Rapid but Incomplete Degradation of Residual Visual Representations over Time

Tom Mercer and Anna-Maria Markova

School of Psychology

University of Wolverhampton

Author Note

Tom Mercer https://orcid.org/0000-0001-8402-1762

Correspondence concerning this article should be addressed to Tom Mercer, Centre for Psychological Research, School of Psychology, University of Wolverhampton, Wolverhampton, UK, WV1 1LY. Email: t.mercer2@wlv.ac.uk. Telephone: +44(0)1902 321368.

The data that support the findings of this study are openly available in the Open Science Framework project "Rapid but Incomplete Degradation of Residual Visual Representations over Time" at https://doi.org/10.17605/OSF.IO/4JBSV

Participant incentives in Experiment 4 were provided through an Experimental Psychology Society Small Grant Scheme ("Tracking Proactive Interference Over Time"). We are grateful to Professor John Krantz for hosting Experiment 3 on Psychological Research on the Net (https://psych.hanover.edu/research/exponnet.html). The stimuli used in this

experiment were provided courtesy of Michael J. Tarr, Carnegie Mellon University, http://www.tarrlab.org/.

Running head: Proactive interference over time.

Word Count: 12,996, excluding Title, Abstract, Tables, Figures and References.

Abstract

While visual working memory has a short lifetime, residual representations can persist and disrupt currently maintained information. This phenomenon is known as proactive interference, and the present study investigated whether the representations underpinning item-specific proactive interference lose details over time. This would be expected if the memories underlying proactive interference are susceptible to temporal processes such as decay, which is strongly disputed. In four experiments, a modified version of the recent probes task was used, requiring participants to determine whether a probe matched one of two recently presented targets. The probe sometimes matched an untested target from a previous trial, or varied in its resemblance to it, and the amount of time separating trials varied. Results revealed that proactive interference was specific and highly disruptive at very short intervals, but its effect diminished over time. At longer intervals, a milder form of proactive interference was present and produced by probes that were only similar to a recently encountered target. In summary, residual visual representations may remain accurate for a few seconds after encoding, before losing precise details and continuing to endure in an inexact state.

Keywords: Proactive interference, visual working memory, forgetting, time, decay.

Rapid but Incomplete Degradation of Residual Visual Representations over Time

One of the defining characteristics of visual working memory (VWM) is its short lifetime. Numerous studies have demonstrated that VWM is lost over brief delays lasting seconds (e.g., Gold et al., 2005; Krill et al., 2018; Kuuramo et al., 2022; Pertzov et al., 2013; Rademaker et al., 2018; Ricker & Cowan, 2010, 2014), even if just one item must be retained over an interference-free retention interval (Mercer & Barker, 2020). While there is debate as to the mechanisms producing this forgetting (e.g., Nilsson, 2020; Ricker et al., 2014; Souza & Oberauer, 2015; Zhang & Luck, 2009), the brief lifetime of VWM is as much a feature of this system as its limited capacity (Luck & Vogel, 1997).

Nonetheless, the current contents of VWM can also be disrupted by re-exposure to events experienced in the recent past, suggesting the presence of residual representations of old information (e.g., Hartshorne, 2008; Jonides & Nee, 2006; Makovski & Jiang, 2008). This is known as proactive interference (PI), which has two forms (Postle & Brush, 2004; Postle et al., 2004). Item-nonspecific PI builds up over the course of an experimental session and is produced by exposure to a general class of stimuli (e.g., Wickens et al., 1963). In VWM, the effect of item-nonspecific PI was documented by Endress and Potter (2014), who developed the Repeated-Unique paradigm (see Shoval & Makovski, 2022). This procedure is a modification to the change detection task, in which the unique condition only uses stimuli once per trial and the repeating condition continually relies on a small set of stimuli across trials. Endress and Potter estimated that the capacity of VWM was greatly reduced in the repeating compared to unique condition, potentially due to stronger PI in the former condition. This effect has been broadly replicated (e.g., Endress, 2022; Endress & Siddique, 2016; Shoval et al., 2020; Shoval & Makovski, 2021, 2022), though Shoval et al. found that PI was much stronger with a heterogeneous rather than homogenous set of stimuli (and in their fourth experiment, PI was absent in the homogenous condition), while Shoval and

Makovski (2022) reported stronger PI for meaningful, rather than meaningless, stimuli. Other studies, using the canonical change detection task, have suggested a limited impact of item-nonspecific PI on VWM capacity (e.g., Balaban et al., 2019; Lin & Luck 2012).

There is another form of PI – item-specific – that can also damage the retention of visual information (Makovski & Jiang, 2008; McKeown et al., 2014, 2020; but see Oberauer et al., 2017, Experiment 4). Item-specific PI is produced by re-exposure to a specific event encountered in the very recent past, such as an object from a previous trial (e.g., Berman et al., 2009), and it has primarily been demonstrated using Monsell's (1978) recent probes task. In this procedure, participants must remember an array of targets and then determine whether a probe matched any of those targets. Recent Negative (RN) probes match a target from the *previous* trial (Trial *N*-1), and this is usually compared against a Non-Recent Negative (NRN) probe – i.e., an item experienced earlier than Trial *N*-1 (e.g., Makovski & Jiang, 2008; Mercer & Duffy, 2015), a novel probe not previously encountered (e.g., Berman et al., 2009, McKeown et al., 2020, Experiments 2-5), or a mixture of the two (e.g., McKeown et al., 2014). Typically, the RN probe lowers accuracy and/or increases response time in comparison to the NRN probes, which provides a behavioral indicator of PI (see Berman et al., 2009). Furthermore, this occurs regardless of whether the RN probe is compared against a novel stimulus, or a stimulus presented much earlier in the experiment (Mercer & Fisher, 2022).

The lifetime of item-specific PI provides insights into the fate of the underlying representation and its susceptibility to forgetting. However, there are competing views about the persistence of PI, with some accounts expecting a loss of PI over time and others predicting much more durable PI. Temporal models of forgetting, such as theories relying on a decay process or temporal distinctiveness, are consistent with the former possibility, predicting that PI will disappear over time, as old representations become less accessible. For

example, in the time-based resource-sharing (TBRS) model – a contemporary decay theory – maintenance of memory requires attention (Barrouillet et al., 2004; Barrouillet & Camos, 2021). If attention is directed elsewhere, the memory will undergo time-based decay. In the recent probes task, as there is no requirement to maintain old representations encountered on a previous trial, they should undergo decay and PI itself should diminish over time. In this decay-based interpretation, PI may be caused by a carryover of redundant content still lingering in VWM, and this should rapidly decay, in line with the lifetime of the VWM system (see Makovski & Jiang, 2008, and Shoval & Makovski, 2022, for discussion).

In temporal distinctiveness theories such as SIMPLE (Brown et al., 2007), memory traces do not decay, but isolating representations from each other in time increases temporal distinctiveness and reduces the likelihood that they will be confused. As such, events encountered on a previous trial should produce less PI if they are temporally detached from events in the present.

Alternatively, PI may be insensitive to the passage of time. This is predicted by McKeown et al.'s (2014) active/passive conception of memory. According to this view, actively maintaining information in VWM adds noise to the memory trace and degrades it, which accounts for rapid demonstrations of time-dependent forgetting. Conversely, noise is not added to passively held memories, permitting old representations to endure for tens of seconds and produce a long-lasting form of PI.

Current evidence appears more consistent with the active/passive theory, rather than with temporal models of forgetting. This has been determined through studies that have varied the inter-trial interval separating trials in the recent probes task (see Berman et al., 2009). The RN-NRN difference can then be assessed at different inter-trial interval lengths, with temporal models of forgetting predicting a reduction in PI at longer delays. In two studies using this approach, McKeown et al. (2014) reported PI in VWM regardless of the

inter-trial interval length. McKeown et al. (2020) documented similar time-insensitive PI across several experiments, with different stimuli and varying interval lengths. In one experiment, the inter-trial interval was extended to 32 s but responses to RN probes were still less accurate than the NRN probes (though see Mercer & Duffy, 2015, for evidence of a temporal recovery from PI). In another recent probes task study (Mercer et al., 2022), PI was resistant to different manipulations designed to lessen its effect and it was generally durable over time. Hartshorne (2008) also discovered that PI extended over several trials, providing further evidence for its durability and robustness.

This long-lasting PI appears difficult to reconcile with the notion that PI reflects a "carryover" effect from VWM and conflicts with the rapid forgetting of actively maintained visual stimuli found in numerous other studies (e.g., Gold et al., 2005; Krill et al., 2018; Kuuramo et al., 2022; Pertzov et al., 2013; Rademaker et al., 2018; Ricker & Cowan, 2010, 2014). It thereby poses problems for temporal models of short-term forgetting, such as TBRS or SIMPLE. Conversely, the discovery of durable PI is consistent with the notion that passively maintained memories are not distorted over time, and they must therefore be preserving specific and fine details of the original event.

Importantly, however, there may be subtle degradation of residual representations as time passes, even if memories of events from previous trials possess sufficient fidelity to produce PI – i.e., there may have been partial degradation of the old representation. This interpretation offers another way of conceptualizing the PI effect and is compatible with fuzzy trace theory, which distinguishes precise, veridical representations from blurred, gist-based traces (e.g., Brainerd & Reyna, 2002). It is also supported by evidence from Oberauer et al. (2017), who argued that PI in VWM is produced by a familiarity signal from long-term memory (LTM) which affects the decision process.

In verbal working memory, there is already convincing evidence for semantic distortion. For example, in two recognition experiments, Atkins and Reuter-Lorenz (2008) asked participants to study 48 lists containing four semantically related target words. After a short interval, participants responded to a single probe that could be positive (matching one of the targets), novel or a lure. The lures were semantically related to one of the targets and they increased incorrect responses and slowed reaction times in comparison to novel probes. This happened regardless of whether an attention-demanding distractor task was presented during the retention interval. Two subsequent experiments replicated this effect in a recall task, showing that semantic errors (recalling an item related to those on the list) were much more common than phonological errors or other mistakes. Strikingly these mistakes happened despite retention interval lengths of just 3-4 s, and other evidence has shown that semantic distortion may be related to PI (Atkins et al., 2011).

Cases of enduring PI in VWM (e.g., Hartshorne, 2008; McKeown et al., 2014, 2020) may therefore be caused by residual representations that have undergone some distortion over the passage of time, which could be compatible with an (incomplete) decay process. The discovery of such distortion would be important, as the active/passive conception (McKeown et al., 2014, 2020) does not appear to predict any temporally influenced distortion in PI: the rapid decay of the actively maintained representation is replaced with a detailed, specific long-term representation that produces robust, unchanging PI.

Knowing about the fidelity of PI over time is therefore crucial, but it has not been possible to investigate this issue in the standard recent probes task as the PI effect is based on a comparison between RN and NRN probes. As such, the impairment in accuracy on RN trials could be because the memory of the RN probe remains unchanged and resilient to forgetting processes, or it could be because the underlying memory has undergone some degradation, yet the RN probe possesses more familiarity than NRN probes, and hence is

capable of increasing errors. The standard recent probes task cannot be used to gain insights into the state of old representations because RN probes may simply act as a coarse reminder of a recent event.

In summary, understanding the fate of old visual memories (as expressed via PI) has important theoretical implications for understanding short-term forgetting, yet this issue has not been adequately explored due to methodological issues. The notion of a gradual and continuous decline in PI over time, as expected by decay and temporal distinctiveness processes, is insufficient, as several studies have shown PI to be unaffected by different delay periods (e.g., Hartshorne, 2008; McKeown et al., 2014, 2020; Mercer et al., 2022). Nonetheless, while these findings are congruent with the view of durable, time-invariant PI, they cannot be taken as conclusive evidence that residual VWM resist forgetting as previous studies were incapable of detecting more subtle changes to the representations underpinning PI. Indeed, it may be most plausible that residual memories undergo some modest degradation over time, perhaps via temporal decay, but this forgetting is incomplete and allows a broadly accurate representation to endure over time. The purpose of this study was to assess this possibility, considering whether there is some loss in the fidelity of old representations producing PI over time, versus the active/passive interpretation in which PI is entirely insensitive to the passage of time.

