DF (IMDF) studies, F- and R-instructions are delivered following each item, producing robust impairment of F-items. Traditionally, this impairment has been interpreted in terms of selective rehearsal of R-items, in which the F-cue leads to terminating of rehearsal of those items, whereas the R-cue leads to encoding those items more elaborately (Basden et al., 1993; Bjork, 1970; MacLeod, 1999). In contrast to this passive view, more active views of IMDF suggest that effortful processes are engaged to accomplish forgetting of F-items. For example, reaction times on the secondary tasks performed with IMDF are slower during the execution of the F-instruction, indicating that attention is actively withdrawn after the F-cue, and that forgetting is effortful (Fawcett & Taylor, 2008, 2012; Taylor, 2005; Wylie et al., 2008). In another study, participants engaged in a stop-signal task while studying words that were followed by R- or F-cues (Fawcett & Taylor, 2010). Participants were more successful in stopping performance after F-cues compared to R-cues, suggesting that F-cues might engage inhibitory mechanisms similar to stop signals. In addition to behavioral studies, multinomial modeling approaches that disentangle encoding and retrieval parameters of memory effects indicate that worse memory of F-items in IMDF is driven by impaired retrieval (in addition to impaired encoding of F-items), thereby suggesting that IMDF does not merely reflect a failure to encode information, but that it is also driven by retrieval factors, which presumably could be voluntarily controlled (Rummel et al., 2016; Marevic, Arnold, & Rummel, 2018; Marevic & Rummel, 2020). Finally, neural evidence indicates that successful IMDF recruits additional processes beyond those associated with incidental forgetting, and that these processes are likely inhibitory in nature (Anderson & Hanslmayr, 2014; Ludowig et al., 2010; Nowicka et al., 2011; Paz-Caballero et al., 2004; Rizio & Dennis, 2013; Oehrn et al., 2018; Van Hooff et al., 2009; Wylie et al., 2008). Our first study was inspired by this growing body of evidence, indicating that passive decay of F-items, which has historically dominated the field, is no longer the widely accepted view, and that inhibitory explanations are being entertained in the IMDF literature.

In list-method DF (LMDF) studies, F- and R-instructions are delivered once between the study of two lists, and they refer to an entire first list of items. Such procedures typically produce impaired memory for List 1 items, and enhanced memory for List 2 items in the F-group compared to the R-group. In contrast to IMDF, which leads to memory impairment in tests of recognition and recall, LMDF is typically not obtained in recognition (it is mostly found in free recall tests). To reconcile the discrepant findings in recognition between the IMDF and LMDF, the inhibitory explanation of LMDF was proposed, where the null effect in recognition was attributed to release of inhibition of F-items due to their re-presentation at test (e.g., Bjork, 1989; Bjork & Bjork, 2003; Geiselman et al., 1983). Much later, an alternative interpretation of LMDF was offered by the context-change account (Sahakyan & Kelly, 2002), which explained the null effect in recognition by pointing out the diminished role of contextual information in recognition tests, in contrast to the substantial effect of contextual information in free recall, where LMDF is robust. For a more detailed review of theoretical explanations across IMDF and LMDF, the reader is referred to Sahakyan (in press) or Sahakyan and Foster (2013). In this section we briefly reviewed various accounts, noting that our investigation was prompted by the debate in IMDF (although we also examined LMDF using eye movements).

1.3. Paradigm for assessing eye movements

We employed a cornerstone paradigm from the neurocognitive literature that assesses relational memory. The latter involves the ability to learn arbitrary associations between a set of elements that constitute a memory episode (Cohen & Eichenbaum, 1993; Eichenbaum & Cohen, 2001). Unlike single-item recognition tests, which involve deciding whether items had been studied, relational memory tests require nuanced distinctions involving how elements in an episode were related. The paradigm adapted in our investigation has been extensively used in assessment of relational memory impairment in amnesia (Hannula et al., 2007, 2015, 2006; Ryan et al., 2000; Ryan and Cohen, 2004), aging (Ryan et al., 2007), patients with schizophrenia (Williams et al., 2010; Hannula, Althoff, et al., 2010), as well as the benefits of aerobic fitness on relational memory (Baym et al., 2014; Monti et al., 2012). Importantly, eye-movement behavior has been extensively examined in this paradigm (e.g., for a review, see Hannula, Ranganath, et al., 2010), including with concurrent neuroimaging techniques (Hannula & Ranganath, 2009).

The details of the paradigm are demonstrated in Fig. 1. Briefly, participants are exposed to arbitrary item-scene pairings, one pair at a time, and are subsequently tested by being shown a previously exposed scene, followed by a 3-item display superimposed on that scene. One of the items was paired with that scene during encoding (target), whereas the remaining two items are familiar from the study phase, but were studied with different scenes (lures). Participant's task is to select the item that was paired with the presented scene at encoding. Because all presented items and scenes are familiar, correct identification of the target hinges on relational memory.

In a series of experiments with this task, Hannula et al. (2007) examined eye movements during the test phase. The viewing patterns demonstrated disproportionately greater viewing towards the target that was previously studied with the test scene relative to a selected lure, a phenomenon known as *preferential viewing* (Hannula et al., 2012, 2007; Hannula & Ranganath, 2009; Ryan et al., 2007; Baym et al., 2014). Critically, the difference in viewing behavior is between the two items which were selected behaviorally (i.e., selecting a Target represents a correct trial, and selecting a Lure represents an incorrect trial). Therefore, differences in viewing observed at the time of test are not driven by selection of an object, but rather reflect the influence of memory of the target having been studied with that particular background scene. Previous research demonstrates that preferential viewing

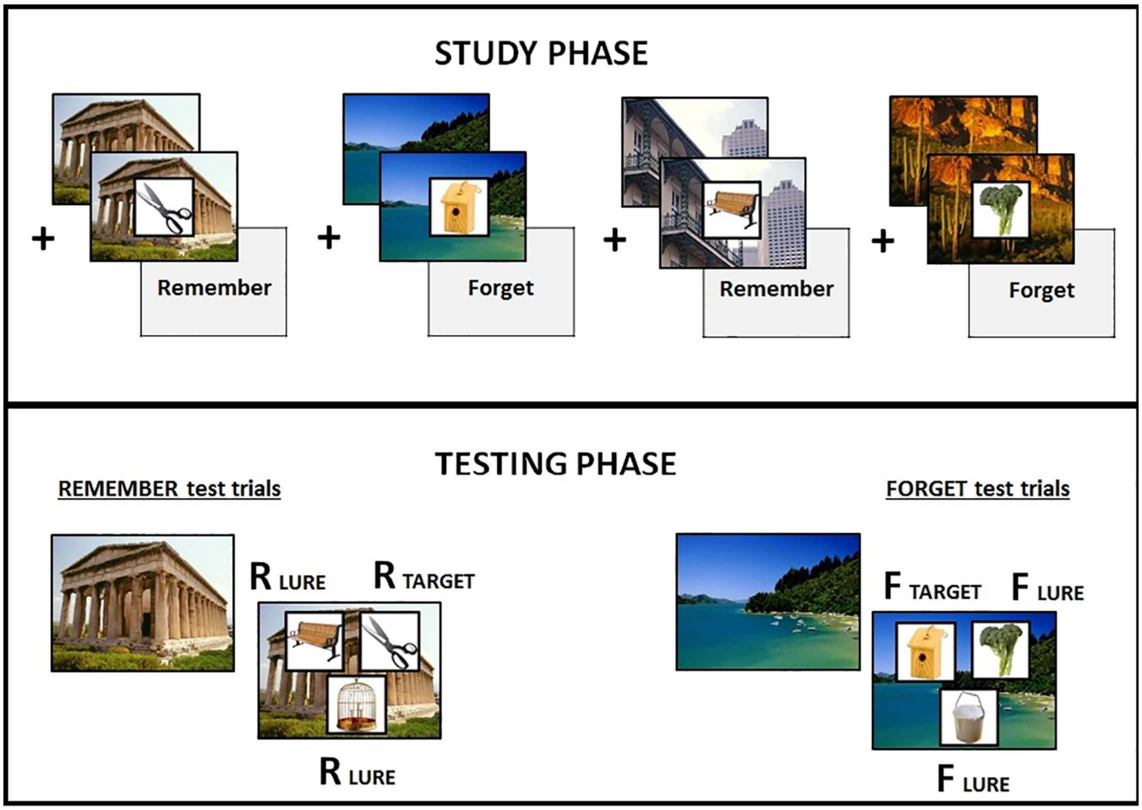

Fig. 1. Eye-movement relational memory paradigm. Participants are presented with a fixation cross, followed by an unobstructed picture of a scene, followed by a centrally superimposed object on that scene, and concluding with a memory instruction indicating whether the object needs to be remembered or forgotten (top row). At test, participants see a fixation cross, followed by unobstructed view of a previously studied scene, which in turn is followed by a three-object display consisting of three previously studied objects superimposed on that scene (bottom row). One of the objects had previously been paired with the scene (target), whereas the remaining two were paired with different scenes (lures). Participants are asked to indicate which object was paired with the presented scene. Note, that although the figure shows R and F instructions, which is what we implemented in our investigation, in the original paradigm, all study trials are "R" trials, without the actual presentation of the instruction.

emerges extremely rapidly, rising and peaking within approximately 500–750 ms after the onset of the 3-item test display. This time course is similar to ERP recordings showing brain activity that discriminated studied pairs from repaired items beginning approximately 600 ms following stimulus onset (Donaldson & Rugg, 1998, 1999). Also similar to ERP studies, viewing behavior can be shifted to align with the overt behavioral response (i.e., response-locked analysis), where preferential viewing was shown to peak approximately 1000–500 ms prior to item selection. The time course of eye-movement behavior is quite robust across many studies. It is uninfluenced by task demands and emerged even when viewing the target was counterproductive to the task (Ryan et al., 2007).

1.4. Current studies and predictions

Relational memory is impaired by IMDF, as demonstrated in previous behavioral studies (Bancroft et al., 2013; Hockley et al., 2016; Hourihan et al., 2007; Marevic et al., 2018). Our investigation aimed to assess how impaired relational memory is expressed in eye movements. More specifically, the first goal was to assess how IMDF affects the magnitude of preferential viewing. As discussed above, preferential viewing is defined as disproportionate amount of viewing directed towards the Selected Target than towards the Selected Lure, reflecting the expression of memory in eye movement behavior. We had reasons to suspect that preferential viewing would be reduced in IMDF in part because various populations with impaired relational memory were examined using this paradigm, and showed reduced preferential viewing. Thus, we suspected that viewing to the Selected Target may be reduced in the F-condition.

Our second goal was to assess viewing patterns on *incorrect trials*, when participants select the Lure. The critical question is what happens to viewing to the Unselected Target on those trials, and whether it differs between the cue conditions. Evidence of retained memory traces for target items (despite selecting the lure) would be reflected in greater viewing directed towards the Unselected Target compared to the Unselected Lure. Given that on incorrect trials participants select a lure, there is no reason to expect viewing differences between the remaining two unselected objects (i.e., target and the second lure), unless there was lingering memory for the target item, leading to greater viewing of that Unselected Target compared to the Unselected Lure. We aimed to examine whether viewing patterns to the Unselected Target distinguish intentional forgetting from incidental forgetting.

