AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

2024, Vol. 60, No. 3, 582–594
https://doi.org/10.1037/dev0001686

Can't Get It Out of My Head: Proactive Interference in the Visual Working Memory of 3- to 8-Year-Old Children

Mollie Hamilton, Tessyia Roper, Erik Blaser, and Zsuzsa Kaldy
Department of Psychology, University of Massachusetts Boston

Proactive interference (PI) occurs when previously learned memories compete with currently relevant information. Despite extensive literature investigating the effect in adults, little work has been done in young children. In three preregistered studies ($N = 38$, 35, 172; convenience samples from the Northeastern United States), first, we showed that 3-year-old toddlers are highly sensitive to the effect of PI in visual working memory and second, that these effects can originate from the reactivation of previously encoded information. Third, we tested how the ability to cope with PI changes between 2.5 and 7.5 years of age. Besides providing an estimate for the size of the interference effect at the youngest age to date, our findings have an important methodological implication: paradigms that repeat items across trials potentially underestimate young children's working memory abilities.

Public Significance Statement

Performing tasks requires the constant updating of working memory (WM), which is especially challenging when previously relevant memories interfere with current ones ("Did I add a teaspoon of salt already, or was that the baking powder?"). In adults, this proactive interference (PI) has been well established as a fundamental limitation of memory performance. Characterizing PI in early childhood has three potential impacts. First, it can inform theoretical models of WM and its development. Second, it can lead to applied insights, such as a better understanding of children's learning in the classroom. Lastly, it has important methodological impacts. WM capacity is typically measured by presenting a series of trials with similar, or repeated, stimuli—a practice that promotes PI and may lower performance. Without updating paradigms to account for, or avoid, this interference, we may be systematically underestimating children's WM capacity.

Keywords: working memory, proactive interference, capacity, children, cognitive control

Supplemental materials: https://doi.org/10.1037/dev0001686.supp

Working memory (WM) is the cognitive mechanism responsible for keeping information active in order to solve a task (Baddeley, 1992), from baking a cake to solving math problems. The updating and manipulation of relevant information require inhibiting information that is no longer relevant. In the cognitive neuroscience literature, such interference resolution is discussed as an example of a crucial cognitive control skill (Badre & Wagner, 2007; Bunge & Wright, 2007; O'Reilly & Frank, 2006). In contrast with long-term memory (LTM), WM is very limited; adults are only able to maintain three to four items in mind before they start to slow down or make errors (Cowan, 2001). What gives rise to these WM capacity limitations? Unlike temporal decay theory where information is forgotten as a

Erik Blaser https://orcid.org/0000-0002-9968-8946
Zsuzsa Kaldy https://orcid.org/0000-0002-0679-4505

This research was supported by the National Institutes of Health (Grant R15HD086658). The authors would like to thank Martha Ann Bell for helpful suggestions; Sangya Dhungana and other members of the Early Minds Lab at UMass Boston for their help with data collection; the staff of the Children's Museum of New Hampshire, especially Xanthi Gray, for making it possible for us to recruit our participants for Experiment 3; and the hundreds of families who volunteered to participate in their studies. The data, the stimuli, and the code for this study are publicly accessible (see also references to Hamilton et al., 2022a, 2022b, 2023; Experiment 1: https://osf.io/sfxg5/, Experiment 2: https://osf.io/5bdxg/, Experiment 3: https://osf.io/39w7u/). All three experiments were preregistered on Open Science Framework (see https://osf.io/3eg6h, https://osf.io/cjukf, https://osf.io/dxf49, respectively).

Mollie Hamilton served as lead for data curation, project administration, and writing–original draft. Tessyia Roper served in a supporting role for data curation. Erik Blaser served in a supporting role for writing–review and editing. Zsuzsa Kaldy served as lead for supervision and writing–review and editing and served in a supporting role for project administration. Mollie Hamilton, Erik Blaser, and Zsuzsa Kaldy contributed equally to formal analysis and visualization. Erik Blaser and Zsuzsa Kaldy contributed equally to conceptualization and funding acquisition.