Investigating this issue required alterations to the recent probes task, and an approach used by Mercer and Fisher (2022) was adopted. In this study, probes closely resembling a previously seen target in terms of state/arrangement – a “Similar” RN probe – produced PI. Unfortunately, this experiment did not manipulate the amount of time separating trials, meaning it was not possible to determine whether changes to the PI effect were influenced by the passage of time, but further refining this approach allowed such changes to be investigated in the present study.

Table 1

Methodological Arrangements in Experiments 1-4

Exp	Type	*N*	Probes	Retention Interval	Inter-trial Intervals	Decay Intervals	Additional Information
1	Lab	25	RN Similar RN Dissimilar RN Novel Positive	1 s	500 ms or 8 s	4.8 s or 12.3 s	Maximum decay intervals were 6.8 s and 14.3 s.
2	Lab	25	RN Similar RN Dissimilar RN Novel Positive	750 ms or 4.5 s	1 s	4.6 s or 12.4 s	Maximum decay intervals were 6.8 s and 14.3 s.
3	Lab and online	26	RN Similar RN Dissimilar RN Positive	1 s	200 ms or 3.5 s	3.5 s or 6.8 s	On critical trials of interest, there was no probe on Trial *N*-1.
4	Online	93	RN Similar RN Dissimilar RN NRN Positive	400 ms	400 ms, 1.6 s or 4.9 s	2.3 s, 3.5 s or 6.8 s	On critical trials of interest, there was no probe on Trial *N*-1. This study also used a smaller set of blue Fribbles.

Note. N was based on the sample used in the data analysis, excluding any outliers.

In four experiments, a modified recent probes task was employed that introduced different types of RN probe (see Figures 1 and 2). These RN probes varied in the degree to which they matched a previously encountered target, allowing different predictions about PI to be tested. The amount of time separating trials also varied so that changes to PI over time could be recorded. This was intended to explore the strength of the PI effect across different periods of time and determine whether residual representations are fully resistant to change, in line with the passive conception of PI, or whether they do lose some detail and undergo subtle changes over time. A summary of the methodological arrangements in all four experiments is shown in Table 1, and see https://doi.org/10.17605/OSF.IO/4JBSV for diagrams depicting the trial structure and probe types.

Experiment 1

Participants encoded two targets on each trial and then determined whether a single probe matched either of the current targets. The probe type and inter-trial interval (500 ms or 8 s) varied, and the intervals were chosen to broadly match the delays used in some previous studies. Specifically, shorter delays have typically been 500 ms or less, with longer delays around 6-10 s (e.g., McKeown et al., 2014, 2020; Mercer et al., 2022). In one of these studies (Mercer & Duffy, 2015), an 8.3 s inter-trial interval was sufficient to find a reduction in the PI effect. In addition, these two delay lengths are clearly differentiated and should allow plentiful time for any forgetting processes to occur. To calculate a decay interval, the amount of time from the offset of the targets on Trial *N*-1 to the onset of the probe on Trial *N* was calculated. When considering median response time, a short inter-trial interval led to a decay interval of 4.8 s whereas for the longer inter-trial interval the equivalent decay period was

12.3 s[1]. Any reduction in PI over this decay interval would evidence the loss of redundant information from memory.

To assess changes to the representations producing PI, different types of RN probe were used. RN probes matched an untested target from the previous trial, but not the current trial. Similar RN probes had not previously been studied but closely resembled a previously seen (but untested) target (see Mercer & Fisher, 2022). Dissimilar RN probes had also not previously been studied and possessed only a broad overlap with a previously seen target, which served as the main comparison for the RN and Similar RN conditions. Crucially, however, all three of these probe types shared the same level of overlap with one of the current targets. That is, the Similar and Dissimilar RN probes – while novel – were the same distance from the current target as the RN probe. This allowed inferences about the state of old representations to be inferred more directly than the standard recent probes task (images to illustrate the adapted task and types of RN probe can be found at https://doi.org/10.17605/OSF.IO/4JBSV).

If these old representations have undergone some degradation, both RN and Similar RN probes should lower task accuracy below the Dissimilar RN probe. Furthermore, any temporal loss in the fidelity of old representations should lead to poorer performance in the RN condition at the short decay interval, followed by a recovery in this condition at the longer decay interval. Likewise, if the representations causing PI only undergo degradation after sufficient time has passed, PI from the Similar RN probe should only occur at the long decay interval.

A secondary aim was to test for total collapses in VWM. To do so, a Novel probe type was created, which was completely different to both the current targets and the previous

[1] The maximum decay interval was 6.8 s and 14.3 s for the short and long inter-trial intervals, respectively.

targets. Such an obvious mismatch should be easy to detect, but these probes were included as a way of measuring the complete loss of previously viewed targets.

Method

Data for this experiment, and all other experiments, can be accessed at https://doi.org/10.17605/OSF.IO/4JBSV.

Participants. A recent probes study using the same stimuli (Mercer & Duffy, 2015) reported an interaction between the probe type and inter-trial interval ($\eta_p^2 = 0.17$). Based on 80% power and .05 significance in a fully repeated measures design, at least 22 participants were required to detect this effect. The final sample included 28 participants, but one participant was removed due to numerous missing responses (over 40% of trials). The remaining 27 participants (21 female) were aged between 18 and 57 ($M = 28.48$, $SD = 12.44$) and included undergraduate and postgraduate psychology students from the University of Wolverhampton, as well as members of the public. All participants provided written informed consent before undertaking the procedure.

Materials. The Fribble stimulus set was used as targets and probes (https://sites.google.com/andrew.cmu.edu/tarrlab/stimuli?authuser=0; Stimulus images courtesy of Michael J. Tarr, Carnegie Mellon University, http://www.tarrlab.org/). Fribbles are three-dimensional objects containing a body and four appendages, with the color and shape of the appendages varying. Fribbles are divided into three distinct families, and each family contains four species. Items within a species share the same body but have very different appendages, meaning these stimuli are ideally suited for systematically manipulating the similarity between the targets and probes.

In total, 456 Fribbles were used on experimental trials and 14 on practice trials. They were presented on a white background and the three families were equally represented.

However, each specific Fribble was used as a target on only one occasion. Two targets were used on each trial and the pairings were pseudo-random – any two targets could be paired provided they were not from the same family.

Positive probes matched one of the current targets whereas RN probes matched an untested target from the previous trial, as in previous studies (e.g., McKeown et al., 2014, 2020; Mercer et al., 2022). Similar RN probes were selected from the same family and species as one of the untested targets from the previous trial, but its four appendages differed. Dissimilar RN probes came from the same family as one of the previous untested targets, but not the same species. This ensured there was a match between the Fribble families used across successive trials and, crucially, meant that RN, Similar RN, and Dissimilar RN probes all had the same relationship to the target on the current trial – they came from the same family, but not the same species. They were therefore matched in their similarity to the current targets. Novel probes differed in both family and species to the current targets.

Other stimuli included a black fixation cross (Tahoma size 42) and a pure tone (4.8 kHz) used to denote the start of each trial. The experiment was built using SuperLab 5 software (Cedrus, https://www.cedrus.com/superlab/) and run on a PC. The experiment was displayed on a HannsG HP191 19" LCD monitor (Hannspree Europe Holdings B. V.; https://www.hannspree.eu) and the tone was presented via internal or external speakers at approximately 65 dB. The procedure was undertaken in an experimental cubicle with participants seated approximately 60 cm from the monitor. Responses were recorded through a keyboard.

Design and Procedure. The experiment employed a 2 (decay interval: 4.8 s vs. 12.3 s) x 3 (probe type: RN vs. Similar RN vs. Dissimilar RN) repeated measures design. Positive and Novel trials were assessed separately, as they are uninformative about the PI effect.

The procedure was explained to participants, and they completed six practice trials before the main experiment. Each trial began with the simultaneous presentation of a tone and central fixation cross for 300 ms. Two targets were then presented for 1 s and had to be remembered. After an unfilled 1 s delay, a single probe was presented in the center of the screen. Participants had to determine whether it matched either target presented 1 s ago, using the "M" key to denote a match and the "Z" key to denote a mismatch. Up to 3 s was given for a response, and the probe disappeared after a key press. The next trial then began after an inter-trial interval of 500 ms or 8 s.

On half of trials, the probe did match one of the targets. On the remaining trials there was a mismatch, divided equally between the four negative probe types. The experiment contained 192 trials completed in three blocks. For each inter-trial interval, the four negative probe types occurred on 12 trials. This was based on Berman et al. (2009), who used 12 negative probe types in their experiments, as did other recent probe task studies (e.g., McKeown et al., 2014, 2020), indicating that 12 trials would be sufficient to find a PI effect. Trials within a block were presented in a fixed order, but the order of blocks was randomized. The procedure lasted approximately 40 minutes (including two opportunities for breaks), participants were tested individually, and no feedback was given. The experiment was approved by a Faculty Ethics Committee.

Results

Approach to data analysis. Performance on the recent probes task is usually explored through the proportion of correct responses (e.g., Makovski & Jiang, 2008; McKeown et al., 2014, 2020) and/or the response time (e.g., Berman et al., 2009; Campoy, 2012). In a previous study using the recent probes task (McKeown et al., 2020), PI was primarily manifested in correct responses, rather than response times. A similar effect was

found here, so the analyses focused on the proportion of correct responses (though response time analyses are shown in the online supplementary materials). The decision to focus on the proportion of correct responses, rather than the proportion of errors, was to allow comparison with prior studies, which have emphasized accurate responding (e.g., Makovski & Jiang, 2008; McKeown et al., 2014, 2020; Mercer & Duffy, 2015; Mercer et al., 2002; Mercer & Fisher, 2022). However, the error rate has been reported as the main variable of interest in other studies, such as Hartshorne (2008), and the error rate in the present experiments are shown in Table 2.

The main PI effect was assessed by comparing the three RN probe types. These were suitable comparisons as they all possessed the same degree of overlap with one of the current targets, controlling for the difference between a target and probe on Trial *N*. The RN, Similar RN and Dissimilar RN probes then systematically differed from an untested target on Trial *N*-1, becoming increasingly distinct from that target. This allowed the specificity of PI to be tested. Novel probes were assessed separately as they differed from both the current and previous targets, and were used to address a different research question concerning total collapses in VWM. This experiment also followed prior experiments in assessing Positive trials and Negative trials independently (e.g., Berman et al., 2009; Campoy, 2012; McKeown et al., 2022). This is because Positive and Negative trials may be affected by different retention strategies and the former are less insightful about PI. Additionally, in some previous VWM studies (e.g., McKeown et al., 2022; Mercer & Fisher, 2022), performance on Positive trials was the lowest of all conditions (though here, performance in the RN condition tended to be poorer than the Positive condition – see General Discussion).

The PI effect was explored using a frequentist ANOVA, with confidence intervals being calculated according to the method of Jarmasz and Hollands (2009). This was supplemented with the equivalent Bayesian ANOVA, calculated using JASP software (JASP

Team, 2018), which generated a Bayes factor (BF_{10}) for each main effect and interaction[2]. BF_{10} values of 3, 10, 30 or 100 were taken to denote moderate, strong, very strong or extreme support for the alternative hypothesis (Wagenmakers et al., 2018). BF_{10} values lower than 0.33 offer good support for the null hypothesis whereas values between 0.33 and 3 were classed as insensitive.