The debate in IMDF literature centers around whether impaired memory is driven by passive withdrawal of rehearsal in response to forget cues, or due to active inhibitory processes that downregulate memory for F-items. By examining viewing patterns on incorrect trials, we tested two different predictions of these accounts. The inhibitory account predicts diminished viewing to the Unselected Target in the Forget compared to the Remember condition because according to this account active inhibitory processes are engaged to downregulate the processing of F-items. Therefore, successfully forgotten F-items result from the consequences of inhibitory processes impairing memory, as opposed to forgotten R-items, which arise from selective encoding differences. In contrast, the selective rehearsal account attributes DF to selective encoding, such that withdrawing rehearsal from R-items turns them functionally into F-items, and therefore forgotten R-items should be similar to F-items in the viewing patterns on incorrect trials.

These predictions were examined in Experiment 1 using an IMDF paradigm. In Experiments 2 and 3, we addressed alternative interpretations of IMDF findings, by eliminating the DF manipulation and examining eye movements in strong and weak memories. Finally, Experiment 4 involved LMDF to assess if eye movements might reveal processes that go undetected in behavioral accuracy, and whether they differentiate DF in general from other strength-based differences in memory.

2. Experiment 1: item-method DF

2.1. Methods

2.1.1. Participants & sample size

Participants were 31 undergraduates from the University of Illinois who were compensated with course credit. Eye-movement recordings were not recorded for one participant due to equipment failure, and eye-movement analyses were conducted on 30 participants, whereas recognition accuracy was based on 31 participants. The study was approved by the Institutional Review Board of University of Illinois at Urbana-Champaign and complied with APA ethical standards in the treatment of participants. All participants gave informed consent prior to inclusion in the study.

We estimated sample size based on a priori power analysis for repeated measures ANOVA, in which we were planning to test an interaction between Cue (F vs. R) and viewing to Selected Objects (Target vs. Lure). We used G*Power package (Faul et al., 2007). The power analysis for a small effect size interaction showed that we needed 30 participants at power equal to 0.80, with $\alpha = 0.05$ (two-tailed). However, as part of the initial revision of the manuscript, we switched from ANOVAs to employing multilevel modeling. Our sample size based on our original unilevel power analysis should be sufficient given that multilevel models are more powerful than ANOVAs for this type of data (Brysbart & Stevens, 2018).

2.1.2. Apparatus

Eye position was recorded at a rate of 1000 Hz using an Eyelink 1000 eye-tracking system (SR Research). After the study block and prior to the test block, eye position was calibrated using a 3 × 3 spatial array. Calibration ended with participants fixating on a centrally located cross-hair, which began the test block. The computer screen resolution was set to 1280 × 1024.

2.1.3. Stimuli

The stimuli consisted of 108 colored images of real-world, identifiable objects selected from various online sources, including Google Images (sized to 300 × 300 pixels), and 108 colored images of real-world scenes (sized 800 × 600 pixels) selected from the Fine-Grained Image Memorability (FIGRIM) dataset (Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015). Object stimuli were images of everyday objects, such as office supplies, food items, kitchen utensils, toys, musical instruments, etc. Scenes consisted of both indoor and outdoor scenes comprised of urban, rural, and natural landscapes, whereas indoor scenes were equally comprised of rooms in houses and buildings. Object-scene pairings were randomly determined across participants, and were equally assigned Remember and Forget instructions across counterbalancing conditions. In the test display, objects were presented in the top left, top right, and bottom middle positions (see Fig. 1), and across counterbalancing conditions, target objects appeared in each spatial position an equal number of times.

2.1.4. Procedure

The details of the paradigm are shown in Fig. 1. There was an initial practice block consisting of 6 study trials and 2 test trials. The practice block was identical to the actual study and testing procedure, with the exception that Forget/Remember memory instructions were not introduced during practice. The purpose of the practice block was to familiarize participants with encoding and testing procedures, and learn the mapping of responses on the keyboard with the spatial arrangement of test items in the object display. The encoding phase consisted of 108 study trials, in which participants were presented a fixation point for 1 s, followed by an unobstructed scene for 2 s. Afterwards, an object was superimposed on the scene for 4 s, and participants were instructed to think of how well the object and scene went together. This is a typical instruction in this paradigm to ensure encoding of relational information, to permit observing its expression in eye movements.[1] Afterwards, a Forget or Remember memory instruction was presented for 2 s. Half of the study trials were followed by a Remember, the other half by a Forget, instruction, and were randomly assigned with the exception that no more than three memory instructions of the same type occurred in a row. Participants were told that objects followed by a Remember instruction would appear on the memory test, and they should keep them in mind, while objects followed by a Forget instruction would not appear on the memory test, and that participants should do what they can to forget them.

All eye tracking data was collected during the testing phase. At test, participants were told they would be tested on all objects from the encoding phase, regardless of the memory instruction. That is, participants were to indicate the object they remember being paired with the scene even if they were previously instructed to forget that object. Testing phase consisted of 36 trials, half including all F-items and the other half including all R-items. Each test trial began with a fixation point for 1 s, followed by an unobstructed scene preview for 2 s. Afterwards, a 3-object display was superimposed on the scene for 6 s. One of the objects had been previously paired with the scene, while the other two had been previously paired with other scenes. Participants' task was to indicate which of the three objects had been previously paired with the scene. Participants made their selection via key presses on the computer keyboard, while their eyes were fixated on the computer screen. Importantly, within the 3-object display, all objects came from the same memory instruction, thereby controlling for item strength within the test trial (see Fig. 1, bottom panel). Within each testing phase and throughout counterbalancing conditions, the Target objects appeared equally often in any of the three test locations on the screen. Each test trial concluded after participants were given 4 s to provide a binary confidence judgment (Low or High), indicating their confidence that the object they selected had been previously paired with the test scene.

2.2. Results

2.2.1. Analytic approach

The study procedure and data collection was conducted using *E*-Prime 2.0 (Psychology Software Tools, Inc, 2015). Eye tracking data was extracted for analysis using MATLAB software (MATLAB, 2018). All statistical analyses were done using R software (R Development Core Team, 2008). The eye movement analyses (i.e., proportion of viewing time on individual trials) were performed using mixed effects models, fitted with the lmer function in the lme4 package (Bates et al., 2015) as well as the lmerTest package (Kuznetsova et al., 2017). In evaluating proportion of viewing, Cue (F vs. R), Selected Object (Target vs. Lure) and Bin (four 500 ms time bins amounting to 2 s prior to response selection) were used as the fixed effects, and Participants as random intercepts. The Cue and Selected Object variables were mean centered and contrast coded. Bin was coded using a linear mean centered contrast coding, as an increasing linear trend was expected in the

[1] Studies in our lab indicate that the practice session familiarizes participants with the task demands so that they engage in associative encoding on their own, even if encoding instructions are not explicitly delivered.

Bin variable based on previous research with this paradigm (e.g., Hannula et al., 2007; Hannula & Ranganath, 2009; Baym et al., 2014. Further details about analyses are explained below). Models were fit by maximum likelihood using the lme4 package in R (Bates et al., 2015), and Wald's z-scores were computed for each coefficient to test for significance of fixed effects (whenever appropriate). Significance testing for each coefficient was done using the lmerTest package in R (Kuznetsova et al., 2017).

Our analyses centered around two broad goals of this investigation. The first goal was to assess how DF instructions affect the magnitude of preferential viewing. Prior research indicates that preferential viewing unfolds over time, rising and peaking *prior* to the overt behavioral response, after which it diminishes and levels off. Therefore, first we identified the time point at which preferential viewing is fully established (i.e., when viewing peaks and is at its maximum), using response-locked analyses. Afterwards, we performed targeted hypothesis testing examining the effect of Cue on preferential viewing in that critical time bin. The second goal of this investigation was to assess viewing patterns on incorrect trials, when participants select the lure. The critical question is what happens to viewing to the Unselected Target on those trials, and whether it differs between the cue conditions.

2.2.2. Recognition accuracy

Recognition accuracy was analyzed with a multilevel logistic regression model, which predicted whether participants made a correct or incorrect recognition response on a trial-by-trial basis. In this analysis, *Cue* (R vs. F) was used as a fixed effect and *Participants* was treated as a random intercept for the fixed effect of cue. Recognition accuracy replicated the DF effect – participants were more likely to select the target in the Remember condition ($M = 0.81$, $SD = 0.39$) than the Forget condition ($M = 0.73$, $SD = 0.44$), $\beta = 0.48$, $SE = 0.15$, $z = 3.17$, $p = .002$.

2.2.3. Eye movement analyses

2.2.3.1. Preferential viewing analyses. For eye movement analyses, we followed the established practices in the literature and calculated a measure of viewing time to an element of a test display as a proportion of total time viewing all three elements (“proportion of viewing time”). In order to measure proportion of viewing time, three distinct regions of interest (ROI) were defined, where each ROI indicated the area on the screen where one of the objects was presented during the test. On trials where an accurate response was made (the target was selected), viewing time was calculated as the proportion of time viewing that target out of total time viewing all three objects. On trials when an inaccurate response was made (a lure was selected), viewing time was calculated as the proportion of time viewing that lure out of total time viewing all three objects. Thus, the number of correct/incorrect trials that contribute to estimates of viewing time was not expected to be equivalent across conditions, and therefore multilevel models were used to evaluate the results.

Fig. 2 summarizes mean proportion of viewing (averaged across participants and across trials) to a Selected Object (Target vs. Lure), by Cue (F vs. R), and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response on each trial (i.e., response-locked figure). To assess preferential viewing, a mixed effects model was fit to the proportion of viewing on each trial, using *Cue* (F vs. R), *Selected Object* (Target vs. Lure), and *Bin* (four 500 ms time bins, equating to 2 s prior to response) as fixed effects, and *Participants* as a random intercept. There was a significant effect of Selected Object, $\beta_{\text{object}} = 0.13$, $SE = 0.02$, $t = 8.09$, $p < .001$, indicating disproportionately more viewing to the Selected Target than to the Selected Lure, confirming preferential viewing. There was also an effect of Bins, $\beta_{\text{bin}} = 0.18$, $SE = 0.02$, $t = 11.49$, $p < .001$, confirming that overall viewing increased over time. Critically, these effects were qualified by a significant Selected Object × Bin interaction, $\beta = 0.22$, $SE = 0.03$, $t = 7.02$, $p < .001$, indicating that the magnitude of preferential viewing varied across the time bins (the variance associated with random intercept of participants was $\sigma^2 = 0.003$, $SD = 0.05$). Preferential viewing peaked in the 500 ms bin prior to response, with approximately 72% of the total viewing time being devoted to the Selected Target over the Selected Lure, $\beta_{\text{object}} = 0.23$, $SE = 0.03$, $t = 7.55$, $p < .001$.