Correspondence concerning this article should be addressed to Zsuzsa Kaldy, Department of Psychology, University of Massachusetts Boston, 100 Morrissey Boulevard, Boston, MA 02125, United States. Email: zsuzsa.kaldy@umb.edu

function of time (Barrouillet et al., 2004; Towse & Hitch, 1995), interference theory postulates that forgetting is due to competition between similar items during retrieval (Brown et al., 2007; Endress & Szabó, 2017; Oberauer et al., 2016). Computational model comparisons suggest that the interference theory better explains WM capacity limitations than the decay theory (H.-Y. Lin & Oberauer, 2022; Oberauer et al., 2016). When interference occurs between currently relevant and previously relevant information, it is called proactive interference (PI). In some WM models, PI has been hypothesized as a principal cause of WM limitations (Anderson & Neely, 1996; Bunting, 2006; Endress & Potter, 2014; Kane & Engle, 2000). The goal of the present study is to explore the early development of PI resolution abilities in children between 3 and 8 years of age.

PI is a classic phenomenon first demonstrated by Keppel and Underwood (1962). Participants heard a three-letter trigram (e.g., "CXJ"), followed by a 3-s delay (during which they had to count backward). Participants were able to recall the trigram with about 80% accuracy on the first trial, but performance dropped to approximately 50% by the second trial (Keppel & Underwood, 1962, Experiment 1). Even if encoding was made easier by introducing a simultaneous visual presentation, increasing performance to near ceiling of 100% correct, performance still dropped to about 80% by the second trial (Experiments 2–4). Importantly, this was not due to fatigue, as participants could be "released" from PI—with performance restored to near ceiling on a subsequent trial—if the stimulus was changed, for example, from a trigram of letters to numbers (Wickens et al., 1963).

In repeated tasks over the same set of items and actions, the leading account of how PI exerts its effect is based on temporal distinctiveness (Crowder, 1976; Glenberg & Swanson, 1986). Since no longer relevant information (from previous executions of the task) has been stored along with currently relevant representations, the difficulty lies in retrieving the information with the correct "time stamp" for the task at hand ("Did I add sugar to my coffee already, or am I thinking of yesterday morning?"). The temporal distinctiveness account has been supported by multiple studies (Brown et al., 2007; Souza & Oberauer, 2015).

More recently, the effects of PI have been demonstrated using visual stimuli (Endress & Potter, 2014; Makovski & Jiang, 2008; Mercer & Fisher, 2022). For example, Makovski and Jiang (2008) found that participants in a change detection task were more likely to make an error if the color and location of the current probe matched a distractor item from a previous trial. This effect, however, appeared to be spatially specific: the participants were affected by PI only when the current probe was in the same location as the previous distractor.

While some studies have disputed the size of the PI effect in traditional change detection paradigms (Balaban et al., 2019; P.-H. Lin & Luck, 2012), others have provided robust evidence that visual working memory (VWM) is affected by PI (Endress, 2022; Endress & Potter, 2014; Makovski, 2016). For example, Endress and Potter (2014) showed participants a series of real-world objects (at the same location) and then asked them to judge whether a probe object had been present in the series. Then, they either induced PI by recycling images from a common set from trial-to-trial, or sidestepped PI by using novel images on each trial. They found that PI reduced performance dramatically. Participants could remember up to 30 items in the absence of PI, but the presence of PI reduced capacity to the typical three- to four-item limit (Cowan, 2001). A recent series of experiments by Endress (2022) and studies from our own group (Donenfeld et al., under review) have shown that the effects of PI in VWM are not entirely location-dependent. Thus, in adults, PI has been well established as a fundamental limitation of VWM.

Despite the theoretical significance and extensive cognitive and neuroscientific literature on the mechanisms underlying PI in adults (Badre & Wagner, 2005; Gray et al., 2003; Jonides & Nee, 2006; Postle et al., 2004), little is known about how PI resolution develops in young children, leaving a major gap in the literature (for a review, please see Hamilton et al., 2022). This is particularly relevant as WM capacity in infants is low (Kaldy & Leslie, 2005; Kaldy et al., 2016; Kibbe & Leslie, 2013; Ross-Sheehy et al., 2003) and only approaches adult levels by middle childhood (Ahmed et al., 2022; Gathercole et al., 2004). Some important prior work with 4- to 7-year-olds measured the effects of PI in LTM in a paired-associate learning task (Darby & Sloutsky, 2015; Yim et al., 2013), and found that the recall accuracy of younger children was more affected by PI than of older children. Other studies with older, school-age children (8- to 14-year-olds) found that WM was affected by PI, and that the ability to resolve PI improved with age (Kail, 2002; Loosli et al., 2014). Most work on PI (with both older children and adults) has used verbal stimuli (e.g., word lists) and/or extensive training protocols, which are not suitable for very young children. To the best of our knowledge, no prior studies have directly measured the ability to resolve PI in children under the age of 4, and no studies have measured how PI affects VWM in children below 8 years of age.