Preliminary analysis. Overall accuracy on the procedure, based on the proportion of correct responses across all probe types, revealed two low scoring individuals. Indeed, these two individuals were close to chance (0.53 and 0.57, respectively) and over 2.5 *SD*s below the mean (0.86). They were therefore removed, as were trials on which participants had not responded, or responded very quickly (< 150 ms). However, these comprised less than 1.5% of total trials.

PI effects. The proportion of correct responses for each probe type is shown in Figure 1. Responses to RN probes were less accurate than Dissimilar RN probes at both decay intervals, but there was also evidence of disruption from Similar RN probes.

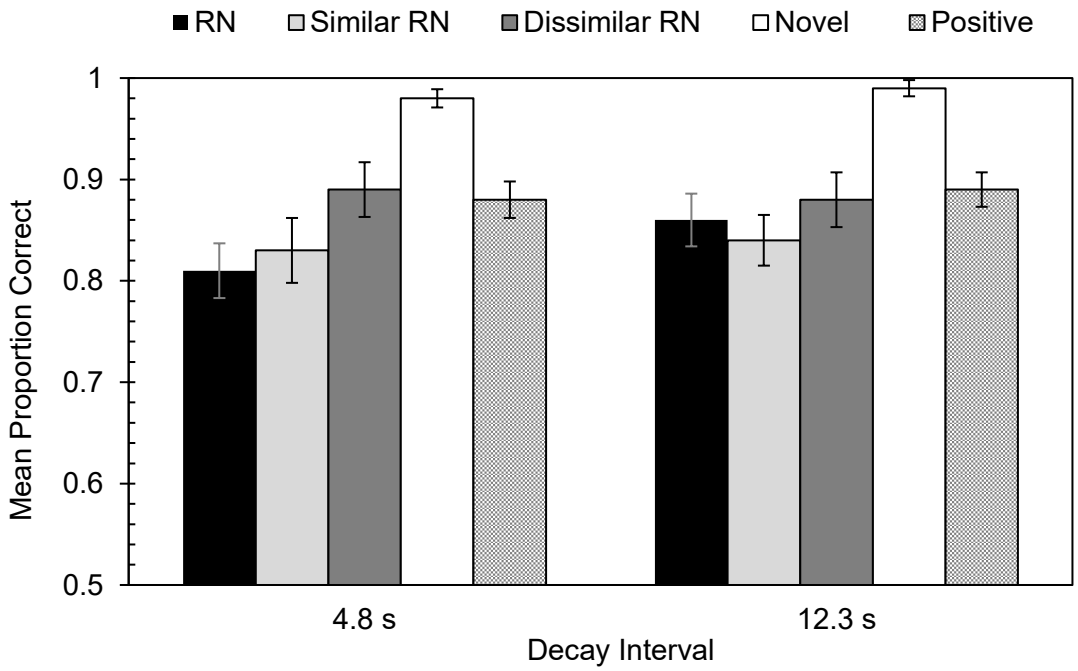
A 2 (decay interval: 4.8 s vs. 12.3 s) x 3 (probe type: RN vs. Similar RN vs. Dissimilar RN) repeated measures ANOVA was performed and there was a significant effect of probe type, with moderate support for the alternative hypothesis, $F(2, 48) = 5.16$, $MSE = 0.01$, $p = .009$, $\eta_p^2 = 0.18$, $BF_{10} = 6.82$. Šidàk and Bayesian post-hoc tests showed that responses to RN probes ($M = 0.83$ [0.80, 0.86]) were less accurate than Dissimilar RN probes ($M = 0.89$ [0.86, 0.92], $p = .02$, $d = 0.68$, $BF_{10} = 3.76$). Responses to Similar RN probes ($M = 0.83$ [0.80, 0.86]) were also less accurate than Dissimilar RN probes ($p = .009$, $d = 0.74$, $BF_{10} = 22.01$), whereas RN and Similar RN conditions did not differ ($p = 1.00$, $d = 0$, $BF_{10} = 0.15$).

[2] In two-way Bayesian ANOVAs in JASP, BF_{10} is calculated for the combined main effect model (effect A + effect B) as well as the combined main effect model plus interaction (effect A + effect B + effect A x effect B). The BF_{10} for the interaction alone can be determined by dividing BF_{10} for the model including the interaction by the model with both main effects. This approach was used here.

The main effect of decay interval was non-significant, $F(1, 24) = 1.37$, $MSE = 0.01$, $p = .254$, $\eta_p^2 = 0.05$, $BF_{10} = 0.34$, as was the interaction, $F(2, 48) = 1.36$, $MSE = 0.01$, $p = .265$, $\eta_p^2 = 0.05$, $BF_{10} = 0.25$, and they were more compatible with the null than alternative hypothesis.

Figure 1

Mean Proportion of Correct Responses in Experiment 1, According to Probe Type and Decay Interval



Note. Error bars show +/-1 *SE*.

Positive and Novel probes. Accuracy on Positive trials was generally very high and comparable to the Dissimilar RN condition. Two-tailed paired-samples frequentist and Bayesian *t*-tests (again conducted using JASP) found little difference between the two decay intervals, $t(24) = -1.64$, $p = .114$, $d = 0.32$, $BF_{10} = 0.68$. Participants were also highly adept at

rejecting Novel probes and responding did not differ according to the decay interval, $t(24) = -1.06$, $p = .298$, $d = 0.21$, $BF_{10} = 0.35$. Both these effects were insensitive but more congruent with the null hypothesis.

Discussion

PI was found in Experiment 1, with the RN and Similar RN conditions lowering task accuracy compared to the Dissimilar RN condition. These data therefore indicated that a stimulus closely resembling a previously seen target could produce PI, suggesting some degradation to memories of events from Trial *N*-1, but there was little evidence that this was affected by the passage of time (the interaction was non-significant and compatible with the null hypothesis). Overall, residual memories underpinning PI appeared to have undergone distortion even at the short decay interval. The memory trace may have been inaccurate following encoding and the data from Experiment 1 do not suggest that any changes to the representations underpinning PI were affected by temporal factors. However, there was also minimal evidence for any total collapses in VWM, as shown by the high performance on Novel trials.

Experiment 2

Experiment 2 varied the retention interval rather than the inter-trial interval. The gap between the targets and the probe was 750 ms or 4.5 s, and these retention intervals were completed in separate blocks. These delay lengths were chosen to match the decay intervals used in Experiment 1[3]. PI effects may be manifested differently when the retention interval is

[3] The maximum decay intervals used in Experiments 1 and 2 were identical. When calculating the median decay interval, based on participant response time, they were also very

varied as this delay affects active maintenance (see McKeown et al., 2014), whereas inter-trial interval manipulations affect passive retention of residual information.

Once again, responses to the RN, Similar RN and Dissimilar RN probe types were of primary interest. If there is a gradual loss in representations from the previous trial, performance should be lowest for RN probes at the short delay. However, after a longer delay PI should be equivalent on RN and Similar RN trials, as the representation undergoes some distortion.

Additionally, performance was expected to decline as the retention interval was lengthened, following previous demonstrations of time-dependent forgetting (e.g., Gold et al., 2005; Krill et al., 2018; Mercer & Barker, 2020; Pertzov et al., 2013; Rademaker et al., 2018; Ricker & Cowan, 2010, 2014). McKeown et al. (2020, Experiment 1) also varied the retention interval within the recent probes task but found that the decline in accurate responding was strongest on Positive trials, with no change on RN and NRN trials. In another study, based on Endress and Potter's (2014) Repeated-Unique paradigm, Shoval and Makovski (2021) reported that changes to the retention interval did not interact with the PI manipulation.

These previous studies suggested that PI is unaffected by the retention interval duration. Experiment 2 assessed the replicability of these findings using the modified recent probes task employed in Experiment 1, which allowed the state of the representations underpinning PI to also be interrogated.

Method

similar. Median decay intervals were 4.8 s and 12.3 s in Experiment 1, and 4.6 s and 12.4 s in Experiment 2.

Participants. Effort was made to recruit a sample size similar to Experiment 1. In total, 26 participants (16 female) aged between 19 and 49 (M = 25.88, SD = 6.85) completed the experiment. Participants were psychology students from the University of Wolverhampton or members of the public, and none had taken part in Experiment 1. All participants provided written informed consent before beginning the procedure.

Materials. Stimuli and equipment were identical to Experiment 1.

Design and Procedure. The arrangements matched Experiment 1, except the retention interval separating the targets and the probe varied – 750 ms and 4.5 s. The inter-trial interval had a fixed length of 1 s, and the short and long retention intervals were completed in separate blocks. The experiment was approved by a Faculty Ethics Committee.

Results

Preliminary analysis. The total proportion of correct responses was again used to assess task accuracy. One individual performed over 2.5 *SD*s below the mean and was removed from the analysis, leaving 25 participants. Missing responses were also removed (these comprised less than 1% of trials, in total).

PI effect. The proportion of correct responses for each probe type is shown in Figure 2. Both RN and Similar RN probes disrupted accuracy, compared to the Dissimilar RN condition, with lowest accuracy for RN probes. This was formally assessed using 2 (retention interval: 750 ms vs. 4.5 s) x 3 (probe type: RN vs. Similar RN vs. Dissimilar RN) frequentist and Bayesian repeated measures ANOVAs. There was a significant effect of probe type and extreme support for the alternative hypothesis, $F(2, 48) = 19.84$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = 0.45$, $BF_{10} = 81{,}487.83$, with Šidàk and Bayesian post-hoc tests finding that accuracy in the RN condition (M = 0.74 [0.71, 0.77]) was poorer than both Similar RN (M = 0.82 [0.79, 0.85], $p = .008$, $d = 0.66$, $BF_{10} = 20.52$) and Dissimilar RN (M = 0.89 [0.86, 0.92], $p < .001$, d

= 1.12, BF_{10} = 81,444.53) conditions. The latter two conditions also differed (p = .004, d = 0.80, BF_{10} = 12.44).

The effect of retention interval was non-significant, $F(1, 24) = 2.11$, $MSE = 0.02$, p = .159, $\eta_p^2 = 0.08$, $BF_{10} = 0.56$, as was the interaction, $F(2, 48) = 1.48$, $MSE = 0.01$, p = .238, $\eta_p^2 = 0.06$, $BF_{10} = 0.29$. Both outcomes were more compatible with the null hypothesis, though the retention interval effect was insensitive.

Positive and Novel probes. Responses to Positive probes at the 750 ms retention interval were generally accurate and equivalent to the Dissimilar RN probes. However, accurate responding notably declined over the retention interval and paired-samples frequentist and Bayesian t-tests showed a significant decrease in accuracy and extreme support for the alternative hypothesis, $t(24) = 5.62$, $p < .001$, $d = 1.12$, $BF_{10} = 2{,}457.60$. Conversely, responses to Novel probes were highly accurate at both retention intervals, with no evidence for any change in performance, $t(24) = -0.18$, p = .858, $d = 0.00$, $BF_{10} = 0.21$.