These analyses confirmed the established effects in the literature that preferential viewing peaks prior to the behavioral selection, establishing itself by 500 ms prior to response. In order to assess how DF affected preferential viewing, we examined the effect of Cue on Selected Objects in the 500 ms time bin prior to response with a Mixed Effects Model. In addition to the significant effect of Selected Object, $\beta_{\text{object}} = 0.24$, $SE = 0.03$, $t = 7.82$, $p < .001$, there was a significant Selected Object × Cue interaction, $\beta = 0.12$, $SE = 0.06$, $t = 2.08$, $p = .038$. Although preferential viewing was observed in both cue conditions as evidenced by the significant effect of selected object in the Forget ($\beta_{\text{object}} = 0.18$, $SE = 0.04$, $t = 4.57$, $p < .001$) and in the Remember conditions ($\beta_{\text{object}} = 0.30$, $SE = 0.05$, $t = 6.68$, $p < .001$), preferential viewing was reduced in the Forget condition. This was evidenced by the different effects of cue within selected objects. Namely, more viewing was directed towards the Selected Lure in the Forget compared to the Remember condition ($\beta_{cue} = -0.12$, $SE = 0.06$, $t = 2.19$, $p = .030$), whereas viewing to the Selected Target was the same between the cue conditions ($\beta_{cue} < 0.001$, $SE = 0.03$, $t = 0.03$, $p = .977$). Thus, preferential viewing was reduced in the Forget condition because more viewing time was devoted to the Selected Lure. Taken together, the viewing pattern in the Remember condition replicated the established findings in the literature. The novel findings in the Forget condition indicated that preferential viewing was reduced by the DF manipulation, and that it was driven by enhanced viewing on incorrect trials, rather than reduced viewing on correct trials.

2.2.3.2. Analyses of retained traces. Previous literature has identified eye movement behavior as a marker of retained memory traces in the absence of conscious recollection. Even when behavioral accuracy failed, eye movements were shown to indicate the influence of memory on viewing patterns, suggesting that they are a more sensitive (albeit indirect) marker of memory compared to explicit accuracy. Our interest was whether viewing towards the Unselected Target would distinguish between incidental forgetting and intentional forgetting, which would indicate differential retention of memory traces for Targets across Forget and Remember conditions.

Fig. 3 summarizes average proportion of viewing on incorrect trials (averaged across participants) to Unselected Objects (Target vs. Lure), by Cue (F vs. R), and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response (i.e., response-locked figure). Although Bin was not a variable of interest in this analyses, we presented the results across the time bins for completeness. A mixed effects model was fit to proportion of viewing time on incorrect trials (collapsed across bins) using *Cue* and *Unselected Object* as fixed effects, and *Participants* as a random intercept. There was an effect of Unselected Object, $\beta_{\text{object}} = 0.10$, $SE = 0.02$, $t = 4.74$, $p < .001$, indicating greater viewing to the Unselected Target than the Unselected Lure (the variance associated with random intercept of participants was $\sigma^2 < 0.001$, $SD < 0.001$). Although the Cue × Unselected Object interaction did not reach statistical significance ($\beta = 0.07$, $SE = 0.04$, $t = 1.71$, $p = .088$), there was a significant effect of Cue in viewing to the Unselected Targets, $\beta_{\text{cue}} = 0.07$, $SE = 0.03$, $t = 2.43$, $p = .02$, but no effect of Cue in viewing to the Unselected Lures, $\beta_{\text{cue}} < 0.01$, $SE = 0.003$, $t = 0.17$, $p = .865$. These findings indicate that on incorrect trials, participants tend to view the Unselected Target substantially less in the Forget condition than in the Remember condition, indicating reduced retention of memory traces of target objects as a result of successful DF.

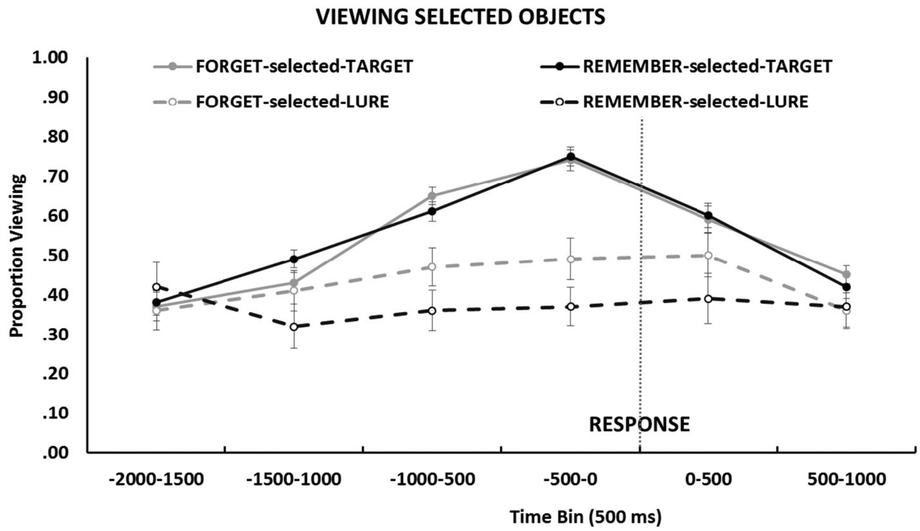

Fig. 2. Mean proportion of viewing to the Selected Target and Selected Lure across Cue and Time Bins (grouped by 500 ms time.) in Experiment 1. Error bars represent SE of the mean.

2.3. Discussion

In this experiment, we obtained a DF effect in recognition accuracy, indicating impaired relational memory as a result of the Forget cue. Eye movements revealed reduced preferential viewing in the Forget compared to the Remember condition. This was driven by enhanced viewing towards the Selected Lure in the Forget compared to the Remember condition, whereas viewing to the Selected Target remained invariant across the cue conditions. This pattern indicates that when participants failed to forget the target (despite being told to forget), their eye movements were indistinguishable from the Remember condition. On the other hand, when participants forgot the target (consistent with the instruction to forget), they viewed those missed targets substantially less in the Forget condition compared to the missed targets in the Remember condition. The assessment of viewing on incorrect trials indicated that there was reduced evidence of retained memory traces for the Unselected Targets in the Forget condition compared to the Remember condition. Thus, eye-movements provided a new marker of intentional forgetting, distinguishing *intentional* from *incidental forgetting*.

The eye movement findings in the Forget condition provide support for the inhibitory account of IMDF. However, an alternative explanation might be that the observed viewing differences between the Forget and Remember conditions are not specific to inhibition, per se, but rather reflect more general viewing differences that emerge between any strong and weak memories, with R-items reflecting strong memories, and F-items reflecting weak memories (regardless of what produces those strength-based differences). Interestingly, published data to date do not provide guidance to the question of how preferential viewing differs between strong and weak conditions of the experiment. This is because past research with this paradigm has been geared towards testing various populations with compromised hippocampal activity (e.g., older adults, schizophrenia patients, or amnesic patients), or examining the neural mechanisms of preferential viewing. Given that a third variable problem cannot be ruled out in studies with population differences, we aimed to directly examine preferential viewing in strong and weak memories. If what was observed in the current study is reflective of more general viewing differences produced by strong and weak memories, then we should observe similar viewing behaviors without DF manipulations.

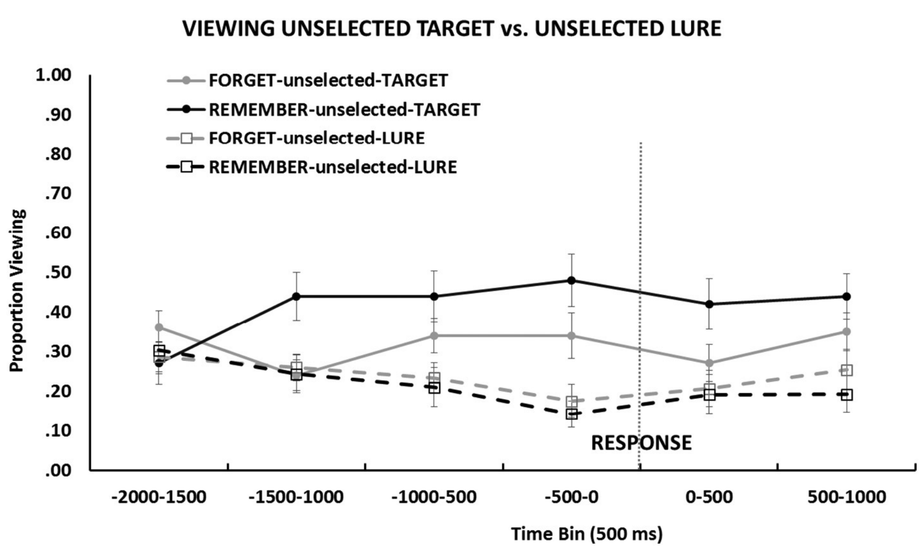

Fig. 3. Mean proportion of viewing time on incorrect trials directed towards the Unselected Target and Unselected Lure as a function of Cue and Time Bins in Experiment 1. Error bars represent SE of the mean.

Initially, we re-examined Experiment 1 findings as a function of participant-provided confidence ratings obtained in the experiment, and found robust differences in memory accuracy between high and low confidence conditions, consistent with the well-known relationship between confidence and accuracy (e.g., Tekin et al., 2018; Wixted & Wells, 2017). Unfortunately, confidence judgments were timed, resulting in participants failing to provide many confidence ratings (despite not missing the actual object selection). This precluded meaningful analyses of the eye movements as a function of confidence, despite some promising preliminary trends. We revisited this issue in Experiment 2.

3. Experiment 2: subjective memory strength

In this study, we assessed viewing behavior as a function of participant-provided confidence ratings in the absence of any DF manipulations. The procedure was similar to Experiment 1, with the exception that the DF manipulation was eliminated, and the entire procedure was divided into several mini study-test blocks, which is a common practice in preferential viewing literature. After encoding a block of object-scene pairs, participants received a test on those items, before moving on to the next study block. There were a total of four study-test blocks. Importantly, in the current experiment, confidence judgments were untimed, and for more nuanced discrimination, participants provided Low/Medium/High confidence ratings, as opposed to a binary confidence rating used previously. Thus, confidence was used as a proxy for memory strength. We aimed to assess whether preferential viewing varied across memory strength conditions, and how these differences compared to what was observed in Experiment 1.

3.1. Methods

3.1.1. Participants & sample size

Participants were 30 students from the University of Illinois who were compensated with course credit. The study was approved by the Institutional Review Board of University of Illinois at Urbana-Champaign and complied with APA ethical standards in the treatment of participants. All participants gave informed consent prior to inclusion in the study. Sample size determination was similar to the previous experiment.

3.1.2. Apparatus

Eye position was recorded at a rate of 1000 Hz using an Eyelink 1000 Plus eye-tracking system (SR Research). The calibration stage and the computer screen resolution were identical to Experiment 1. Calibration took place prior to the beginning of each test block, for a total of four times. Each calibration ended with participants fixating on a centrally located cross-hair, which began the test block.

3.1.3. Stimuli

The stimuli included all materials from Experiment 1, and were supplemented by an additional set of 36 objects and 36 scenes selected from the same sources as in the previous experiment, resulting in a total of 144 background scenes and 144 objects.

3.1.4. Procedure

The details of the paradigm and the encoding instructions are similar to Experiment 1 with a few exceptions. First, there were no DF instructions (all trials were "Remember" trials). Second, the entire experiment was divided into four study-test blocks, each block consisting of encoding 36 object-scene pairs, and receiving 12 trials of a subsequent test. Each test trial contained 3-objects per test display, with a Target and two Lures (similar to Experiment 1). In total, participants were exposed to 144 object-test pairings, and received a total of 48 test trials. The third change to the procedure involved collecting three confidence judgments (Low, Medium, or High), and the ratings were untimed.