Our preregistered studies (Open Science Framework [OSF], see links in General Method) have three goals. Our first goal was to quantify the effect of PI on WM performance at the youngest age to date (in 2.5- to 3.5-year-old toddlers). Second, we wanted to identify the source of the interfering information. Interference could stem from information in LTM (i.e., from any previous trial) that is activated during retrieval, or, from information in WM that remained active from the just-completed trial (if interference stems from information lingering in WM, then it is not truly PI). To tease these two possibilities apart, we employed a novel experimental design that "flushes" WM between trials; eliminating the potential for interference from information lingering in WM. If PI remains, then, it must be that activated information from LTM is a factor. Third, our goal was to see if we could detect age-related changes, between the ages of 2.5 and 7.5 years of age in a large ($N = 172$) cross-sectional sample, in the ability to resolve PI.

Characterizing the development of PI resolution has three potential impacts. First, it can inform theoretical models of WM and its development. If PI in WM in young children is especially strong, that means that the development of PI resolution abilities is one of the major factors underlying the development of WM capacity. This suggestion was first formulated, but not tested, 30 years ago by Dempster (1993). He made the conjecture that "resistance to interference is a major factor in cognitive development and that it is a basic processing mechanism linked to the efficiency of the frontal lobes" (p. 21). Second, it can lead to applied insights, such as a better understanding of children's learning in the classroom. For instance, it has already been shown that decreasing interference leads to better retention in school-age children, see, for example, Carvalho and Goldstone (2019) and Rohrer et al. (2015). Lastly, it could have important methodological impacts. For instance, in the developmental literature, WM is typically measured by presenting a series of trials with repeated (or similar) stimuli and calculating average performance across trials. If children are more affected by PI than adults, it could mean that we have been systematically underestimating their WM capacity.

General Method

All data and experimental scripts are available for download at OSF (see also references to Hamilton et al., 2022a, 2022b, 2023; Experiment 1: https://osf.io/sfxg5/, Experiment 2: https://osf.io/5bdxg/, Experiment 3: https://osf.io/39w7u/). For preregistrations, please see https://osf.io/3eg6h, https://osf.io/cjukf, and https://osf.io/dxf49, respectively. All three studies were approved by the Institutional Review Board of the University of Massachusetts Boston.

Power Analysis

We conducted the required a priori sample size calculations using G*Power (Faul et al., 2007). This analysis specified that a sample size of 34 would be sufficient to obtain 80% power to detect a medium effect size ($d = 0.5$; $\alpha = .05$) in a paired-sample t test. Both Experiments 1 and 2 met this criterion. For Experiment 3, which was aimed at capturing age-related changes, a sample size of 152 was deemed sufficient to obtain 80% power to detect a small effect size ($r^2 = .05$, $f^2 = 0.0526$; $\alpha = .05$) in a linear regression model with one predictor. Our final sample in Experiment 3 included 172 children.

General Design

All experiments in the current study used a Delayed Match Retrieval paradigm (Kaldy et al., 2016) tailored for a touchscreen tablet (see Figure 1a). The paradigm tests "what was where" object-location bindings in VWM. All stimulus presentation and response collection were controlled by OpenSesame 3.2.7 (Mathôt et al., 2012). We have previously used a (gaze-based) version of this paradigm to test various mechanisms of VWM development from 8 months to 2.5 years of age (Cheng et al., 2019a, 2019b, 2020; Fitch et al., 2021; Kaldy et al., 2016), and others have adopted it as well (Hochmann et al., 2016). Based loosely on the card game memory, after having seen a number of virtual cards briefly exposed (flipped face-up to expose an object on the face of the card, then flipped back face-down) in a fixed set of locations, the participant is shown a sample card and asked to choose the location of its match among the previously exposed, now face-down cards. In the present study, children made their choice by touching the card on the tablet. PI in the current study was generated by repeating items (match and nonmatch cards) across trials (see Figure 1b), such that a participant may mistakenly base a decision about the match card's location on where it had appeared on a previous trial as opposed to where it was in the current trial.