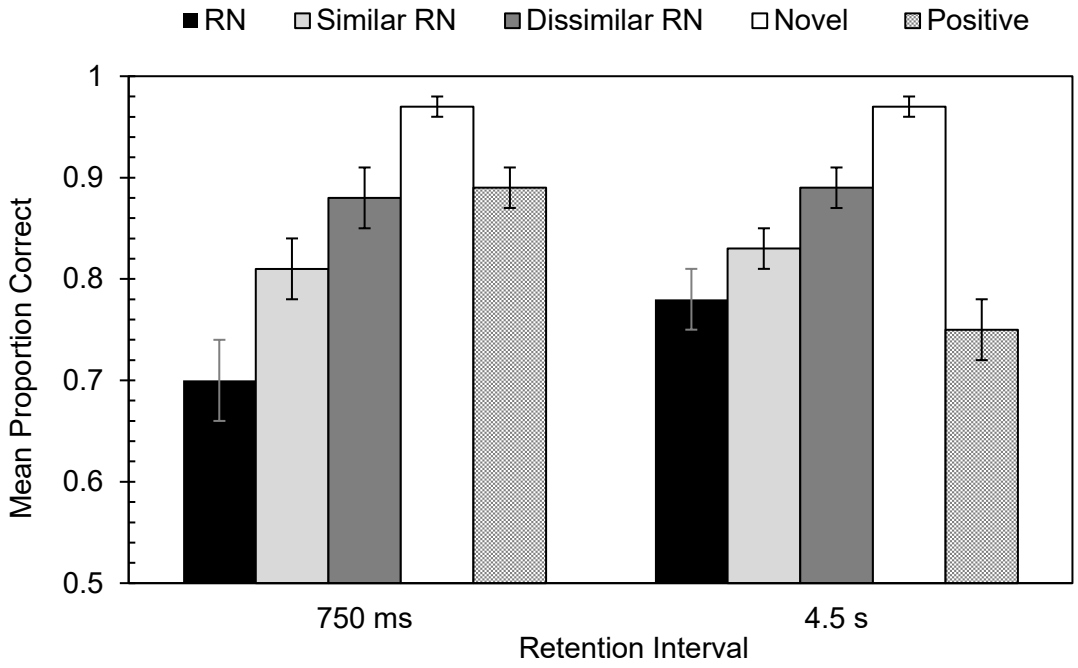
Discussion

PI was present and it did not reliably change according to the retention interval (i.e., there was evidence for persistent PI once again), supporting McKeown et al. (2020) and Shoval and Makovski (2021). Additionally, the Similar RN probe was capable of disrupting task performance in comparison to the Dissimilar RN probe, matching Experiment 1. These data suggest that representations of old events do lose accuracy, allowing PI to be caused by stimuli that resemble a prior event. The Positive trial data were also compatible with this view. Specifically, the ability to detect target-probe matches strongly declined over the retention interval, suggesting that a precise target memory does degrade over a relatively brief period. Yet participants had no trouble rejecting Novel probes at either retention

interval, indicating that any trace degradation may reflect gradual, minor changes, rather than a complete collapse of the target memory (see also Rademaker et al., 2018).

Figure 2

Mean Proportion of Correct Responses in Experiment 2, According to Probe Type and Retention Interval



Note. Error bars show +/-1 *SE*.

Unlike Experiment 1, the RN probe was more damaging than the Similar RN probe. This may have been influenced by the blocked nature of Experiment 2, in which short and long retention interval conditions were completed separately, whereas Experiment 1 intermixed the inter-trial intervals within a block. As such, when the retention interval was short in Experiment 2, trials progressed at a rapid pace, and this may have increased intrusion errors. Indeed, the RN/Similar RN difference was most pronounced at the short retention

interval and there were hints of stronger PI from RN probes at the shortest decay interval. Nonetheless, the crucial interaction was non-significant and changes to old representations may have occurred too rapidly to have been reliably detected in Experiments 1 and 2. Experiment 3 therefore reduced the length of the decay interval.

Experiment 3

In Experiment 1, averaged responses to the RN probe at the short decay interval were 5% lower than the longer interval. In Experiment 2, responses to the RN probe increased by 8% as the delay was lengthened. Conversely, there was no change within the Similar RN condition, hinting that specific features from the representation underlying PI may degrade quickly over time. This possibility has implications for understanding the fate of old representations, as temporal models of forgetting such as decay would expect temporal degradation to occur, but accounts such as the active/passive model do not.

To assess more rapidly occurring changes, Experiment 3 shortened the inter-trial intervals further, using a modification to the recent probes task outlined by Campoy (2012; see also McKeown et al., 2014, Experiment 2). This was achieved by introducing trials without a probe, allowing the decay intervals to be reduced to 3.5 s and 6.8 s, respectively. Some evidence indicates that the loss of VWM may occur rapidly, within several seconds (e.g., Kuuramo et al., 2022; Ricker & Cowan, 2010, 2014), and with verbal material there is evidence for forgetting occurring within around 4 s (Mueller & Krawitz, 2009) or less (Campoy, 2012). By shortening the decay interval, any rapid degradation to memories causing PI could be measured. If there was a loss of fine details in an old target memory, the RN probe should produce the strongest PI effect at the short decay interval, followed by a

recovery in performance, as features in the RN probe representation was forgotten. Likewise, the Similar RN probe should only damage performance at the longer decay interval.

Method

Participants. Effort was made to obtain a sample size similar to Experiment 1. Nine participants were recruited for a laboratory-based version of the experiment, before a national lockdown in response to Covid-19 was introduced in the UK. At this point, the experiment was moved online and a further 24 participants were recruited through the website "Psychological Research on the Net" (https://psych.hanover.edu/research/exponnet.html). Of these 33 participants, one asked to withdraw their responses, two were removed due to numerous missing responses and one had technical problems during the procedure.

The final sample included 29 participants (17 female) aged between 18 and 57 (M = 27.25, SD = 10.34; one age not reported). Participants were University of Wolverhampton psychology students or members of the public, and none had taken part in Experiment 1 or 2. All participants provided informed consent before beginning the procedure.

Materials. Stimuli and equipment for the laboratory-based version of the study matched Experiment 1, but a new end marker image was created to indicate the end of a trial, given that the probe was not always present. This consisted of three white hashtags presented against a red background.

The online version of the experiment was run using the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020), but participants could only complete the study using a desktop computer or laptop. The online experiment was identical to the laboratory version, except the tone was removed (due to challenges ensuring that participants had speakers turned on at the intended volume).

Design and Procedure. Another 2 (decay interval: 3.5 s vs. 6.8 s) x 3 (probe type: RN vs. Similar RN vs. Dissimilar RN) repeated measures design was used. The major events of interest contained 144 pair of trials, in which the first trial did not include a probe and the second did. Of these, there were 12 trials for each RN probe at each decay interval, matching Experiments 1 and 2, and 36 trials with a Positive probe. The first trial started with a fixation cross (300 ms) followed by the two targets (displayed for 1 s) and the retention interval (1 s). Participants were then shown the end marker for 200 ms, indicating that the trial was over. Another trial followed that did include a probe, displayed for up to 3 s. The 3.5 s decay interval had no gap separating the end marker from the start of the new trial, whereas the 6.8 s decay interval had a blank 3.3 s delay separating the end marker from the start of the next trial. The longer decay interval matched the maximum decay periods in Experiments 1 and 2, and as little change to PI was detected beyond this time period, it was deemed suitable for use here.

To avoid predictability, there were 36 trial pairs where neither trial included a probe, as well as 36 trials following the traditional structure (i.e., a standalone trial with a probe). This included six trials for each RN probe type (three at each decay interval) and 18 trials with a Positive probe (nine at each decay interval). These extra trials acted as fillers and were designed to prevent a situation in which a trial with a probe always followed a trial without a probe (though there were not more than three successive trials without a probe).

RN, Similar RN, and Dissimilar RN probes comprised half of all trials. The remaining trials were Positive, with Novel probes being removed given the high performance seen in Experiments 1 and 2.

Participants completed the experiment in three blocks of 72 trials. The order of blocks was randomized, and each block included an equal mixture of probe types, trial types and inter-trial intervals. There were four practice trials before the main experiment, which lasted

approximately 40 minutes and there were two opportunities for breaks. In the online experiment, participants also had to confirm they were willing to submit their responses for analysis at the end of the procedure. The experiment was approved by a Faculty Ethics Committee.

Results

Preliminary analysis. Using the total proportion of correct responses, three participants were found to have performed below chance (0.50) and were removed. This left 26 participants in the final analysis. Missing/invalid responses were also removed, but these comprised less than 1% of total trials.

Filler trials. While the filler trials were primarily introduced to make the trial arrangements less predictable, standalone filler trials with a probe can be used to determine whether participants attempted to memorize the first set of targets in the trial pair. The overall proportion of correct responses across these filler trials was high ($M = 0.77$, $SD = 0.15$), suggesting there was effort to encode targets.

PI effect. The proportion of correct responses is shown in Figure 3, where Trial *N*-1 did not include a probe and Trial *N* did. PI was produced by both RN and Similar RN probes, though the effect on RN trials was strongest at the short interval. To assess this, 2 (decay interval: 3.5 s vs. 6.8 s) x 3 (probe type: RN vs. Similar RN vs. Dissimilar RN) frequentist and Bayesian repeated measures ANOVA were conducted.

There was a significant effect of probe type and extreme support for the alternative hypothesis, $F(2, 50) = 9.87$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = 0.28$, $BF_{10} = 637.14$. Šidàk and Bayesian post-hoc tests found that accuracy for Dissimilar RN probes ($M = 0.87$ [0.84, 0.90]) was higher than both the RN ($M = 0.78$ [0.75, 0.81], $p = .003$, $d = 0.72$, $BF_{10} = 338.45$) and

Similar RN (M = 0.79 [0.76, 0.82], p = .001, d = 0.85, BF_{10} = 2,010.34) conditions. The latter two conditions did not differ (p = .936, d = 0.09, BF_{10} = 0.18).

The decay interval effect was also significant, $F(1, 25) = 8.20$, MSE = 0.01, p = .008, η_p^2 = 0.25, BF_{10} = 7.32, with moderate support for the alternative hypothesis. This effect was caused by more accurate responding at the long (M = 0.84 [0.81, 0.87]) rather than the short (M = 0.79 [0.76, 0.82]) delay.

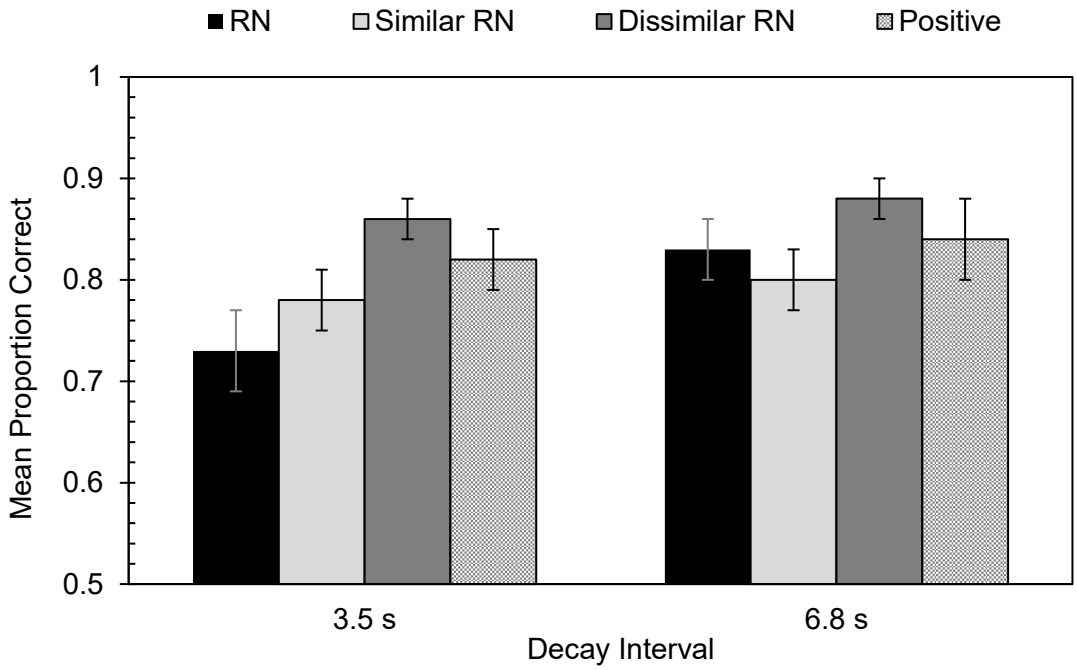
Lastly, the interaction was significant, $F(2, 50) = 5.03$, MSE = 0.01, p = .010, η_p^2 = 0.17, BF_{10} = 1.18, though insensitive from the Bayesian perspective. As the model was conventionally significant, the interaction was deemed worthy of further investigation given the specific changes predicted in the RN condition.