3.2. Results

3.2.1. Recognition accuracy

Recognition accuracy was analyzed with a multilevel logistic regression model using *Confidence* (Low vs. Medium vs. High) as a fixed effect and *Participants* as a random intercept for the fixed effect of confidence. Overall recognition accuracy was 82%, and it varied across the participant-provided confidence ratings (Low Vs. *medium* vs. High), $\beta = 2.64$, SE = 0.22, $z = 11.85$, $p < .001$. Specifically, accuracy was higher in the High confidence condition ($M = 0.93$, $SD = 0.25$) than in a Medium confidence condition ($M = 0.65$, $SD = 0.48$), $\beta = 3.96$, SE = 0.41, $z = 9.69$, $p < .001$, which in turn was higher than in the Low confidence condition ($M = 0.53$, $SD = 0.50$), $\beta = 1.17$, SE = 0.45, $z = 2.58$, $p = .01$. Thus, higher confidence ratings were associated with more accurate memories. We therefore used confidence as a proxy for subjective memory strength and examined viewing behavior as a function of confidence.

3.2.2. Eye movement analyses

3.2.2.1. Preferential viewing analyses. Fig. 4 summarizes mean

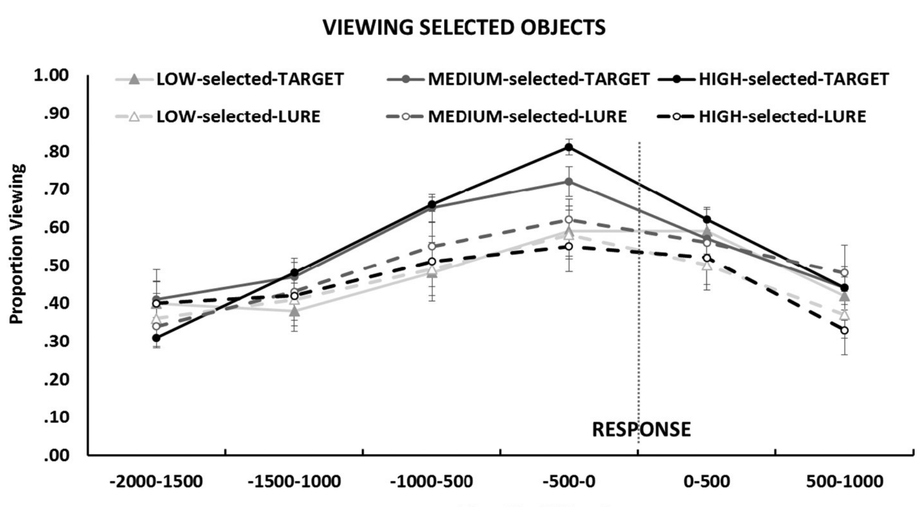

Fig. 4. Mean proportion of viewing time to the Selected Target and Selected Lure as a function of Confidence and Time Bins (grouped by 500 ms) in Experiment 2. Error bars represent SE of the mean.

proportion of viewing (averaged across participants and across trials) to a Selected Object (Target vs. Lure), by Confidence (Low vs. Medium vs. High) and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response on each trial. To assess preferential viewing, a mixed effects model was fit to the proportion of viewing time on each trial, using *Confidence* (Low vs. Medium vs. High), *Selected Object* (Target vs. Lure), and *Bin* (four 500 ms time bins, equating to 2 s prior to response) as fixed effects, and *Participants* as a random intercept. There was a significant effect of Selected Object, $\beta_{\text{object}} = 0.13$, $SE = 0.02$, $t = 7.18$, $p < .001$, indicating disproportionately more viewing to the Selected Target than to the Selected Lure, confirming preferential viewing. There was also an effect of Bin, $\beta_{\text{bin}} = 0.17$, $SE = 0.02$, $t = 9.86$, $p < .001$, indicating that overall viewing increased over time. These effects were qualified by a significant Selected Object × Bin interaction, $\beta = 0.18$, $SE = 0.03$, $t = 5.29$, $p < .001$, indicating that the magnitude of preferential viewing varied across the time bins (the variance associated with random intercept of participants was $\sigma^2 = 0.004$, $SD = 0.06$). Viewing reached its peak in the 500 ms time bin prior to response, with approximately 79% of the total viewing time being devoted to the Target over the Lure, $\beta_{\text{object}} = 0.17$, $SE = 0.03$, $t = 6.26$, $p < .001$. These results confirm the established findings in the literature as well as the findings of the first experiment.

To assess whether viewing patterns varied across confidence levels, we examined proportion of viewing in the 500 ms time bin prior to response with a mixed effects model, using *Confidence* (Low vs. Medium vs. High) and *Selected Object* (Target vs. Lure) as fixed effects, and *Participants* as a random intercept. In addition to the significant effect of Selected Object, $\beta_{\text{object}} = 0.11$, $SE = 0.03$, $t = 3.83$, $p < .001$, there was a significant Selected Object × Confidence interaction, $\beta = 0.16$, $SE = 0.07$, $t = 2.39$, $p = .017$. This interaction arises because of a significant effect of Confidence on viewing to the Selected Target, $\beta_{\text{confidence}} = 0.17$, $SE = 0.03$, $t = 5.03$, $p < .001$, indicating greater viewing to the Selected Target in the High confidence condition compared to the Medium confidence or Low confidence conditions. However, there was no effect of Confidence on viewing to the Selected Lure, $\beta_{\text{confidence}} = -0.01$, $SE = 0.08$, $t = 0.07$, $p = .943$. The net result was that the magnitude of preferential viewing decreased across decreasing levels of confidence – whereas preferential viewing was robust in the High confidence condition, $\beta_{\text{object}} = 0.22$, $SE = 0.04$, $t = 5.15$, $p < .001$, it was attenuated although still significant in the Medium confidence condition, $\beta_{\text{object}} = 0.11$, $SE = 0.05$, $t = 2.17$, $p = .031$, whereas in the Low confidence condition, there was no preferential viewing, $\beta_{\text{object}} = 0.01$, $SE = 0.07$, $t = 0.15$, $p = .879$.

3.2.2.2. Analyses of retained traces. Fig. 5 summarizes mean proportion of viewing on incorrect trials (averaged across participants and across the trials) to Unselected Objects (Target vs. Lure), by Confidence (Low vs. Medium vs. High) and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response. The confidence provided for the selected object on incorrect trials was used to assign the memory strength of the unselected targets and lures (on incorrect trials, participants select one of two lures and assign it a confidence judgment; the unselected target and unselected lure on that trial is being assigned the proxy strength of the selected lure).

The same mixed effects model was used to assess the proportion of viewing time on incorrect trials as in Experiment 1. There was only an effect of Unselected Object, $\beta_{\text{object}} = 0.04$, $SE = 0.02$, $t = 1.93$, $p = .054$, indicating greater viewing to the Unselected Target than to the Unselected Lure (the variance associated with random intercept of participants was $\sigma^2 = 0.001$, $SD = 0.04$). There was no effect of Confidence in viewing to Unselected Target, $\beta_{\text{confidence}} = 0.05$, $SE = 0.05$, $t = 1.24$, $p = .217$, although there was an effect of Confidence in viewing to the Unselected Lure, $\beta_{\text{confidence}} = -0.07$, $SE = 0.04$, $t = 2.10$, $p = .036$. The Confidence × Unselected Object interaction was significant, $\beta = 0.11$, $SE = 0.05$, $t = 2.11$, $p = .035$. This interaction reflected an effect of Unselected Object in the High confidence condition, $\beta_{object} = 0.13$, $SE = 0.04$, $t = 3.23$, $p = .001$, but not the Medium confidence condition, $\beta_{object} = -0.01$, $SE = 0.03$, $t = 0.20$, $p = .845$, or the Low confidence condition, $\beta_{object} = 0.01$, $SE = 0.04$, $t = 0.32$, $p = .749$. Thus, the differential retention of memory traces across levels of confidence was due to reduced viewing to the Unselected Lures in the High confidence condition compared to the lower confidence conditions. This is the opposite of what was observed in Experiment 1, where there was reduced viewing to the Unselected Target.

3.3. Discussion

The current experiment revealed recognition accuracy differences as a function of confidence, confirming that memory strength varied across confidence levels. Eye movements revealed that preferential viewing was reduced across confidence conditions. Participants viewed the Selected Target more in the High confidence condition, and they viewed it decreasingly less in the Medium and Low confidence conditions. At the same time, there were no differences in viewing to the Selected Lure across confidence conditions. This pattern of viewing behavior is the opposite of what was observed in Experiment 1, where reduced preferential viewing in the weak/Forget condition was driven by increased viewing towards the Selected Lure, rather than decreased viewing of the Selected Target.

In terms of the viewing patterns on incorrect trials (i.e., retained traces analyses), viewing to the Unselected Target was invariant across the confidence levels. This finding also contrasts with Experiment 1, where viewing to the Unselected Target was reduced in the weak/Forget condition. We interpret this finding with caution because in Experiment 2, the confidence judgments that were provided for the selected lures on incorrect trials were used to assign proxy strength for unselected target and unselected (second) lure on that trial. In contrast, in Experiment 1, the unselected targets and unselected lures had objectively defined strength. Nevertheless, the overall pattern of findings across preferential viewing analyses and the retained traces analyses suggests that strength-based differences that arise when the participants' goal is to remember all items (as in Experiment 2) cannot explain the viewing patterns of Experiment 1, where the goal of forgetting some items in IMDF produced a different set of viewing behaviors, despite strength based differences between F and R items.

Importantly, the confidence judgments used to group the unselected objects in Experiment 2 were taken from those assigned to the incorrectly selected objects on those trials, and therefore the conclusions should be taken with caution. In Experiment 3 we sought to overcome this limitation by objectively manipulating the memory strength of all objects.