Experiment 1

Participants

Data were collected from 38 toddlers between the ages of 2.5 and 3.5 years of age ($M = 2.89$ years, $SD = 0.29$ years, 20 boys). Seven additional toddlers were tested, but their data were excluded: four due to not meeting the training criterion (see below), two due to experimenter error, and one was excluded as an outlier (i.e., the PI effect (see below) was more than 2.5 median absolute deviations (MAD) from the median (Leys et al., 2013). The outlier-based exclusion was not part of our preregistration of Experiment 1, but we added it for the preregistrations of subsequent experiments, and applied it here for the sake of consistency. Overall, participants contributed 529 valid trials out of the potential 608 trials (87.0%). Caregivers reported that seven participants were multiracial, three were Hispanic or Latino, seven did not wish to report demographic information, and the rest of the sample ($n = 21$) was reported to be white.

Stimuli

The study was designed and run using OpenSesame 3.2.7 (Mathôt et al., 2012). The backs of the virtual cards were solid-colored (green for the sample and blue for the target cards, see below). The faces of the cards depicted images of hard-to-name, unfamiliar 3D objects drawn from the Novel Object and Unusual Name (NOUN) database (Horst & Hout, 2016). We chose to use hard-to-name objects to minimize verbal rehearsal strategies such as naming objects out loud or subvocally. (Subvocal verbal rehearsal emerges in children as young as five and becomes adult-like between 7 and 10 years of age; Elliott et al., 2021; Poloczek et al., 2019). While it is important to note that in typical, everyday situations verbal and visual coding can go on in parallel thereby facilitating remembering in service of solving a task (Overkott et al., 2023), our focus here was on visual WM processes.

Procedure

The toddler was seated at a table with the experimenter in front of a Microsoft Surface Go tablet in a laboratory testing room ($n = 35$) or in a quiet room in the child's home ($n = 3$). The task was explained verbally as well as demonstrated by the experimenter. The caregiver was seated next to the toddler and was asked not to talk or gesture to the toddler during testing. Laboratory sessions were videotaped.

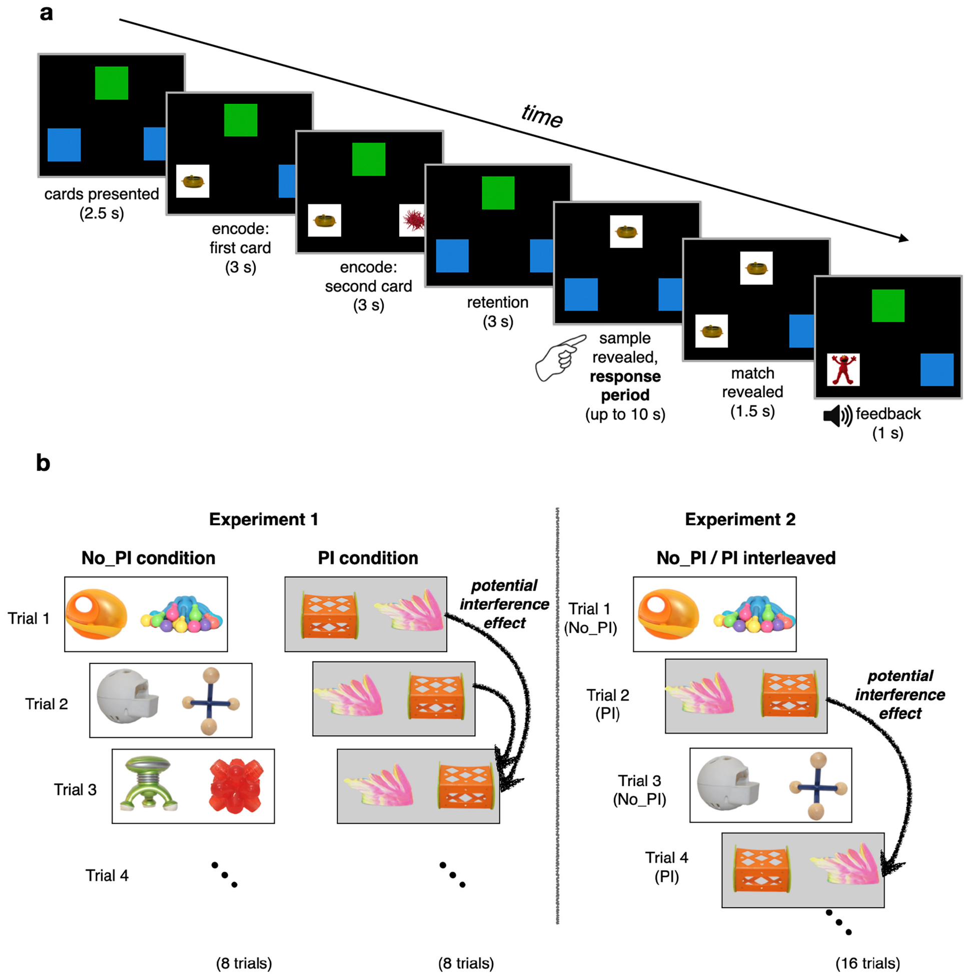
The testing session started with a directions phase (four trials). First, toddlers saw three face-up cards appear on the tablet in a triangular arrangement. The bottom two cards were two different target cards and the top card was the sample that matched one of the targets (i.e., the match card) and did not match the other (i.e., the nonmatch card). The experimenter asked the toddler to pick the card that matched the sample card, and took turns with the toddler selecting and touching the matching card. If the toddler demonstrated difficulty touching the screen, he or she was permitted to point to the selected card and the experimenter touched the card that the toddler indicated.