Holm-Šidàk corrected and (two-tailed) Bayesian paired-samples t-tests were conducted. There was no reliable change in accuracy as the decay interval was lengthened for Similar RN (t[25] = -0.83, p = .415, d = 0.20, BF_{10} = 0.28) and Dissimilar RN (t[25] = -1.20, p = .424, d = 0.15, BF_{10} = 0.40) conditions. However, in the RN condition there was a significant improvement in accuracy as the decay interval was lengthened, and strong support for the alternative hypothesis, $t(25) = -3.73$, p = .003, d = 0.66, BF_{10} = 34.67.

Positive probes. Responses to Positive probes were generally accurate and above the RN and Similar RN conditions. Paired-samples frequentist and Bayesian t-tests found no change in accuracy scores according to the decay interval, and the effect was compatible with the null hypothesis, $t(25) = -0.69$, p = 0.494, d = 0.18, BF_{10} = 0.26).

Figure 3

Mean Proportion of Correct Responses in Experiment 3, According to Probe Type and Decay Interval



Note. Error bars show +/-1 *SE*.

Discussion

Matching Experiment 1, the Similar RN condition produced a notable PI effect that was similar in magnitude to the RN condition. Yet there was also some evidence that the decay interval affected performance in the RN condition, as accurate responding in this condition increased by 10% as the decay interval was extended. Equivalent changes were not reported in the Similar RN and Dissimilar RN conditions; hence these findings are compatible with the idea that specific details are rapidly lost from VWM. This may represent minor changes to the representation, rather than a complete collapse. An imprecise representation of prior events may then endure, continuing to produce PI over time.

This diminishing effect of the RN probe does, however, contrast with McKeown et al.'s (2014) second experiment. They also introduced some trials without a probe but found the inter-trial interval length to have no effect. There are several methodological differences that may account for this, but McKeown et al.'s participants were required to remember

simpler objects and performed better on the task overall, hence it may have been more difficult to detect relatively subtle changes to the PI effect. There may be additional challenges in remembering the complex stimuli used in the present experiment, especially over delays (Mercer & Barker, 2020), leading to a rapid loss of fine details and changes to the PI effect in the RN condition. However, it should be noted that while the interaction was significant in the present experiment, Bayesian evidence for this interaction was less clear and yielded an insensitive outcome.

Experiment 4

By shortening the decay interval to just a few seconds, Experiment 3 found an initially strong and specific form of PI, produced by the RN probe, diminished over time. This was manifested in a conventionally significant interaction yet, as noted above, the Bayesian analysis was ambiguous regarding the interaction. An additional replication effort with greater power was therefore warranted. This was attempted in Experiment 4, which also further shortened the decay intervals. This decision was motivated by the need to assess very rapid changes that may be occurring in under 3.5 s, particularly as Experiment 3 suggested that performance on RN trials may be especially low at short delays.

Two of the decay intervals matched those used in Experiment 3 (3.5 s and 6.8 s), but an even shorter decay interval of just 2.3 s was added. The length of this interval was influenced by procedural considerations – a very brief delay was desired, but it was also important to avoid adversely affecting task performance by making too many changes to the configuration of the procedure. A decay period of 2.3 s was deemed suitable and it provided a better opportunity to detect rapidly occurring changes to PI than the previous three experiments.

Additionally, methodological concerns in the previous experiments were addressed. The main comparison for assessing PI in Experiments 1-3 was the Dissimilar RN probe, which was designed to have a specific relationship with RN/Similar RN probes and a target on Trial *N* and *N*-1. However, the Dissimilar RN probe was novel and not previously encoded. While this arrangement has been used previously in the recent probes task (e.g., Berman et al., 2009; McKeown et al., 2020), a fairer comparison may be a "true" NRN probe: i.e., a target previously encoded in the experiment, but more distantly than Trial *N*-1. Such a probe was added to Experiment 4, which matched an untested target from at least Trial *N*-6. The Dissimilar RN probes from Experiments 1-3 were retained to assess whether it produced any PI in comparison to NRN. As the NRN probe matched a much more distant target, it may have been largely forgotten due to the multiple intervening objects separating the original presentation of the object and its reoccurrence (see Berman et al., 2009). Conversely, the Dissimilar RN probe coarsely matched a target from the immediately preceding trial, and therefore had a closer relationship to the Trial *N*-1 target than the NRN condition did. If the PI effect generalized widely (i.e., through a coarse reminder effect), the Dissimilar RN probe should lead to lower performance than the NRN condition.

Experiments 1-3 had also adopted an arrangement in which the relationship between the targets/probes on two successive trials was carefully controlled, as the stimuli on Trials *N*-1 and *N* were both selected from the same Fribble family. This helped ensure that the requirements for RN probes could be achieved and managed the degree of overlap between successively presented stimuli, but it also made the task suitably demanding (as Fribbles across families are extremely distinct in color, a probe from a different family to the current targets was expected to be very easy to reject, as shown in the Novel condition). However, it also raised a concern that these procedural arrangements could have influenced the PI effect. Specifically, as the RN probes were selected from the same family as one of the current and

previous targets, PI may have been exaggerated due to the difficulty of the task. In Experiment 4, a separate set of uniform blue Fribbles was used, which includes three exemplars from each family/species. This meant that the requirements for the three RN probe types could still be met, while the uniform color of these stimuli made it easier to remove the requirement for targets on Trial *N*-1 and Trial *N* to be selected from the same family. In addition, the task could remain challenging and avoid ceiling effects (see Mercer & Duffy, 2015). That is, a probe could not be rejected on the basis of color alone. In sum, these changes meant the wider generalizability of PI could be assessed.

Given the results of Experiment 3, the strongest source of PI was expected in the RN condition at the shortest delay. However, if representations of events from Trial *N*-1 underwent modest but rapid changes, the PI effect should become less strong and specific as the decay interval lengthened. This should be manifested in two ways: 1) The impairment on RN trials, as compared to NRN trials, should decrease at longer decay intervals; 2) A PI effect in the Similar RN condition should only emerge at longer delays.

Method

Participants. Based on the interaction effect size from Experiment 3, 33 participants were required based on a repeated measures design with increased power (95%). However, there was an opportunity to recruit a larger sample for this experiment and boost power using the recruitment service Prolific Academic (https://www.prolific.co/). One hundred participants were paid £7.50 to complete the 45-minute experiment, though five were subsequently removed due to having 15% or more responses missing. The final sample included 95 participants (72 women, 21 men, one non-binary/third gender and one gender unreported) aged between 18 and 75 (M = 37.81, SD = 11.91, two ages unreported). No participants had taken part in the previous experiments.

Materials. This online study was again designed and run using the Gorilla Experiment Builder and could only be completed on a desktop computer or laptop. The set of 36 blue Fribbles were used to create the targets and probes, which provide three exemplars from each of the three family's four species. Each specific Fribble was used between 35 and 43 times. The use of these blue Fribbles allowed arrangements for RN, Similar RN and Dissimilar RN probes to match previous experiments. NRN probes showed an untested target that had been presented between six and nine trials earlier and arrangements for Positive probes also matched previous experiments. The end marker depicted three red hashtags (Calibri size 60) on a white background.

Design and Procedure. The 240 trials of interest were organized into pairs, where Trial *N*-1 did not have a probe and Trial *N* did. The restriction concerning the stimuli used on each trial of Experiments 1-3 was also removed. In those experiments, the relationship between stimuli on Trial *N*-1 and Trial *N* was carefully manipulated, ensuring a high level of control over the relationship between target/probes on successive trials. This was intended to reduce ceiling effects (the Novel probes showing how easy it was to reject a probe very different to either target), but it meant that the RN probe types were always closely related to the current targets. Any PI effect could have been highly specific to the experimental set up, but the blue Fribbles employed here meant there was no longer need to continue with this arrangement. Given the use of Fribbles from different families within the target pair, half of trials with a mismatching probe presented a stimulus from a different family to the current target, allowing the wider generalizability of PI to be tested.

Other arrangements matched Experiment 3, except the fixation cross was presented for 100ms, the retention interval was reduced to 400 ms and the end marker was displayed for 400 ms. There were also three inter-trial interval durations – 400 ms, 1.6 s and 4.9 s – which

included the end marker and unfilled delay prior to the new trial. This created decay intervals of 2.3 s, 3.5 s, and 6.8 s.

A 3 x 4 repeated measures design was employed, manipulating decay interval (2.3 s vs. 3.5 s vs. 6.8 s) and probe type (RN vs. Similar RN vs. Dissimilar RN vs. NRN). There were 12 trials for each probe type at each decay interval, and 32 Positive trials. There were also 30 trial pairs without probes and another 30 regular standalone trials with a probe, equally divided across the three decay intervals. These again acted as fillers and were designed to reduce expectations that a trial without a probe would always be followed by one with a probe.

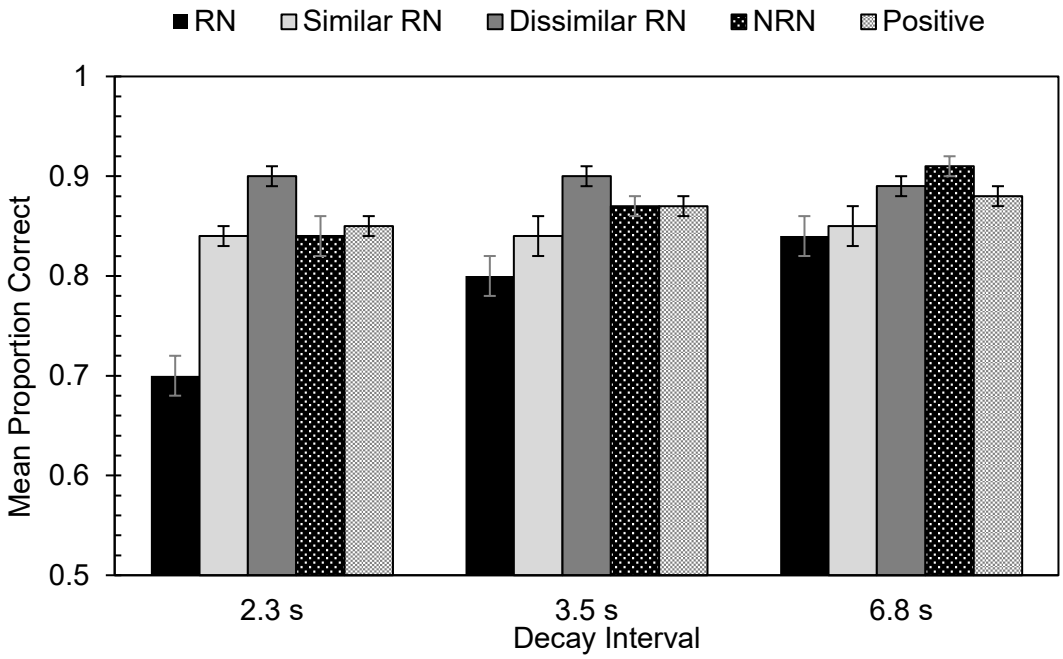
The experiment was completed in three blocks of 100 trials, where trials within a block followed a set pattern but the order of blocks was randomized, and the first block was preceded by 12 practice trials. The experiment was approved by a Faculty Ethics Committee.