4. Experiment 3: objective repetitions

The results of the previous two studies tentatively suggest that the eye movement patterns observed in Experiment 1 and Experiment 2 may reflect the influence of two fundamentally different mechanisms on viewing behavior — namely, that the viewing patterns observed in Experiment 1 may be specific to DF, and may not generalize to other strength-based differences in memory, despite both experiments producing differences in memory accuracy. However, memory strength was not objectively manipulated in Experiment 2, where confidence served as a proxy for memory strength. Therefore, in Experiment 3 we aimed to further examine strength-based differences in memory and their reflection in the viewing behavior by objectively varying strength through repetitions, given that they have a robust influence on strengthening memory (e.g., Hintzman, 1974). In this study, there was no DF manipulation, and half of the object-scene pairs were shown twice, whereas the other half of object-scene pairs were presented once. The objective strengthening manipulation would also permit more

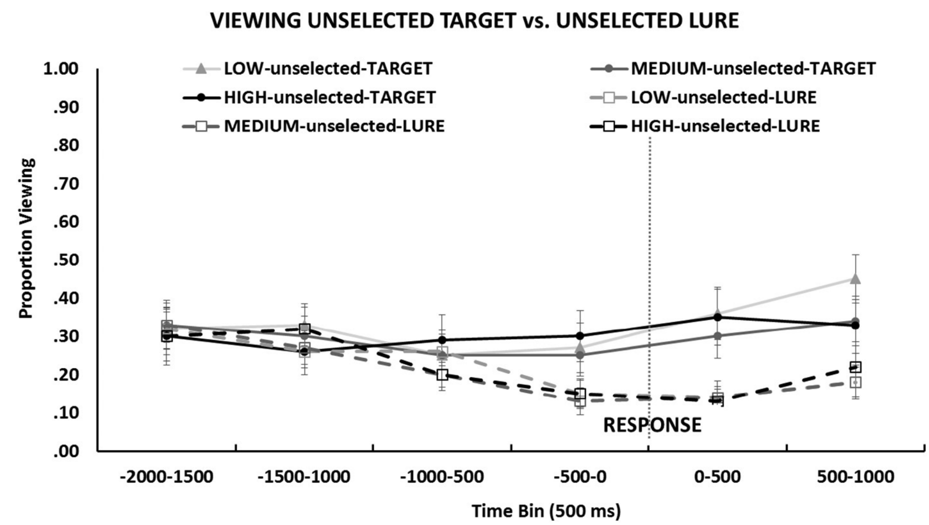

Fig. 5. Mean proportion of viewing time on incorrect trials directed towards the Unselected Target and Unselected Lure as a function of Cue and Time Bins in Experiment 2. Error bars represent SE of the mean.

unequivocally interpreting the findings from the retained traces analyses in Experiment 2.

4.1. Methods

4.1.1. Participants & sample size

Participants were 30 students from the University of Illinois who were compensated with course credit. The study was approved by the Institutional Review Board of University of Illinois at Urbana-Champaign and complied with APA ethical standards in the treatment of participants. All participants gave informed consent prior to inclusion in the study. Sample size determination was similar to the previous experiments.

4.1.2. Apparatus

The apparatus, calibration procedures, and screen resolution was identical to Experiment 2.

4.1.3. Stimuli

The stimuli were the same as in Experiment 1.

4.1.4. Procedure

The experimental procedure was similar to Experiment 1 with a few exceptions. Instead of the DF manipulation, we used repetition manipulation to strengthen some pairs. We dropped the DF instructions and employed “Remember-all trials”. Half of the object-scene pairs were presented once, whereas the remaining half were shown twice. Participants studied 54 object-scene pairs using the same timing and encoding instructions as in Experiment 1. Afterwards, they were shown 108 object-scene pairs, half of which were presented for the first time, whereas the remaining half of the pairs were presented the second time, as they had been shown previously (the pairing of objects with their scenes was preserved across repetitions). The selection of Once- and Twice-presented pairs was determined randomly, and so was their presentation order. The test procedure was similar to the previous experiments. The test displays consisted of a scene along with either all Once-presented or all Twice-presented objects, thereby controlling for item strength within the test display. After selecting the object, participants were asked to provide one of three confidence judgments regarding their object selection (Low, Medium, or High), and they had unlimited time to provide their confidence rating.

4.2. Results

4.2.1. Recognition accuracy

Recognition accuracy was analyzed with a multilevel logistic regression model using *Presentations* (Once vs. Twice) as a fixed effect and *Participants* as a random intercept for the fixed effect of presentations. Recognition accuracy was significantly higher for twice-presented (M = 0.88, SD = 0.33) than once-presented items (M = 0.71, SD = 0.45), β = 1.22, SE = 0.18, z = 6.72, p < .001, replicating successful manipulation of memory strength through repetitions.

4.2.2. Eye movement analyses

4.2.2.1. Preferential viewing analyses. Fig. 6 summarizes mean proportion of viewing (averaged across participants and across trials) to a Selected Object (Target vs. Lure) by Presentation Status (Once vs. Twice) and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response on each trial. To assess preferential viewing, a mixed effects model was fit to the proportion of viewing time on each trial, using *Presentations* (Once vs. Twice), *Selected Object* (Target vs. Lure), and *Bin* (four 500 ms time bins, equating to 2 s prior to response) as fixed effects, and *Participants* as a random intercept. There was a significant effect of Selected Object, β_{object} = 0.10, SE = 0.02, t = 5.39, p < .001, indicating disproportionately more viewing to the Selected Target than to the Selected Lure, confirming preferential viewing. There was also an effect of Bin, β_{bin} = 0.22, SE = 0.02, t = 12.24, p < .001, indicating that overall viewing increased over time. These effects were qualified by a significant Selected Object × Bin interaction, β = 0.11, SE = 0.04, t = 3.13, p = .002, indicating that the magnitude of preferential viewing varied across the time bins (the variance associated with random intercept of participants was σ^2 = 0.005, SD = 0.07). Viewing reached its peak in the 500 ms time bin prior to response, with approximately 76% of the total viewing time being devoted to the Target over the Lure, β_{object} = 0.16, SE = 0.03, t = 5.27, p < .001. These results confirm the established findings in the literature and the previous two experiments.

To assess whether viewing patterns varied across once or twice presented items, we examined proportion of viewing in the 500 ms time bin prior to response with a mixed effects model, using *Presentations* (Once vs. Twice) and *Selected Object* (Target vs. Lure) as fixed effects, and *Participants* as a random intercept. In addition to the significant effect of Selected Object, β_{object} = 0.18, SE = 0.03, t = 5.38, p < .001, there was a significant Selected Object × Presentations interaction,

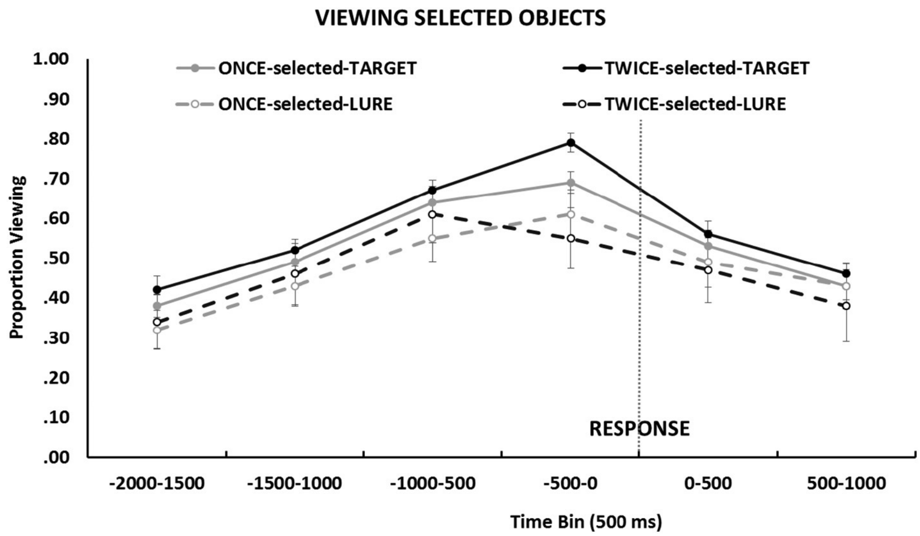

Fig. 6. Mean proportion of viewing time to the Selected Target and Selected Lure as a function of Presentation Status and Time Bins (grouped by 500 ms) in Experiment 3. Error bars represent SE of the mean.

$\beta = 0.15$, $SE = 0.06$, $t = 2.34$, $p = .020$. This interaction arises because of a significant effect of Presentations on viewing to the Selected Target, $\beta_{presentations} = 0.05$, $SE = 0.02$, $t = 2.15$, $p = .032$, indicating greater viewing to Selected Target in the twice-presented condition compared to once-presented condition. However, there was no effect of Presentations on viewing to the Selected Lure, $\beta_{presentations} = -0.07$, $SE = 0.07$, $t = 1.03$, $p = .304$. The net result was that the magnitude of preferential viewing decreased in once-presented compared to twice-presented condition, although it was significant in both cases (in once-presented: $\beta_{object} = 0.10$, $SE = 0.04$, $t = 2.33$, $p = .020$; in twice-presented: $\beta_{object} = 0.28$, $SE = 0.05$, $t = 5.50$, $p < .001$).

4.2.2.2. Analyses of retained traces. Fig. 7 summarizes mean proportion of viewing on incorrect trials (averaged across participants and across trials) to Unselected Objects (Target vs. Lure), by Presentations (Once vs. Twice) and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response. A mixed effects model was fit to proportion of viewing time on incorrect trials (collapsed across bins) using *Presentations* and *Unselected Object* as fixed effects, and *Participants* as a random intercept. There was only an effect of Unselected Object, $\beta_{object} = 0.05$, $SE = 0.02$, $t = 2.31$, $p = .021$, indicating overall greater viewing to Unselected Target than Unselected Lure (the variance associated with random effect was $\sigma^2 = 0.002$, $SD = 0.04$). None of the remaining effects were significant. There was neither effect of Presentations in viewing to Unselected Targets, $\beta_{presentations} < 0.01$, $SE = 0.03$, $t = 0.13$, $p = .895$, nor in viewing to Unselected Lures, $\beta_{presentations} = 0.02$, $SE = 0.03$, $t = 0.611$, $p = .541$.

4.3. Discussion

Experiment 3 revealed strength-based differences in recognition accuracy resulting from objective repetitions. The eye movements revealed reduced preferential viewing in the Once-compared to Twice-presented condition that stemmed from decreased viewing to the Selected Target in the weaker memory condition. There were no differences in viewing to the Selected Lure across strong and weak conditions. Note, that this viewing pattern was similar to Experiment 2, where weaker conditions (i.e., Low or Medium confidence) were associated with less viewing to the Selected Target compared to the stronger condition (i.e., High confidence). Furthermore, no differences in

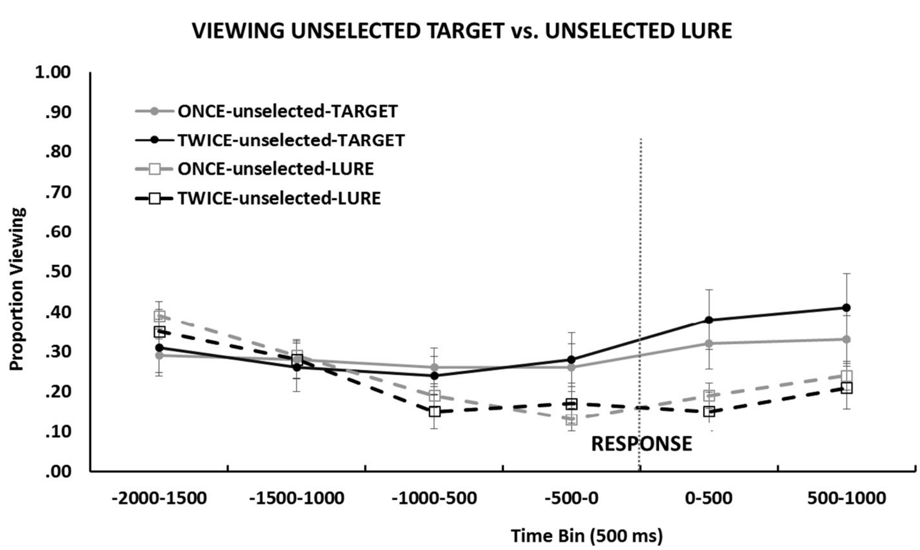

Fig. 7. Mean proportion of viewing time on incorrect trials directed towards the Unselected Target and Unselected Lure as a function of Cue and Time Bins in Experiment 3. Error bars represent SE of the mean.

viewing to the Unselected Target were obtained in Experiment 3 across presentation conditions, a finding that confirms what was observed in Experiment 2. Therefore, in Experiment 3, where we objectively manipulated memory strength through repetitions, the viewing behavior resembled what was observed in Experiment 2, and it was the opposite of what was observed in Experiment 1.