Upon completion of the directions phase, the training phase (three to eight trials) began. Here, the cards were still presented face-up (no memorization was needed) and the toddler had up to 10 s to respond. If the toddler chose the correct card three trials in a row, the test phase began. If the toddler could not get three trials correct in a row within eight trials, the test phase began, but data from these participants ($n = 4$) were discarded.

The test phase consisted of two sets of eight trials, blocked by condition. Here, toddlers saw three face-down cards appear on the screen in a triangular arrangement. The bottom two cards were the to-be-memorized target cards and the top card was the sample.

Figure 1

Examples of Test Events and Trial Sequences in Experiments 1 and 2



Note. (a) Sequence of events in a test trial in Experiments 1 and 2. (b) Example of trial sequences in Experiment 1 (blocked), and Experiment 2 (interleaved) with sample target stimuli (NOUN database, Horst & Hout, 2016). NOUN = Novel Object and Unusual Name; PI = proactive interference. NOUN database is adapted from "The Novel Object and Unusual Name (NOUN) Database: A Collection of Novel Images for Use in Experimental Research," by J. S. Horst and M. C. Hout, 2016, *Behavior Research Methods*, *48*(4), pp. 1393–1409 (https://doi.org/10.3758/s13428-015-0647-3). Copyright 2016 by SpringerNature. See the online article for the color version of this figure.

First, the target cards were revealed, sequentially (these two cards always showed different images). The sequential reveal ensured that children looked at both images. The first card was exposed for 3 s, then the second one for another 3 s, and then both cards flipped back face-down. After a 3 s retention period, the sample was exposed, which always matched one of the two (now-face-down) target cards. The toddlers had up to 10 s to touch the target card that they thought was the match to the sample (response period). If a toddler's attention wandered or there was apparent confusion, the participant was reminded to find and touch the card that matched. When the correct answer was chosen, the card was flipped over to reveal the match. Additionally, a picture of Elmo appeared accompanied by a correct-answer sound ("brrng!") as feedback before the next trial. If the toddler chose incorrectly, the next trial began without feedback.

The match location (left or right) was pseudorandom, not appearing on the same side more than twice in a row. Trials were blocked by condition. In the PI condition block, the same pair of items was repeated in each of the eight trials as targets (with left/right location also pseudorandomized). In the No_PI condition block, novel items were presented on each of the eight trials. The order of the two blocks was counterbalanced such that half of the toddlers saw the PI block first and half of the toddlers saw the No_PI block first. The two blocks were separated by a 30-s child-friendly cartoon.

Questionnaires were given to caregivers after the completion of the study about the typical frequency of touchscreen device use by the toddler and their prior experience with the game memory (or concentration). Results were entered into exploratory analyses (see the online supplemental materials).