Results

Preliminary analysis and filler trials. Based on the total proportion of correct responses, two participants were found to have performed below chance overall and over 3 *SD*s below the mean. These participants were removed, leaving 93 participants in the data set. Missing/invalid responses, comprising less than 1.25% of total trials, were again removed. Following Experiment 3, performance on standalone filler trials was used to determine whether participants attempted to memorize the first set of targets in the trial pair. The overall proportion of correct responses across these filler trials was high ($M = 0.82$, $SD = 0.11$), again suggesting there was effort to encode all targets.

Figure 4

Mean Proportion of Correct Responses in Experiment 4, According to Probe Type and Decay Interval



Note. Error bars show +/-1 *SE*.

PI effect. The proportion of correct responses are shown in Figure 4, showing data where Trial *N*-1 did not include a probe. In comparison with the NRN control, RN probes were particularly disruptive to performance at the shortest decay interval, but this diminished as time passed. Additionally, in comparison to NRN, evidence for PI in the Similar RN condition began to emerge at the intermediate decay interval.

Next, 3 (decay interval: 2.3 s vs. 3.5 s vs. 6.8 s) x 4 (probe type: RN vs. Similar RN vs. Dissimilar RN vs. NRN) frequentist and Bayesian repeated measures ANOVA were conducted. There was a significant effect of probe type, $F(2.60, 239.50) = 56.03$, $MSE = 0.01$, $p < .001$ (Greenhouse-Geisser corrected), $\eta_p^2 = 0.38$, $BF_{10} = 2.22 \times 10^{33}$, with extreme support for the alternative hypothesis. Šidàk and Bayesian post-hoc tests found that performance on RN trials ($M = 0.78$ [0.77, 0.79]) was worse than all other probe types

(Similar RN: $p < .001$, $d = 0.46$, $BF_{10} = 7.26 \times 10^6$; Dissimilar RN: $p < .001$, $d = 0.95$, $BF_{10} = 2.28 \times 10^{23}$; NRN: $p < .001$, $d = 0.72$, $BF_{10} = 6.14 \times 10^{17}$). Similar RN probes ($M = 0.84$ [0.83, 0.85]) also led to lower accuracy than Dissimilar RN ($M = 0.90$ [0.89, 0.91], $p < .001$, $d = 0.44$, $BF_{10} = 9.73 \times 10^6$) and NRN ($M = 0.87$ [0.86, 0.88], $p = .015$, $d = 0.23$, $BF_{10} = 20.26$) trials, whereas performance on Dissimilar RN trials was higher than NRN trials ($p = .001$, $d = -0.22$, $BF_{10} = 26.19$).

There was a significant effect of the decay interval too, $F(2, 184) = 25.27$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = 0.22$, $BF_{10} = 2.91 \times 10^6$, with extreme support for the alternative hypothesis. Šidàk and Bayesian post-hoc tests showed an improvement in accurate responding as the decay interval lengthened (2.3 s vs. 3.5 s: $p < .001$, $d = 0.26$, $BF_{10} = 331.90$; 2.3 s vs. 6.8 s: $p < .001$, $d = 0.44$, $BF_{10} = 8.03 \times 10^7$; 3.5 s vs. 6.8 s: $p = .012$, $d = 0.19$, $BF_{10} = 6.43$), with worst performance at the very shortest delay.

Finally, the interaction was significant and this time there was extreme support for the alternative hypothesis, $F(6, 552) = 12.51$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = 0.12$, $BF_{10} = 1.33 \times 10^9$. To explore the interaction, a series of Holm-Šidàk adjusted frequentist and Bayesian paired-samples *t*-tests were used to compare probe types at each decay interval.

At the 2.3 s decay interval, the RN probe had lower accuracy than all other probe types (RN vs. NRN: $p < .001$, $d = 0.82$, $BF_{10} = 1.43 \times 10^{11}$; RN vs. Similar RN: $p < .001$, $d = 0.84$, $BF_{10} = 5.27 \times 10^{10}$; RN vs. Dissimilar RN: $p < .001$, $d = 1.31$, $BF_{10} = 1.32 \times 10^{18}$). However, the Dissimilar RN condition had higher accuracy than both Similar RN ($p < .001$, $d = -0.50$, $BF_{10} = 30{,}067.91$) and NRN ($p < .001$, $d = -0.46$, $BF_{10} = 11{,}113.63$) probe types, whereas Similar RN and NRN did not differ ($p = .944$, $d = 0.01$, $BF_{10} = 0.12$).

At the 3.5 s decay interval, performance in the RN condition was again lower than NRN ($p < .001$, $d = 0.48$, $BF_{10} = 236.81$) and Dissimilar RN ($p < .001$, $d = 0.70$, $BF_{10} = 405{,}480.35$) probes. The RN vs. Similar RN comparison was on the threshold of significance

(p = .049) but represented a small effect (d = 0.22) and was insensitive (BF_{10} = 1.35). The Similar RN condition also led to lower accuracy than Dissimilar RN (p = .001, d = 0.43, BF_{10} = 126.84), but the comparison with the NRN condition was more ambiguous, being on the threshold of significance and insensitive (p = .050, d = 0.22, BF_{10} = 0.75). Correct responses for Dissimilar RN probes exceeded NRN probes (p = .033, d = -0.24), but there was only anecdotal evidence for the alternative hypothesis (BF_{10} = 2.70).

At the 6.8 s decay interval, responding on NRN trials was higher than both RN (p < .001, d = 0.49, BF_{10} = 10,959.42) and Similar RN conditions (p = .002, d = 0.38, BF_{10} = 44.11), whereas the latter two probe types did not differ (p = .446, d = 0.08, BF_{10} = 0.15). Accuracy was also lower for RN than Dissimilar RN probes (p = .005, d = 0.33, BF_{10} = 20.47), but the other comparisons were non-significant and insensitive (Similar RN vs. Dissimilar RN: p = .097, d = 0.23, BF_{10} = 1.04; NRN vs. Dissimilar RN: p = .129, d = 0.16, BF_{10} = 0.59).

Lastly, four Holm-Šidàk corrected frequentist and Bayesian one-way ANOVAs assessed changes within the decay intervals. There was no effect of decay interval for Similar RN probes, $F(1.85, 170.14) = 0.99$, MSE = 0.01, p = .603 (Greenhouse-Geisser corrected), η_p^2 = 0.01, BF_{10} = 0.09, and Dissimilar RN probes, $F(2, 184) = 0.88$, MSE = 0.01, p = .418, η_p^2 = 0.01, BF_{10} = 0.08. There was a change in the RN condition, $F(2, 184) = 37.95$, MSE = 0.01, p < .001, η_p^2 = 0.29, BF_{10} = 4.10 x 10^{11}, with Holm-Šidàk corrected and Bayesian t-tests showing that accuracy on RN trials improved as the decay interval was lengthened (2.3 s vs. 3.5 s: $t[92] = 6.00$, p < .001, d = 0.57, BF_{10} = 306,741.33; 2.3 s vs. 6.8 s: $t[92] = 8.13$, p < .001, d = 0.86, BF_{10} = 4.27 x10^9; 3.5 s vs. 6.8 s: $t[92] = 2.71$, p < .001, d = 0.28, BF_{10} = 3.53). Responding to NRN probes also became more accurate as the decay interval was lengthened, $F(1.66, 152.36) = 14.42$, MSE = 0.01, p < .001 (Greenhouse-Geisser corrected), η_p^2 = 0.14, BF_{10} = 9,493.45. Holm-Šidàk corrected and Bayesian t-tests showed an increase in correct

responses from 2.3 s to 6.8 s (t[92] = 5.16, $p < .001$, $d = 0.49$, $BF_{10} = 10{,}028.88$) and 3.5 s to 6.8 s (t[92] = 3.92, $p < .001$, $d = 0.30$, $BF_{10} = 113.73$). There was also a significant improvement from 2.3 s to 3.5 s (t[92] = 2.07, $p = .041$, $d = 0.22$), though the Bayesian analysis yielded an insensitive outcome ($BF_{10} = 0.88$)[4].

Positive probes. Responses to Positive probes were assessed using one-way ANOVA. There was a significant effect of decay interval, $F(2, 184) = 4.84$, $MSE = 0.003$, $p = .009$, $\eta_p^2 = 0.05$, and this was 2.8 times more likely under the alternative, than null, hypothesis ($BF_{10} = 2.82$; thus, while the evidence leaned more towards the alternative hypothesis, it was insensitive). Šidàk and Bayesian post-hoc tests showed an improvement from the shortest 2.3 s decay interval to the longest 6.8 s interval ($p = .012$, $d = 0.26$, $BF_{10} = 6.64$), but no other differences were found (2.3 s vs. 3.5 s: $p = .135$, $d = 0.18$, $BF_{10} = 0.79$; 3.5 s vs. 6.8 s: $p = .737$; $d = 0.07$, $BF_{10} = 0.17$).

Discussion

The final experiment further reduced the length of the shortest decay interval while making methodological improvements. At the very shortest decay interval, only the RN probe reliably lowered accuracy below NRN, indicating that a few seconds after encoding, the residual representation of a Trial *N*-1 target retained specific and fine details and produced strong PI. Similar effects were found at the intermediate decay interval, but at the longest decay interval, PI was also produced from the Similar RN probe. This would be expected if the old representation had undergone distortions over the passage of time, leading

[4] While there was too little data in the laboratory-based condition to determine whether the experimental set up (online vs. offline) affected performance, descriptive statistics indicated similar trends. Both RN and Similar RN conditions lowered task accuracy compared to Dissimilar RN. Both online and offline versions of the experiment also found a recovery in performance in the RN condition as the decay interval was lengthened, whereas this was much less pronounced in the Similar RN condition.

to a less specific form of PI. Supporting this view, PI at the longest interval was much less potent than at the shortest interval in the RN condition. Correct responding increased by 14% over the delay, suggesting an initially strong and specific form of PI became weaker and coarser over time, perhaps as the underlying memory was gradually forgotten.

Another aspect of the present data worthy of comment concerns the NRN condition (where probes displayed targets not encountered for at least six trials), which led to worse performance than the Dissimilar RN condition (a novel probe not previously experienced). Data from NRN trials may suggest a long-lasting carryover effect across trials, albeit this is less disruptive than trials on which the probe was seen on the immediately preceding trial. Both Hartshorne (2008) and Mercer and Fisher (2022) have shown that probes matching a target from Trial *N*-3 disrupts VWM performance, highlighting that PI can endure over trials, though Hartshorne (2008) found this effect diminished after 4-5 trials. Alternatively, improved performance on Dissimilar RN trials may reflect a heightened ability to detect novelty – i.e., noticing stimuli never previously encountered (see also the high performance on Novel trials in Experiments 1 and 2). Of most importance, however, both RN and Similar RN trials led to worse overall performance than NRN and Dissimilar RN trials.