5. Experiment 4: list-method DF

The previous three experiments revealed viewing differences resulting from DF and those obtained through objectively strengthening memory via repetition, or across subjective memory strength as assessed through confidence ratings. The viewing behavior in DF was the opposite of what was observed in the previous two experiments. If viewing patterns observed in Experiment 1 are unique to DF, they may also emerge in a LMDF procedure. Assessing LMDF through the eye movements may yield new insights because traditionally LMDF has been absent in tests of single item recognition (for reviews, see Sahakyan & Foster, 2013; Sahakyan, in press), and also associative recognition (Hanczakowski et al., 2012). Thus, eye movements may reveal patterns that go undetected in measures of recognition accuracy. There are no studies of LMDF using eye-movements in the literature. Given eye movements' sensitivity to subtle distinctions in memory in ways that behavioral accuracy is not (Hannula et al., 2012; Hannula & Ranganath, 2009; Ryan et al., 2000; Ryan and Cohen, 2004), we wanted to investigate if eye movements might distinguish between Remember and Forget conditions, despite a null LMDF effect in behavioral accuracy. Comparing the eye movement patterns between the two DF studies would be theoretically informative. If the same pattern emerges between Experiment 1 and the current study, it would potentially generalize the effect of DF on the viewing behavior and differentiate DF from other strength-based effects observed in Experiments 2 and 3. In addition, finding reduced evidence of retained memory traces for target objects in the LMDF study (akin to what was observed in Experiment 1) would be consistent with the inhibitory mechanism of LMDF. On the other hand, equivalent viewing towards the Unselected Target across the Forget and Remember condition would be more consistent with non-inhibitory interpretations of LMDF.

5.1. Methods

5.1.1. Participants & sample size

Participants were 80 students from the University of Illinois who were compensated with course credit. The data file was corrupted for one participant, and analyses was conducted on 79 participants. The study was approved by the Institutional Review Board of University of Illinois at Urbana-Champaign and complied with APA ethical standards in the treatment of participants. All participants gave informed consent prior to inclusion in the study. Sample size was determined similar to the previous experiments, except that the nature of critical interaction was between a within-subjects variable (either List, or Selected Object), and a between-subjects variable of Cue. A priori power analyses for a small effect size interaction showed that we needed a minimum of 72 participants at power equal to 0.80, with $\alpha = 0.05$.

5.1.2. Apparatus

Eye position was recorded at a rate of 250 Hz using an Eyelink II eye-tracking system (SR Research). Eye position was calibrated using a 3×3 spatial array, in a process completed after the study block and prior to the test block. Calibration ended with participants fixating on a centrally located cross-hair, beginning the test block. The computer screen resolution was set to 1280×1024.

5.1.3. Stimuli

The stimuli were identical to Experiment 1, consisting of 108 object-scene pairs divided between equally List 1 and List 2 (see procedure for more details).

5.1.4. Procedure

The procedures were similar to Experiment 1 with the exception of employing a list-method DF procedure. This meant splitting the entire stimulus set into two separate lists (List 1 and List 2), each consisting of 54 object-scene pairs, for a total of 108 study trials. At the beginning of encoding, participants were told they would be studying two separate lists of object-scene pairs. After List 1, half of the participants were told they were then going on to study another list of objects (Remember condition), and proceeded to study object-scene pairs. The other half of participants experienced a simulated computer crash at the end of List 1 and were instructed that they would not be tested on that list of objects due to computer failure and that they should forget those objects. They then proceeded to study List 2 with the impression that they would only be tested on objects from that list. The short break between List 1 and List 2 was consistent between the Remember and Forget conditions. At test, DF instructions were cancelled for the group that experienced the computer crash, and all participants were told they would be tested on all object-scene pairs regardless of which list they came from. All three objects in the test display came from either List 1 or List 2, in order to control for item strength within each trial. Upon selection of the object, participants provided binary confidence judgments following the procedures of Experiment 1.

5.2. Results

5.2.1. Recognition accuracy

Recognition accuracy was analyzed with a multilevel logistic regression model using *Cue* (Remember vs. Forget) and *List* (List 1 vs. List 2) as fixed effects, and *Participants* as a random intercept for the fixed effects of cue and list. There was no effect of cue, indicating that recognition accuracy did not differ between the Forget and Remember groups, $\beta_{\text{cue}} = 0.66$, $SE = 0.39$, $z = 1.71$, $p = .09$ (Forget List 1: $M = 0.70$, $SD = 0.46$; List 2: $M = 0.72$, $SD = 0.45$; Remember List 1: $M = 0.76$, $SD = 0.43$; List 2: $M = 0.73$, $SD = 0.44$). There was neither the effect of List, $\beta_{\text{list}} = -0.02$, $SE = 0.09$, $z = 0.27$, $p = .786$, nor the Cue $\times$ List interaction, $\beta = -0.32$, $SE = 0.18$, $z = 1.76$, $p = .08$. These results confirm that there was no DF effect in this study, replicating the null DF effect in associative recognition reported previously (Hanczakowski et al., 2012).

5.2.2. Eye movement analyses

5.2.2.1. Preferential viewing analyses. Fig. 8 summarizes mean proportion of viewing (averaged across participants and across trials) to a Selected Object (Target vs. Lure) by Cue (R vs. F) and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response on each trial. Note that the data is collapsed across the lists in the Forget and Remember conditions. We examined viewing patterns in each list separately, but since the same pattern emerged in both lists, we are presenting the combined figure.

To assess preferential viewing, a mixed effects model was fit to the proportion of viewing time on each trial, using *Cue* (F vs. R), *Selected Object* (Target vs. Lure), and *Bin* (four 500 ms time bins, equating to 2 s prior to response) as fixed effects, and *Participants* as a random intercept. There was a significant effect of Selected Object, $\beta_{\text{object}} = 0.07$, $SE = 0.03$, $t = 2.66$, $p = .008$, indicating disproportionately more viewing to the Target than to the Lure, confirming preferential viewing. There was also an effect of Bins, $\beta_{\text{bin}} = 0.14$, $SE = 0.03$, $t = 5.77$, $p < .001$, indicating that overall viewing increased over time. These effects were qualified by a significant Selected Object $\times$ Bin interaction, $\beta = 0.29$, $SE = 0.05$, $t = 5.87$, $p < .001$, indicating that the magnitude of preferential viewing varied across the time bins (the variance associated with random intercept of participants was $\sigma^2 = 0.003$, $SD = 0.06$). Viewing reached its peak in the 500 ms time bin prior to response, with approximately 67% of the total viewing time

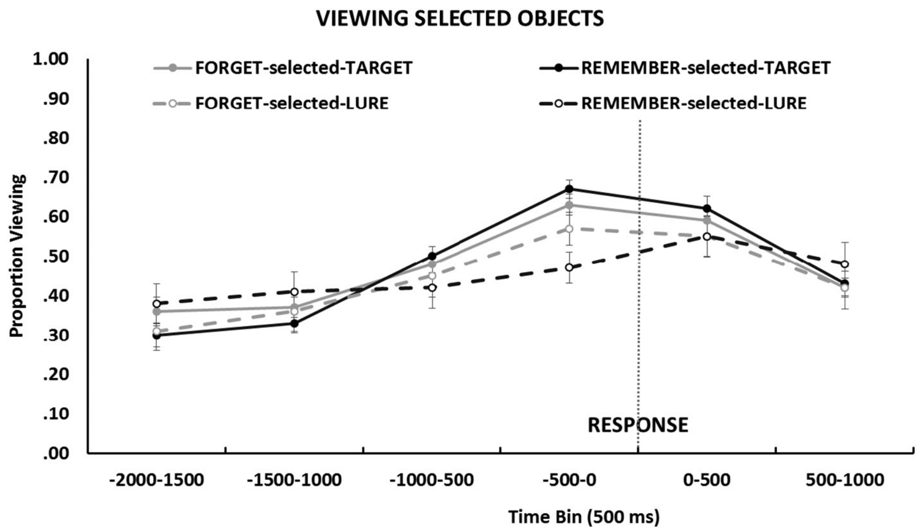

Fig. 8. Mean proportion of viewing time to the Selected Target and Selected Lure as a function of Cue and Time Bin (grouped by 500 ms) in Experiment 4. Error bars represent SE of the mean.

being devoted to the Target over the Lure, $\beta_{\text{object}} = 0.15$, $SE = 0.02$, $t = 7.73$, $p < .001$. These results confirm the established findings in the literature as well as the previous experiments.

To assess whether viewing patterns varied between Forget and Remember conditions, we examined proportion of viewing in the 500 ms time bin prior to response with a mixed effects model, using *Cue* (F vs. R) and *Selected Object* (Target vs. Lure) as fixed effects, and *Participants* as a random intercept. In addition to the significant effect of Selected Object, $\beta_{\text{object}} = 0.15$, $SE = 0.02$, $t = 7.57$, $p < .001$, there was a significant Selected Object × Cue interaction, $\beta = 0.13$, $SE = 0.04$, $t = 3.33$, $p < .001$. This interaction arises because of a significant effect of Cue on viewing to the Selected Lure, $\beta_{\text{cue}} = -0.09$, $SE = 0.04$, $t = 2.15$, $p = .037$, indicating greater viewing to the Selected Lure in the Forget condition compared to Remember condition. However, there was no effect of Cue on viewing to the Selected Target, $\beta_{\text{cue}} = 0.04$, $SE = 0.03$, $t = 1.18$, $p = .244$. The net result was that the magnitude of preferential viewing decreased in the Forget compared to Remember condition, although it was significant in both cases (in Remember: $\beta_{\text{object}} = 0.22$, $SE = 0.03$, $t = 8.13$, $p < .001$; in Forget: $\beta_{\text{object}} = 0.08$, $SE = 0.03$, $t = 2.81$, $p = .005$).

5.2.2.2. Analyses of retained traces. Fig. 9 summarizes mean proportion of viewing on incorrect trials (averaged across participants and across trials) to Unselected Objects (Target vs. Lure), by Cue (F vs. R) and Time Bins (in increments of 500 ms), shifted to align with respect to behavioral response. The same mixed level model was used to assess the proportion of viewing time on incorrect trials as in the previous experiments. There was only an effect of Unselected Object, $\beta_{\text{object}} = 0.03$, $SE = 0.01$, $t = 1.97$, $p = .049$, indicating greater viewing to the Unselected Target than the Unselected Lure (the variance associated with random intercept of participants was $\sigma^2 < 0.001$, $SD = 0.03$). None of the remaining effects were significant. Unlike in Experiment 1, there was no effect of Cue in viewing to Unselected Target in the current experiment, $\beta_{\text{cue}} < 0.01$, $SE = 0.02$, $t = 0.152$, $p = .88$.