Results

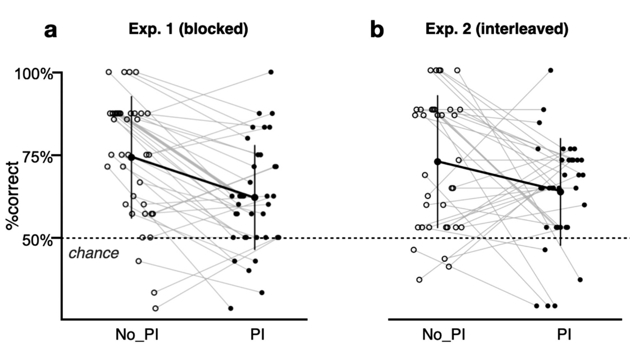
Performance was calculated as the percent of correct trials, averaged within a block for a particular participant, and then across participants within a particular condition. As expected, we found that performance was significantly higher in the No_PI condition ($M = 74.2\%$, $SD = 18.5\%$) than in the PI condition ($M = 62.1\%$, $SD = 15.9\%$, see Figure 2a) indicating that 3-year-old children are susceptible to interference, paired t test, $t(37) = 4.355$, $p < .001$, $d = 0.702$. We then calculated a PI effect, as a difference score between performance in the conditions (No_PI performance − PI performance). The average PI effect was $M = 24.1\%$ ($SD = 34\%$). Whether toddlers were tested in the PI or No_PI condition first did not have a significant effect on the PI effect, independent-samples t test, $t(36) = 0.527$, $p = .601$, $d = 0.172$.

We further tested whether the series of repetitions, trial-by-trial, in the PI condition led to a discernable accumulation of interference. This would be evident if performance dropped in the PI condition as a function of trial, and did so more quickly than in the No_PI condition. Since responses were binary (correct/incorrect) and we had repeated measures, we analyzed these trial-by-trial trends using a generalized estimating equations method to conduct binary logistic regression analyses. In these analyses, response (correct/incorrect) was the dependent variable, and condition and trial number were the independent variables. Both the main effect of condition, Wald $\chi^2(1) = 11.33$, $p < .001$, and trial number were significant, Wald $\chi^2(7) = 31.58$, $p < .001$. That is, toddlers' performance was significantly lower in the PI condition and overall performance declined as the trials went on. As well, the interaction between condition and trial number was significant, Wald $\chi^2(7) = 31.58$, $p < .001$: performance in the PI condition decreased significantly faster than in the No_PI condition (Figure 3a) providing evidence for an accumulation of interference.

Discussion

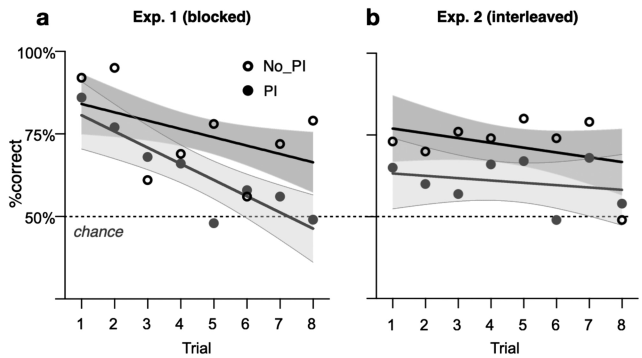
Experiment 1 showed that 3-year-old toddlers are highly sensitive to interference from previously relevant information. While their average performance was near ceiling (87%–92% correct) at the beginning of both the PI and the No_PI blocks, by the end of the block of eight PI trials, toddlers' performance approached chance (50%), while at the end of the eight No_PI trials, performance was still relatively high at 75% correct. This is the first time the effect of PI has been measured in 3-year-olds in a WM task.

Figure 2
Average WM Performance (Percent Correct) in the PI Versus No_PI Trials



Note. (a) Experiment 1 (blocked trials), (b) Experiment 2 (interleaved trials). individual data (gray), group means, and 95% CIs (black) are shown. WM = working memory; PI = proactive interference; CI = confidence interval; Exp. = experiment.

Figure 3
Average WM Performance as a Function of Trial Number (Percent Correct) in the PI Versus No_PI Trials



Note. (a) Experiment 1 (blocked trials), (b) Experiment 2 (interleaved trials). Data in the interleaved design are shown here de-collated, such that trial numbers refer to relative sequencing within the condition. Highlighted regions indicate the 95% confidence bands for the regression fits. WM = working memory; PI = proactive interference; Exp. = experiment.

We also found that WM performance decreased as a function of trial in the PI condition suggesting that interference accumulates as stimuli are repeated over trials. The fact that performance also decreased (to a much lesser extent) over trials in the No_PI condition indicates either general fatigue or item-nonspecific interference (Postle & Brush, 2004) stemming from the general family resemblance of the objects in the NOUN database or interference from the target's location on the previous trial (see, e.g., Makovski & Jiang, 2008).