Table 2

Proportion of Errors According to Probe Type and Decay Interval in Experiments 1-4

	Exp 1		Exp 2		Exp 3		Exp 4		
Probe Type	4.8 s	12.3 s	4.6 s	12.4 s	3.5 s	6.8 s	2.3 s	3.5 s	6.8 s
RN	0.19 (0.14)	0.14 (0.16)	0.30 (0.18)	0.22 (0.16)	0.27 (0.22)	0.17 (0.15)	0.30 (0.17)	0.20 (0.16)	0.16 (0.14)
Similar RN	0.18 (0.13)	0.16 (0.13)	0.19 (0.17)	0.17 (0.12)	0.22 (0.16)	0.20 (0.18)	0.16 (0.14)	0.16 (0.17)	0.15 (0.16)
Dissimilar RN	0.11 (0.13)	0.12 (0.14)	0.12 (0.13)	0.11 (0.11)	0.14 (0.12)	0.12 (0.11)	0.10 (0.12)	0.10 (0.13)	0.11 (0.14)
NRN	-	-	-	-	-	-	0.16 (0.15)	0.13 (0.13)	0.09 (0.12)
Novel	0.02 (0.04)	0.01 (0.04)	0.03 (0.06)	0.03 (0.06)	-	-	-	-	-
Positive	0.12 (0.09)	0.11 (0.09)	0.11 (0.08)	0.25 (0.16)	0.18 (0.14)	0.16 (0.18)	0.15 (0.10)	0.13 (0.10)	0.12 (0.09)

General Discussion

There are competing notions about item-specific PI in VWM, which has important theoretical implications for understanding short-term forgetting processes. One possibility is that PI arises from an immediate carryover of redundant residual information (see Makovski & Jiang, 2008; Shoval & Makovski, 2022). These representations may be susceptible to rapid loss (e.g., via a decay process) and should therefore have a limited lifetime. Alternatively, PI may be produced from robust, passive representations that have qualities closer to long-term memory (McKeown et al., 2014), meaning PI should be insensitive to the passage of time. Findings that PI can endure over lengthy delays (e.g., Hartshorne, 2008; McKeown et al., 2014, 2020; Mercer et al., 2022) appears to support these latter accounts, yet it is likely that more subtle, fast-acting changes to the representations underpinning PI were undetected in previous studies.

The present study explored this issue by modifying the recent probes task. Three different types of RN probe were employed, and these increasingly varied in their similarity to a Trial *N*-1 target. Matching other recent probe task studies (e.g., Hartshorne, 2008; Makovski & Jiang, 2008, McKeown et al., 2014, 2020; Mercer & Duffy, 2015; Mercer et al., 2022), there was clear evidence for PI. This effect was driven by representations from the immediately preceding trial, supporting Shoval and Makovski (2022), and PI was robust in all four experiments. Further evidence for robust PI is derived from comparisons between RN and Positive trials. While this was not the primary emphasis of the present study, such a comparison can provide a measure of how much similar items interfere at retrieval. Previous studies have found accurate responding on RN trials to be generally higher than Positive trials (e.g., McKeown et al., 2022; Mercer et al., 2022; Mercer & Fisher, 2022). Conversely,

the present studies found responses on RN trials were usually lower than Positive trials, especially at very short delays[5].

Of most importance was changes to the specificity of PI over time. Data from the first two experiments showed that PI can be produced by items that both match or closely resemble those encountered in the recent past, and this was unaffected by the passage of time. However, when the decay intervals were shortened in Experiments 3 and 4, a specific and strong form of PI emerged on RN trials, but this declined at longer delays. In Experiment 3, the RN probe was most detrimental at a 3.5 s decay interval, with performance in this condition improving once the decay interval was lengthened. While there was some ambiguity in the interaction underpinning this effect, the improved and highly powered Experiment 4 reported very robust and specific PI at the 2.3 s decay interval, where performance on RN trials was substantially worse than all other conditions. A similar effect was found at the 3.5 s decay interval (matching trends seen in Experiment 3), but by 6.8 s the PI effect was more akin to that reported in Experiments 1 and 2, where both RN and Similar RN probes caused PI. In summary, a strong and specific PI effect appeared at very brief delays, but it became milder and more general at longer delays.

These results are compatible with the notion that representations underpinning PI rapidly lose precise details, but then endure in a broadly accurate state thereafter. As such, these data show that both the temporal interpretation of PI (e.g., in which decay is dominant) and the time-insensitive, passive conception of PI cannot fully explain these data, and a third interpretation which adopts aspects of both these accounts may be best.

---

[5] Separate ANOVAs were used to assess the effect of probe type (RN vs. Positive) and decay intervals in all experiments. While the effect of probe type and the interaction were non-significant in Experiment 1, Experiment 2 did find an interaction. This was due to lower performance on RN than Positive trials at the short, but not long, retention interval. Experiments 3 and 4 also found similar interactions: RN trials led to lower performance at the shorter decay intervals. In Experiment 4, there was also a main effect of probe type, due to lower overall responding on RN than Positive trials.

Evidence for this interpretation comes from three key findings. Firstly, there was a decline in stimulus-specific PI on RN probes over time. In Experiment 4, for example, performance in the RN condition was 14% lower than the NRN condition at the very shortest delay, but this declined to a 7% difference at the very longest delay. Secondly, performance in the Similar RN condition was largely immune to the passage of time, but this probe type was less detrimental than the RN probe at very short delays and only convincingly impaired performance in relation to NRN probes at the longest decay interval. Thirdly, PI did not continue to generalize to other stimuli at long delays, as overall performance in the Dissimilar RN condition was consistently better than RN and Similar RN conditions, and even better than the NRN condition in Experiment 4.

Important theoretical implications arise from these findings. In McKeown et al.'s (2014) account of PI, passively maintained memories do not degrade over time. Zhang et al. (2022) have also proposed a distinction between actively and passively retained memories, with the latter believed to be resistant to forgetting processes. The present data did show that after a certain timepoint, the memory producing PI did not seem to undergo further major changes (at least at the delays tested – it is possible that after a much longer interval, there will be a full recovery from PI, as seen in some verbal memory studies (e.g., Kincaid & Wickens, 1970; Loess & Waugh, 1967; Peterson & Gentile, 1965; Turvey et al., 1970). However, in the first several seconds after encoding, a strong and particularly disruptive form of PI operates, which rapidly declines. It may reflect a short-lived "carryover" of the representation that had until very recently been actively maintained, potentially occurring due to a delay in transferring the contents of VWM from an active to passive store (see Zhang et al., 2022). Overall, the notion that passively maintained residual representations are entirely unaffected by the passage of time was challenged by these data.

However, as that representation quickly loses fidelity, PI takes the form reported by McKeown et al. (2014, 2020), which is long-lasting but milder. This may reflect the existence of a passively maintained memory, as it is unlikely to be subjected to active retention strategies, and it may have been responsible for prior demonstrations of durable PI (e.g., Hartshorne, 2008; Mercer & Fisher, 2022). It is also consistent with Oberauer et al.'s (2017) account, where PI is caused by a familiarity signal from LTM that competes with working memory during the retrieval stage. Nonetheless, the PI effect is not immune to the passage of time, at least at very short intervals.

As such, these data also have implications for the understanding of short-term forgetting, and specifically a temporal decay process. While there was little reduction in PI at decay intervals exceeding 6 s, temporal forgetting operated over a shorter timescale (especially between the 2.3 s and 6.8 s intervals used in Experiment 4). Using verbal material, Berman et al. (2009) found that the inter-trial interval duration had little impact on PI in the recent probes task, but the shortest decay interval tested was 3.3 s. When Campoy (2012) shortened these decay intervals, a rapid reduction in verbal PI was discovered. The present studies, using unfamiliar visual stimuli, are broadly consistent with these findings. That is, PI is only affected by the passage of time at very brief delays – an effect congruent with wider ideas that information held in working memory experiences strongest loss several seconds after encoding (see Mueller & Krawitz, 2009; Ricker & Cowan, 2010).

It is therefore possible that a rapid decay process operates in VWM, leading to the loss of specific details. This is compatible with data from Experiment 2, where the ability to detect target-probe matches on Positive trials declined strongly over delays lasting 4.5 s (Experiment 2). Nonetheless, as the PI effect did not disappear completely, representations underpinning PI do not appear to be entirely lost and instead may undergo more modest changes. These data are therefore not fully consistent with contemporary decay theories such

as the TBRS model. That account should expect a continued recovery from PI over time, as passive representations are not maintained via attention and experience greater time-dependent decay. If there is a decay process, these data suggest it is fast acting, incomplete and limited.

Another possible explanation is offered by temporal distinctiveness theories such as SIMPLE (e.g., Brown et al., 2007). This account can explain a reduction in PI over time as events from Trial *N*-1 can be more easily differentiated from those on Trial *N*. Temporal distinctiveness theory is also supported by data from Experiments 3 and 4, where task performance was generally higher when longer delays separated trials (for example, there were general benefits of trial spacing in Experiment 4, where overall performance improved at each decay interval). Other findings in the VWM literature suggest temporal distinctiveness better explains forgetting than a decay process (e.g., Mercer, 2014; Souza & Oberauer, 2015), though this is not always the case (Ricker et al., 2014). Furthermore, temporal distinctiveness theory seems to hypothesize a continued benefit of longer decay intervals, whereas the present data suggest the recovery from PI is confined to a limited time window.

More generally, the present study highlights the role of temporal factors in short-term forgetting (see also Mathias et al., 2021; Ricker et al., 2020). In relation to the wider debate between time-based and interference/event-based theorists, these data offer some support to the former and suggest that the loss of VWM may not be entirely attributable to interference. Despite this, the present data also indicate that temporal factors do not affect VWM in all circumstances. Responding on Novel probes was unaffected by the decay interval in both Experiments 1 and 2, and long-lasting PI suggests that representations previously held within VWM can persist beyond what might be expected by decay and temporal distinctiveness theories. Indeed, the degraded representation of information previously held in VWM may

last for tens of seconds (McKeown et al., 2020, Experiment 3), if not longer. These data therefore highlight a more complex effect of time on VWM than allowed by decay and temporal distinctiveness theories, though the full experimental outcomes may be most compatible with a rapid decay process that removes fine details from the representation, while leaving a less detailed residual trace intact.

While the present study offers some novel findings, there are potential limitations, and one concern relates to the stimuli used in the task. All experiments used complex three-dimensional shapes that may be difficult to retain, so participants may have been unable to differentiate RN from Similar RN probes. This was not anticipated to be an issue as participants can discriminate Fribble stimuli when differences between them are smaller than those used here, at least at short delays (Mercer, 2014, 2018; Mercer & Barker, 2020). Data from Experiments 2 and 4 then confirmed this, finding that RN probes led to worse performance than Similar RN probes. This provides convincing evidence that they can be differentiated.

A further concern is the loss in experimental control in Experiments 3 and 4, which were conducted online during the Covid-19 pandemic. The environment in which the procedure was conducted was uncontrolled, and participants will have had different set-ups while doing the task. Nonetheless, the PI effect obtained from these studies was robust and the online nature of Experiment 4 allowed for an improvement in statistical power.

A third potential issue was the age range of participants, which was quite wide in all experiments. This was most pronounced in Experiment 4, where the mean age of the sample was notably higher than Experiments 1-3 and there was a greater spread of ages (the oldest participant was 75). However, two-tailed frequentist and Bayesian correlations revealed no reliable relationship between age and the PI effect in Experiment 4, as determined by subtracting RN performance from NRN performance, $r(89) = 0.08$, $p = .442$, $BF_{10} = 0.18$.

There was also no correlation between age and overall task performance, $r(89) = -0.02$, $p = .888$, $BF_{10} = 0.13$, and both these correlations were more compatible with the null hypothesis. This does appear to contradict previous studies, where across different memory paradigms older adults are more susceptible to PI than younger adults (e.g., Carretti et al., 2012; Hasher et al., 2002; Rhodes et al., 2022). In past studies, however, participants were categorized into groups based on their age, with each group being of approximately similar size. Conversely, in the present study older adults were less common. Across all experiments, only 10 adults were aged 55 or over, and only one was aged over 65, which may explain why age appeared to have a limited effect in Experiment 4.