5.3. Discussion

There was no DF effect in behavioral accuracy in Experiment 4, consistent with other findings in the literature indicating the absence of DF in recognition tests, including in associative recognition. The eye movement behavior revealed a reduction in preferential viewing in DF

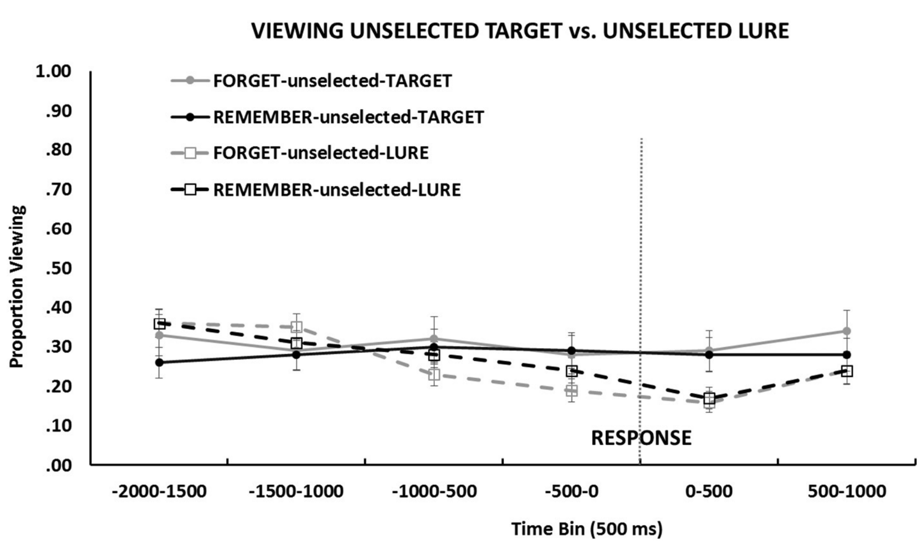

Fig. 9. Mean proportion of viewing time on incorrect trials directed towards the Unselected Target and Unselected Lure as a function of Cue and Time Bins in Experiment 4. Error bars represent SE of the mean.

that was driven by increased viewing to the Selected Lure in the Forget condition compared to the Remember condition. At the same time, there were no differences in viewing to the Selected Target between the cue conditions. Thus, the effect of DF on preferential viewing in the current experiment replicated the findings of Experiment 1 (while also contrasting with those observed in Experiments 2 and 3, which were driven by the opposite patterns of viewing). However, the current experiment contrasts with Experiment 1 in important ways because unlike Experiment 1, no differences emerged in viewing to the Unselected Target between the Forget and Remember conditions in the current experiment. Therefore, the effects of DF on preferential viewing in the current experiment were similar to that in Experiment 1, whereas the effects of DF on retained traces revealed that the magnitude of retained traces for forgotten targets was similar in the Forget and Remember conditions.

6. General discussion

Intentional forgetting can be an adaptive process for navigating a complex world. How exactly intentional forgetting is accomplished is a topic of debate in the literature. In this investigation, we used a novel approach by examining eye movements that accompany behavior to make progress on the theoretical mechanisms of DF. We pursued two major goals – (1) to examine how successful intentional forgetting is expressed in eye movement behavior, and whether it differs from incidental forgetting that happens despite the intention to remember, and (2) how strength based differences in memory are expressed in eye movements in general. In Experiment 1, we varied strength using the item-method DF manipulation. In Experiment 2, memory strength was assessed through subjective confidence ratings. In Experiment 3, strength was varied through objective repetitions, and in Experiment 4, we used the list-method DF procedure. Note, that there was no DF manipulation in Experiments 2 and 3, and they represent more traditional "remember-all" memory studies.

We adapted a paradigm from neurocognitive literature that has been extensively examined using eye movements. The task involved associative recognition, where participants had to select the target object that was paired with the background scene during encoding. The test displays always contained three items from the same memory instruction to control for the strength of items within the test display (i.e., F-Target with two F-Lures, or R-Target with two R-Lures). From the point of behavioral accuracy, correct selection of the F-Target on a given test trial represents a failure of intentional forgetting because memory for the Target "survived" despite the previous instruction to forget it. In contrast, selection of the F-Lure represents successful intentional forgetting, because participants did not select the Target (presumably because it was forgotten). Finally, selection of the R-Lure represents incidental forgetting, because the target was missed despite the intention to remember. Critically, we compared eye movements associated with these behavioral outcomes. To the best of our knowledge, this is the first study to have coupled eye tracking measures with an item-method DF paradigm using associative recognition testing procedures.

In Experiment 1, there was impaired recognition accuracy for F-items, indicating that item-method DF impaired relational memory. This finding is consistent with other behavioral studies in the literature, documenting impaired relational memory in IMDF using word pairs (e.g., Bancroft et al., 2013; Hockley et al., 2016; Hourihan et al., 2007). The eye movements in the Remember condition replicated the established findings of preferential viewing, indicating that participants devoted a disproportionate amount of time viewing the Selected Target compared to the Selected Lure. Greater viewing towards the Target indicates the influence of memory (as opposed to influence of selection) on eye movement behavior.

The viewing behavior in the Forget condition revealed several novel findings. Namely, failure of intentional forgetting was virtually identical to the Remember condition. That is, on the trials when participants did *not* forget the Target and correctly selected it on the recognition test, they viewed that Target as much as in the Remember condition. In contrast, on the trials when they selected the Lure (presumably because they successfully forgot the Target), participants viewed the Selected Lure more in the Forget condition than in the Remember condition. Overall, there was a reduction in preferential viewing in the Forget condition that was driven by enhanced viewing towards the Selected Lure. Note that this viewing pattern is the opposite of our prediction, where we thought that DF may reduce preferential viewing because participants may view the selected Target *less* in the Forget condition. In contrast, the results showed that preferential viewing was reduced because of the enhanced viewing towards the Selected Lure in the Forget condition.

Importantly, eye movements distinguished successful intentional forgetting from incidental forgetting as evident in differential retention of memory traces across the Forget and Remember conditions in Experiment 1. The critical question was what happened to viewing the two remaining objects whenever participants selected the lure (i.e., Unselected Target and Unselected Lure), and whether there were differences between the Forget and Remember conditions. Whenever participants select a lure, there is no reason to expect differences in viewing the two unselected objects, unless a lingering memory for the Unselected Target influenced the viewing behavior, producing greater viewing towards that Unselected Target than the Unselected Lure (i.e., Nickel et al., 2015). The results showed that whenever participants selected a lure, they tended to view that Unselected Target more than the Unselected Lure, confirming the influence of retained memory traces of Unselected Targets on the viewing behavior. Importantly, however, they viewed the Unselected Target substantially less in the Forget condition than in the Remember condition, indicating that successful intentional forgetting impaired memory beyond what was observed in instances of incidental forgetting, producing differences in the viewing behavior between the two instruction conditions.

To the best of our knowledge, this is the first study to establish eye movements as a marker distinguishing successful intentional forgetting from incidental forgetting. From a theoretical viewpoint, the eye movement findings provide support for the inhibitory account of item-method DF, and are inconsistent with the selective rehearsal view. The inhibitory account of IMDF makes a prediction that successful DF should be distinguished from incidental forgetting – in this case, it predicts that successful intentional forgetting should show suppressed viewing to the Unselected Target in comparison with incidental forgetting. In contrast, the selective rehearsal account does not make this prediction because it suggests that impaired memory of F-items arises from terminating rehearsal of those items. Thus, passively forgotten R-items and successfully forgotten F-items should produce similar signatures in eye movements. Our findings were inconsistent with the rehearsal view, and supported the inhibitory view.

With an exception of few important differences, Experiment 1 findings were replicated in Experiment 4. First, there was no impairment in behavioral accuracy between the Forget and Remember conditions in LMDF, consistent with other null findings in the literature (Hanczakowski et al., 2012). Despite the null effect in behavioral accuracy, a reduction in preferential viewing was observed also in Experiment 4, and it was driven by enhanced viewing towards the Selected Lures in the Forget condition, replicating Experiment 1. However, in contrast to IMDF, LMDF revealed *equivalent* retention of memory traces for Unselected Targets between the Forget and Remember conditions. Although both instruction conditions showed greater viewing towards the Unselected Targets than the Unselected Lures, this effect was of the same magnitude between the Forget and Remember conditions. This aspect of divergent findings between the two DF studies is theoretically important because it suggests that LMDF does not share the inhibitory mechanisms that appear to be implicated in IMDF. In Experiment 1, successful intentional forgetting was evident in impaired retention of target memory traces in the Forget condition, consistent with the inhibitory view. This was not the case in Experiment 4, where the Remember and Forget conditions revealed similar evidence of retained memory for Target items as expressed in the viewing behavior.

Viewing behavior across two DF studies were in stark contrast with those observed in strength-based differences in memory in a purely "remember-all" context. Namely, in Experiment 2, recognition accuracy was highest in the high confidence condition, and lowest in the low confidence condition, replicating the well-known relationship between confidence and memory strength (e.g., Tekin et al., 2018; Wixted & Wells, 2017). In Experiment 3, we objectively manipulated memory strength by presenting half of the object-scene pairs twice during the encoding stage, and obtained recognition accuracy differences that tracked the repetitions. Neither Experiment 2 nor Experiment 3 contained any DF manipulation, and both experiments involved strength-based differences in recognition accuracy. Although in general, preferential viewing was reduced in both Experiment 2 and Experiment 3, it was driven by a different mechanism than what was observed in DF studies. Namely, participants were less likely to view the Selected Targets in the "weak" conditions than in the "strong" conditions of Experiment 2 and Experiment 3, whereas viewing to the Selected Lures remained invariant across strengths. This is the opposite of what was observed in DF studies, where reduced preferential viewing stemmed from enhanced viewing towards the Selected Lures, while viewing to the Selected Targets remained invariant across Forget (weak) and Remember (strong) conditions. In addition, when differential memory traces for unselected targets were observed in Experiment 2, it was also driven by a different mechanism than what was observed in IMDF. Namely, participants were less likely to view the Unselected Lures in the "strong" condition than in the "weak" conditions of Experiment 2. In Experiment 3, on the other hand, no differences were observed in retained memory traces of strong and weak targets, which is also inconsistent with IMDF findings, but replicates LMDF findings. Overall, across a range of measures, the viewing patterns between the strong and weak memories in the "remember-all" context were either inconsistent or opposite of what we observed in strong and weak memories in the "forgetting" context, indicating that the goals that participants pursued affected subsequent eye movements observed at retrieval. Assessing memory strength with objective manipulations as well as through subjective assessments is not an uncommon practice in eye-tracking research such as in pupillometry (e.g., Otero et al., 2011). The fact that both objective and subjective memory strength was reflected in similar patterns of viewing behavior, and contrasted with viewing behaviors observed in DF, provides generalizability of findings and strengthens the conclusions.