This experiment left two questions unanswered. First, the blocked design did not allow us to distinguish whether interference arose, as expected, from a reactivation of information drawn from a longer-term store (e.g., "activated LTM"; Cowan, 1988; Lewis-Peacock & Postle, 2008; Oberauer, 2002) or instead from "residual" information still lingering in WM from the just-completed trial. The interference that comes from "residual" information can simply mean that toddlers did not clear their workspace fully between trials. Second, it is possible that toddlers performed more poorly in the PI versus the No_PI condition because they lost interest when playing the game with the same pair of items throughout the block of eight trials (as opposed to the novel pairs used trial-by-trial in the No_PI condition). Experiment 2 was aimed at addressing these two issues with a novel design solution: interleaving PI and No_PI trials in a single mixed block.

Experiment 2

Participants

Data were collected from 35 toddlers between the ages of 2.5 and 3.5 years of age ($M = 3.05$ years, $SD = 0.375$ years, 13 boys), in a testing room in the laboratory ($n = 18$), in a separate room in a children's museum ($n = 16$), or at the toddler's home ($n = 1$). We applied the same exclusion criteria as in Experiment 1 (all were preregistered). Five additional toddlers were tested, but their data were excluded due to not meeting the training criterion (see above). There were no outliers. Overall, participants contributed 486 valid trials out of the potential 560 trials (86.8%). Caregivers reported that three participants were Asian, one was Black/African American, one was Hispanic or Latino, one was multiracial, 22 were white, and eight did not wish to report demographic information.

Stimuli and Procedure

Stimuli and procedures were similar to Experiment 1. The only difference was that the No_PI and PI trials, which had been blocked by condition in Experiment 1, were now interleaved: the eight PI trials were alternated with the eight No_PI trials to create one block of 16 test trials. As before, PI trials always reused the same pair of stimuli from trial-to-trial, but now, given this design, that repetition occurred after an intervening No_PI trial (each using, as before, novel stimuli from trial-to-trial). In this way, like a palate cleanser, the demands of each new No_PI trial flush WM of any information lingering from the just-completed PI trial (see arrows in Figure 1b). If a PI effect still occurs with this design, then that interference must have stemmed from activated LTM.

As well, positive results in this interleaved block design help rule out an alternate explanation for lower performance on PI in Experiment 1, where trials were blocked by condition. One could argue that the lower performance in the PI block could have resulted from a loss of interest given the unbroken repetition of test stimuli trial-by-trial, and not from interference per se. Here, in the current study, PI trials do not form one undifferentiated block, but instead are broken up by the interleaved No_PI trials with their novel stimuli. This mixed design helps to maintain engagement but also means that, even if general engagement does change over the course of a block, it will affect both types of trials.

Results

As expected, a paired-sample t test revealed that toddlers' performance was significantly higher, $t(33) = 2.148$, $p = .039$, $d = 0.36$, in the No_PI condition ($M = 70.9\%$, $SD = 21.2\%$) than in the PI condition ($M = 61.3\%$, $SD = 17.1\%$, see Figure 2b), indicating that toddlers were sensitive to PI even when PI trials were interleaved with other (No_PI) trials. The average PI effect (No_PI performance − PI performance) was $M = 9.3\%$ ($SD = 26.5\%$). As per the request of a reviewer, we also conducted an exploratory analysis on the effect of testing location (museum/home vs. in-lab) on performance. An independent-sample t test showed that overall performance was not significantly different between these sites, $t(33) = 0.467$, $p = .644$, $d = 0.161$.

Trial-by-trial analyses using generalized estimating equations were conducted as in Experiment 1 (Figure 3b). Both the main effect of condition, Wald $\chi^2(1) = 7.05$, $p = .008$, and trial number were significant, Wald $\chi^2(7) = 20.33$, $p = .005$, indicating that PI performance was lower than No_PI performance and that performance overall dropped as a function of trial, but the interaction between condition and trial number was not significant, Wald $\chi^2(7) = 2.86$, $p = .897$, meaning we did not detect a significant accumulation of interference over trials.

Discussion

Experiment 2 showed that 3-year-old toddlers are still sensitive to the effects of PI, even when VWM has been "flushed" between trials thereby eliminating residual information lingering in WM from the just-completed trial. This means that PI can stem from activated LTM representations. To be clear, aspects of interference resolution play out in WM, and WM is the active workplace where "what was where" judgments are made. What our results show is that a source of interfering information is activated LTM. As well, the existence of the PI effect in the present experiment provides evidence that the PI effect in Experiment 1 was not due simply to a loss of interest in the repeated stimuli that comprised the PI block (since here, trials are interleaved, sidestepping the potential for differences in engagement among blocked conditions).