The present study also used a simple match/mismatch option when the probe appeared. This was chosen as it is the standard response in the recent probes task and the intention of these experiments was to make modifications to a proven PI measure. However, future studies could gain further insights into old, residual representations through employing confidence ratings alongside the accuracy measure (e.g., Shoval & Makovski, 2022) or converting the task into a two-alternative forced-choice procedure. Such an approach could avoid an issue with the RN trials of the recent probes task – it is not possible to determine whether the increase in errors on RN trials is due to the simple repetition of an item or by the similarity of the target and RN probe. While the present study mitigated this problem by employing additional RN probes that varied in their similarity to the Trial *N*-1 target, alternative measures for assessing memory would be useful.

Related to this point, further modifications to the recent probes task could allow a more direct comparison between redundant, residual representations that are passively maintained with those that are being actively retained. Future studies would also benefit from exploring individual differences in PI. In the first three experiments, the average PI effect – based on subtracting the proportion of correct responses on RN trials from Dissimilar RN

trials – was 0.10. However, this varied between -0.14 and 0.46. When comparing RN and NRN trials in Experiment 4, the average PI effect was 0.09, but varied from -0.10 to 0.44. Some participants were therefore able to resist PI, hence understanding how they achieved this, and strategies that they may have used, would be worthwhile.

In conclusion, the present study suggested item-specific PI is sensitive to the passage of time, challenging previous findings. Instead, the present results were consistent with the notion that representations producing PI undergo some degradation, at least within a brief window. An initially specific and particularly disruptive form of PI operates for several seconds after encoding, but thereafter the PI effect becomes milder and less specific. This is compatible with a fast-acting but partial and limited decay process.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, *36*(1), 74–81. https://doi.org/10.3758/MC.36.1.74

Atkins, A. S., Berman, M. G., Reuter-Lorenz, P. A., Lewis, R. L., & Jonides, J. (2011). Resolving semantic and proactive interference in memory over the short-term. *Memory & Cognition*, *39*(5), 806–817. https://doi.org/10.3758/s13421-011-0072-5

Balaban, H., Fukuda, K., & Luria, R. (2019). What can half a million change detection trials tell us about visual working memory? *Cognition, 191*, 103984. https://doi.org/10.1016/j.cognition.2019.05.021

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., & Camos, V. (2021). The time-based resource-sharing model of working memory. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: State of the science* (pp. 85–115). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0004

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 317–333. https://doi.org/10.1037/a0014873

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Campoy, G. (2012). Evidence for decay in verbal short-term memory: A commentary on Berman, Jonides, and Lewis (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 1129–1136. https://doi.org/10.1037/a0026934

Carretti, B., Mammarella, I. C., & Borella, E. (2012). Age differences in proactive interference in verbal and visuospatial working memory. *Journal of Cognitive Psychology*, *24*(3), 243–255. https://doi.org/10.1080/20445911.2011.603695

Endress, A.D. (2022). Memory and proactive interference for spatially distributed items. *Memory & Cognition, 50*(4), 782–816. https://doi.org/10.3758/s13421-021-01239-1

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143*(2), 548–565. https://doi.org/10.1037/a0033934

Endress, A. D., & Siddique, A. (2016). The cost of proactive interference is constant across presentation conditions. *Acta Psychologica, 170*, 186–194. https://doi.org/10.1016/j.actpsy.2016.08.001

Farrell, S., Oberauer, K., Greaves, M., Pasiecznik, K., Lewandowsky, S., & Jarrold, C. (2016). A test of interference versus decay in working memory: Varying distraction within lists in a complex span task. *Journal of Memory and Language, 90*, 66–87. https://doi.org/10.1016/j.jml.2016.03.010

Gold, J. M., Murray, R. F., Sekuler, A. B., Bennett, P. J., & Sekuler, R. (2005). Visual memory decay is deterministic. *Psychological Science, 16*(10), 769–774. https://doi.org/10.1111/j.1467-9280.2005.01612.x

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS ONE, 3*(7), e2716. https://doi.org/10.1371/journal.pone.0002716

Hasher, L., Chung, C., May, C. P., & Foong, N. (2002). Age, time of testing, and proactive interference. *Canadian Journal of Experimental Psychology*, *56*(3), 200–207. https://doi.org/10.1037/h0087397

Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology, 62*(2), 124–138. https://doi.org/10.1037/a0014164

JASP Team (2018). JASP (Version 0.9.0.1) [Computer software]. Retrieved from https://jasp-stats.org/

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience, 139*, 181–193. https://doi.org/10.1016/j.neuroscience.2005.06.042

Kincaid, J. P., & Wickens, D. D. (1970). Temporal gradient of release from proactive inhibition. *Journal of Experimental Psychology, 86*(2), 313–316. https://doi.org/10.1037/h0029991

Krill, D., Avidan, G., & Pertzov, Y. (2018). The rapid forgetting of faces. *Frontiers in Psychology, 9*, 1319. https://doi.org/10.3389/fpsyg.2018.01319

Kuuramo, C., Saarinen, J., & Kurki, I. (2022). Forgetting in visual working memory: Internal noise explains decay of feature representations. *Journal of Vision*, *22*(8), 8. https://doi.org/10.1167/jov.22.8.8

Lin, P.-H., & Luck, S. J. (2012). Proactive interference does not meaningfully distort visual working memory capacity estimates in the canonical change detection task. *Frontiers in Psychology, 3*, 42. https://doi.org/10.3389/fpsyg.2012.00042

Loess, H., & Waugh, N. C. (1967). Short-term memory and intertrial interval. *Journal of Verbal Learning and Verbal Behavior, 6*, 455–460. https://doi.org/10.1016/S0022-5371(67)80001-X

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281. https://doi.org/10.1038/36846

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory & Cognition*, *36*(1), 43–52. https://doi.org/10.3758/MC.36.1.43

Mathias, S. R., Varghese, L., Micheyl, C., & Shinn-Cunningham, B. G. (2021). Gradual decay and sudden death of short-term memory for pitch. *Journal of the Acoustical Society of America, 149*, 259. https://doi.org/10.1121/10.0002992

McKeown, D., Holt, J., Delvenne, J. F., Smith, A., & Griffiths, B. (2014). Active versus passive maintenance of visual nonverbal memory. *Psychonomic Bulletin & Review, 21*(4), 1041–1047. https://doi.org/10.3758/s13423-013-0574-1

McKeown, D., Mercer, T., Bugajska, K., Duffy, P., & Barker, E. (2020). The visual non-verbal memory trace is fragile when actively maintained but endures passively for tens of seconds. *Memory & Cognition, 48*(2), 212–225. https://doi.org/10.3758/s13421-019-01003-6

Mercer, T. (2014). The loss of short-term visual representations over time: Decay or temporal distinctiveness? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2281–2288. http://dx.doi.org/10.1037/a0038141

Mercer, T. (2018). Retroactive interference in visual short-term memory. *Experimental Psychology, 65*(3), 139–148. https://doi.org/10.1027/1618-3169/a000401

Mercer, T., & Markova, A.-M. (2023, February 16). *Rapid but incomplete degradation of residual visual representations over time*. https://doi.org/10.17605/OSF.IO/4JBSV

Mercer, T., & Barker, E. (2020). Time-dependent forgetting in visual short-term memory. *Journal of Cognitive Psychology, 32*(4), 391–408. https://doi.org/10.1080/20445911.2020.1767627

Mercer, T., & Duffy, P. (2015). The loss of residual visual memories over the passage of time. *Quarterly Journal of Experimental Psychology, 68*(2), 242–248. https://doi.org/10.1080/17470218.2014.975256

Mercer, T., & Fisher, L. P. (2022). Magnitude and sources of proactive interference in visual memory. *Memory, 30*(5), 591–609. https://doi.org/10.1080/09658211.2022.2032179

Mercer, T., Jarvis, R.-J., Lawton, R., & Walters, F. (2022). Tracking proactive interference in visual memory. *Frontiers in Psychology, 13*, 896866. https://doi.org/10.3389/fpsyg.2022.896866

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*(4), 465–501. https://doi.org/10.1016/0010-0285(78)90008-7

Mueller, S. T., & Krawitz, A. (2009). Reconsidering the two-second decay hypothesis in verbal working memory. *Journal of Mathematical Psychology*, *53*(1), 14–25. https://doi.org/10.1016/j.jmp.2008.11.001

Nilsson, T. (2020). What came out of visual memory: Inferences from decay of difference-thresholds. *Attention, Perception, & Psychophysics, 82*, 2963–2984. https://doi.org/10.3758/s13414-020-02032-z

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(1), 1–22. https://doi.org/10.1037/xlm0000302

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1224–1231. https://doi.org/10.1037/a0030947

Peterson, L. R., & Gentile, A. (1965). Proactive interference as a function of time between tests. *Journal of Experimental Psychology, 70*(5), 473–478. https://doi.org/10.1037/h0022542

Postle, B. R., & Brush, L. N. (2004). The neural bases of the effects of item-nonspecific proactive interference in working memory. *Cognitive, Affective, & Behavioral Neuroscience, 4*(3), 379–392. https://doi.org/10.3758/CABN.4.3.379

Postle, B. R., Brush, L. N., & Nick, A. M. (2004). Prefrontal cortex and the mediation of proactive interference in working memory. *Cognitive, Affective, & Behavioral Neuroscience, 4*(4), 600–608. https://doi/org/10.3758/cabn.4.4.600

Rademaker, R. L., Park, Y. E., Sack, A. T., & Tong, F. (2018). Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 44*(6), 925–940. https://doi.org/10.1037/xhp0000491

Ricker, T. J., & Cowan, N. (2010). Loss of visual working memory within seconds: The combined use of refreshable and non-refreshable features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1355–1368. https://doi.org/10.1037/a0020356

Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 417–428. https://doi.org/10.1037/a0034301

Ricker, T. J., Sandry, J., Vergauwe, E., & Cowan, N. (2020). Do familiar memory items decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(1), 60–76. https://doi.org/10.1037/xlm0000719

Ricker, T. J., Spiegel, L. R., & Cowan, N. (2014). Time-based loss in visual short-term memory is from trace decay, not temporal distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1510–1523. https://doi.org/10.1037/xlm0000018

Rhodes, S., Buchsbaum, B. R. & Hasher, L. (2022). The influence of long-term memory on working memory: Age-differences in proactive facilitation and interference. *Psychonomic Bulletin & Review 29*, 191–202. https://doi.org/10.3758/s13423-021-01981-2

Shoval, R., Luria, R., & Makovski, T. (2020). Bridging the gap between visual temporary memory and working memory: The role of stimuli distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(7), 1258–1269. https://doi.org/10.1037/xlm0000778

Shoval, R., & Makovski, T. (2021). The locus of proactive interference in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 47*(5), 704–715. https://doi.org/10.1037/xhp0000906

Shoval, R., & Makovski, T. (2022). Meaningful stimuli inflate the role of proactive interference in visual working memory. *Memory & Cognition*, *50*(6), 1157–1168. https://doi.org/10.3758/s13421-022-01338-7

Souza, A. S., & Oberauer, K. (2015). Time-based forgetting in visual working memory reflects temporal distinctiveness, not decay. *Psychonomic Bulletin & Review*, *22*(1), 156–162. https://doi.org/10.3758/s13423-014-0652-z

Turvey, M. T., Brick, P., & Osborn, J. (1970). Proactive interference in short-term memory as a function of prior-item retention interval. *Quarterly Journal of Experimental Psychology, 22*(2), 142–147. https://doi.org/10.1080/00335557043000078

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning & Verbal Behavior, 2*(5-6), 440–445. https://doi.org/10.1016/S0022-5371(63)80045-6

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20*(4), 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x

Zhang, J., Ye, C., Sun, H. J., Zhou, J., Liang, T., Li, Y., & Liu, Q. (2022). The passive state: A protective mechanism for information in working memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(9), 1235-1248. https://doi.org/10.1037/xlm0001092