In the current investigation, we adopted a modified paradigm that had been widely used to assess different populations characterized by relational memory impairment, namely in amnesia (Hannula et al., 2007, 2015, 2006; Ryan et al., 2000; Ryan & Cohen, 2004), aging (Ryan et al., 2007), and schizophrenia (Williams et al., 2010; Hannula, Althoff, et al., 2010). To the best of our knowledge, this is the first study to investigate the impact of strength-based differences on preferential viewing as a general principle of memory. Overall, eye movement analyses revealed important differences between strong and weak memories that recognition accuracy could not reveal on its own. First, they distinguished successful and unsuccessful DF, providing support for the inhibitory account of item-method DF, and posing a challenge for the selective rehearsal account. Second, eye movements also distinguished between strong and weak memories in a context of intending to "forget some items" from intending to "remember-all" items of varying strengths. Namely, when the goal is to remember, weaker items lead to reduced preferential viewing compared to strong items, and this is reflected in viewing weaker Selected Targets less than stronger Selected Targets. In contrast, when the goal is to forget some items, weaker/Forget conditions lead to reduced preferential viewing, but this is reflected in greater viewing towards the Selected Lures in the Forget than the Remember condition rather than reduced viewing towards the weak targets. Third, the effect of inhibitory mechanisms in item-method DF on eye movements were not obtained in list-method DF, suggesting that list-method DF does not appear to share inhibitory mechanisms that drive the impairment in item-method DF. The current findings suggest strength-based differences lead to different magnitudes in preferential viewing, and that the pattern of eye movements leading to reduced preferential viewing is dependent on the context in which strength-based differences emerge.

Typical DF studies investigate successful intentional forgetting through items that participants failed to forget, which amounts to studying anti-forgetting despite an intention to do so. More recent investigations using concurrent measures such as EEG and fMRI have begun to dissociate incidental from successful intentional forgetting by identifying differential brain activity between the two types of forgetting (for a review, see Anderson & Hanslmayr, 2014). By doing so, these studies have argued for inhibitory mechanisms associated with successful intentional forgetting, contrasting with the traditional view of successful DF resulting from terminating rehearsal of F-items during encoding. Using eye-tracking as a concurrent measure of behavior allowed shedding light on the theoretical debate of DF, by revealing markers of inhibition in eye movements.

CRediT authorship contribution statement

Jonathon Whitlock:Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing.**Yi-Pei Lo**:Investigation, Software, Methodology, Resources.**Yi-Chieh Chiu**:Conceptualization, Methodology, Formal analysis.**Lili Sahakyan**:Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Appendix A. Supplementary data

Data and syntax from all four experiments are available at https://osf.io/jxcvu/. Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104391.

References

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410*, 366–369.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences, 18*, 279–292.

Bancroft, T. D., Hockley, W. E., & Farquhar, R. (2013). The longer we have to forget the more we remember: The ironic effect of postcue duration in item-based directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 691–699.

Basden, B., Basden, D., & Gargano, G. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 603–616.

Bates, B., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistics Software, 67*, 1–48.

Baym, C. L., Khan, N. A., Pence, A., Raine, L. B., Hillman, C. H., & Cohen, N. J. (2014). Aerobic fitness predicts relational memory but not item memory performance in healthy young adults. *Journal of Cognitive Neuroscience, 26*, 2645–2652.

Bjork, E. L., & Bjork, R. A. (2003). Intentional forgetting can increase, not decrease, residual influences of to-be-forgotten information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 524–531.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*, 255–268.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger3rd, & F. I. M. Craik (Eds.). *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale NJ: Erlbaum.

Bjork, R. A. (2011). On the symbiosis of remembering, forgetting, and learning. In A. S. Benjamin (Ed.). *Successful remembering and successful forgetting: A Fetschrift in honor of Robert A. Bjork* (pp. 1–22). New York: Psychology Press.

Bjork, R. A., Laberge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science, 10*, 55–56.

Brysbart, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Cognition, 1*(1), 1–20.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and

extrinsic effects on image memorability. *Vision Research, 116*, 165–178.
Cohen, N. J., & Eichenbaum, H. B. (1993). *Memory, amnesia, and hippocampal function.* Cambridge: MIT Press.
Donaldson, D. I., & Rugg, M. D. (1998). Recognition memory for new associations: Electrophysiological evidence for the role of recollection. *Neuropsychologia, 36*, 377–395.
Donaldson, D. I., & Rugg, M. D. (1999). Event-related potential studies of associative recognition and recall: Electrophysiological evidence for context dependent retrieval processes. *Cognitive Brain Research, 8*, 1–16.
Eichenbaum, H., & Cohen, N. J. (2001). *Oxford Psychology Series; no. 35. From conditioning to conscious recollection: Memory systems of the brain.* New York, NY, US: Oxford University Press.
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. *Behavioral Research Methods, 39*, 175–191.
Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology: General, 145*, 56–81.
Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition, 36*, 1168–1181.
Fawcett, J. M., & Taylor, T. L. (2010). Directed forgetting shares mechanisms with attentional withdrawal but not with stop-signal inhibition. *Memory & Cognition, 38*, 797–808.
Fawcett, J. M., & Taylor, T. L. (2012). The control of working memory resources in intentional forgetting: Evidence from incidental probe word recognition. *Acta Psychologica, 139*, 84–90.
Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112*, 58–72.
Hanczakowski, M., Pasek, T., & Zawadzka, K. (2012). Context-dependent impairment of recollection in list-method directed forgetting. *Memory, 20*, 758–770.
Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010a). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience, 4*(166), 2–16.
Hannula, D. E., Baym, C. L., Warren, D. E., & Cohen, N. J. (2012). The eyes know: Eye movements as a veridical index of memory. *Psychological Science, 23*, 278–287.
Hannula, D. E., & Ranganath, C. (2009). The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron, 63*, 592–599.
Hannula, D. E., Ranganath, C., Ramsay, I. S., Solomon, M., Yoon, J., Niendam, T. A., ... Ragland, J. D. (2010b). Use of eye movement monitoring to examine item and relational memory in schizophrenia. *Biological Psychiatry, 68*, 610–616.
Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *Journal of Cognitive Neuroscience, 19*, 1690–1705.
Hannula, D. E., Tranel, D., Allen, J. S., Kirchhoff, B. A., Nickel, A. E., & Cohen, N. J. (2015). Memory for items and relationships among items embedded in realistic scenes: Disproportionate relational memory impairments in amnesia. *Neuropsychology, 29*, 126–138.
Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The long and the short of it: Relational memory impairments in amnesia, even at short lags. *Journal of Neuroscience, 26*, 8352–8359.
Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso, & R. L. Solso (Eds.). *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Oxford, UK: Erlbaum.
Hockley, W. E., Ahmad, F. N., & Nicholson, R. (2016). Intentional and incidental encoding of item and associative information in the directed forgetting procedure. *Memory & Cognition, 44*, 220–228.
Hourihan, K., Goldberg, S., & Taylor, T. (2007). The role of spatial location in remembering and forgetting peripheral words. *Canadian Journal of Experimental Psychology, 61*, 91–101.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26.
Ludowig, E., Möller, J., Bien, C. G., Münte, T. F., Elger, C. E., & Rosburg, T. (2010). Active suppression in the mediotemporal lobe during directed forgetting. *Neurobiology of Learning and Memory, 93*, 352–361.
MacLeod, C. M. (1998). Directed forgetting. In J. Golding, & C. M. MacLeod (Eds.). *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.
MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review, 6*, 123–129.
Marevic, I., Arnold, N. R., & Rummel, J. (2018). Item-method directed forgetting and working memory capacity: A hierarchical multinomial modeling approach. *The Quarterly Journal of Experimental Psychology, 71*, 1070–1180.
Marevic, I., & Rummel, J. (2020). Retrieval-mediated directed forgetting in the item-method paradigm: The effect of semantic cues. *Psychological Research, 84*, 1–21.
MATLAB (2018). *9.3.0.713579 (R2017b).* Natick, Massachusetts: TheMathWorks Inc.
Monti, J. M., Hillman, C. H., & Cohen, N. J. (2012). Aerobic fitness enhances relational memory in preadolescent children: The FITKids randomized control trial. *Hippocampus, 22*, 1876–1882.
Nickel, A., Henke, K., & Hannula, D. (2015). Relational memory is evident in eye movement behavior despite the use of subliminal testing methods. *PLoS One, 10*, 1–27.
Nowicka, A., Marchewka, A., Jednorog, K., Tacikowski, P., & Brechmann, A. (2011). Forgetting of emotional information is hard: An fMRI study of directed forgetting. *Cerebral Cortex, 21*, 539–549.
Oehrn, C. R., Fell, J., Baumann, C., Rosburg, T., Ludowig, E., Keller, H., Hanslmayr, S., & Axmacher, N. (2018). Direct electrophysiological evidence for prefrontal control of hippocampal processing during voluntary forgetting. *Current Biology, 28*, 3016–3022.
Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*, 1346–1353.
Paz-Caballero, M. D., Menor, J., & Jiménez, J. M. (2004). Predictive validity of event-related potentials (ERP) in relation to directed forgetting effects. *Clinical Neurophysiology, 115*, 369–377.
Psychology Software Tools, Inc. [E-Prime 2.0] (2015). Retrieved from https://www.pstnet.com.
R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
Rizio, A. A., & Dennis, N. A. (2013). The neural correlates of cognitive control: Successful remembering and intentional forgetting. *Journal of Cognitive Neuroscience, 25*, 297–312.
Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1526–1543.
Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science, 11*, 454–461.
Ryan, J. D., & Cohen, N. J. (2004). Processing and short-term retention of relational information in amnesia. *Neuropsychologia, 42*, 497–511.
Ryan, J. D., Leung, G., Turk-Browne, N. B., & Hasher, L. (2007). Assessment of age-related changes in inhibition and binding using eye movement monitoring. *Psychology and Aging, 22*, 239–250.
Ryan, J. D., & Villate, C. (2009). Building visual representations: The binding of relative spatial relations across time. *Visual Cognition, 17*, 254–272.
Sahakyan, L. (2020). Journey to the heart of directed forgetting. In M. Kahana, & A. Wagner (Eds.). *Handbook of memory*Oxford University Press (in press).
Sahakyan, L., & Foster, N. L. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. *Psychology of learning and motivation. Vol. 59m. Psychology of learning and motivation* (pp. 131–189). Academic Press.
Sahakyan, L., & Kelly, M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1064–1072.
Taylor, T. L. (2005). Inhibition of return following instructions to remember and forget. *The Quarterly Journal of Experimental Psychology, 58*, 613–629.
Tekin, E., Lin, W., & Roediger, H. L. (2018). The relationship between confidence and accuracy with verbal and verbal+ numeric confidence scales. *Cognitive Research: Principles and Implications, 3*, 41–48.
Van Hooff, J. C., Whitaker, T. A., & Ford, R. M. (2009). Directed forgetting in direct and indirect tests of memory: Seeking evidence of retrieval inhibition using electrophysiological measures. *Brain and Cognition, 71*, 153–164.
Williams, L. E., Must, A., Avery, S., Woolard, A., Woodward, N. D., Cohen, N. J., & Heckers, S. (2010). Eye-movement behavior reveals relational memory impairment in schizophrenia. *Biological Psychiatry, 68*, 617–624.
Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest. 18*, 10–65.
Wylie, G. R., Foxe, J. J., & Taylor, T. L. (2008). Forgetting as an active process: An fMRI investigation of item-method–directed forgetting. *Cerebral Cortex, 18*, 670–682.