The size of the PI effect here was smaller than in Experiment 1. This was not unexpected, since the interleaving of No_PI trials likely mitigated the potential for PI in two ways. For one, this condition provides for more distinctiveness for trials. Each No_PI trial breaks up the sequence of PI trials, thereby providing distinct (especially given the novel stimuli) "landmarks" (Unsworth et al., 2008) within the trial sequence. Relatedly, while there are disagreements about when and how elapsed time affects memory (Cowan, 2022), most models of interference would expect the greatest interference from the most recent trials, but here, unlike in Experiment 1, the longer sequence of trials means that there is both more elapsed time and more elapsed trials between potentially interfering PI trials (i.e., PI trial n no longer has PI from trial $n - 1$, but only trial $n - 2$, etc.). Hartshorne observed PI effects in VWM in adults that lasted for four to five trials in a change detection task (Hartshorne, 2008), while our results show that in our 3-year-old children, this window does not seem to reach much beyond the $(n - 2)$th trial. Second, while this Experiment was designed to test, in isolation, whether activated LTM can be a source of PI, of course, information remaining in WM may also be a source of interference, so by removing it, one would expect a smaller PI effect.

Experiment 3

The goal of Experiment 3 was to investigate the developmental trajectory of the ability to resolve PI in WM between the ages of 2.5 and 7.5 years of age (against the backdrop of general age-related improvements in VWM performance). Two prior studies have demonstrated age-related increases in PI resolution ability but they both tested older children, between 8 and 14 years of age (Kail, 2002; Loosli et al., 2014).

Participants

All participants were recruited at the Children's Museum of New Hampshire in Dover, NH, United States. The final sample included 172 children between the ages of 2.5 and 7.5 years ($M = 5.03$ years, $SD = 1.15$ years, 78 boys). Seventy-eight of these children were between the ages of 2.5 and 5.5 years ($M = 4.13$ years, $SD = 0.69$ years, 36 boys) and were tested with the two-card version, and 94 of them were between the ages of 4.5 and 7.5 years ($M = 5.78$ years, $SD = 0.89$ years, 42 boys) who were tested with the three-card version. Caregivers reported that 125 of the 172 participants were non-Hispanic or Latino, 17 were Hispanic or Latino, and 30 did not provide information regarding ethnicity. Additionally, caregivers reported that 117 of the 172 participants were white, four were Black or African American, three were American Indian or Alaska Native, six were Asian, seven were multiracial, six identified as other, and 27 chose not to respond.

An additional 13 children were tested, but their data were excluded due to missing two or more (out of four) practice trials ($n = 2$), missing more than three trials per block ($n = 5$), not wanting to proceed to the test trials after the practice phase ($n = 5$), or because their PI effect was more than 2.5 MAD from the median ($n = 6$; Leys et al., 2013). Overall, participants contributed 1,961 valid trials out of the potential 2,064 trials (95.0%); 867/936 (92.6%) in the two-card version and 1,094/1,128 (97.0%) in the three-card version. (All exclusion criteria were the same as in the previous experiments and were preregistered.)

Stimuli and Procedure

The design was nearly identical to Experiment 1, with four modifications.

1. The training phase trials were identical to test trials. In Experiments 1 and 2, cards in training trials had always been face-up, eliminating the need for memorization. Here training was not face-up because, in a pilot study, the training was too easy for older children. None of the images used in training trials were repeated in test trials.
2. To increase the pace of the session, the presentation of cards (encoding) was shorter (2 s, instead of 3 s).
3. The number of trials per block was reduced from eight to six trials. The children's museum had many other activities that competed for children's and caregivers' time, so it was more challenging to have longer testing sessions versus in the lab setting. Reducing the number of trials ensured that the majority of the children completed testing.
4. We introduced a more difficult version of the task that used three cards, instead of two, just for older children (4.5- to 7.5-year-olds, see Figure 4). This increase in memory load was tailored for the older children so as to avoid ceiling effects in performance (which had been present in our pilot sample). By design, the younger age group (2.5- to 5.5-year-olds) was tested with the same two-card version as in Experiments 1 and 2 and in this way, this served as an internal replication for PI effects in this age range. In order to test whether load on its own affected PI, we had specifically recruited extra participants within a "crossover" age range (between 4.5 and 5.5 years of age). In our analyses, this allowed us to separate out two groups of age-matched participants, one that ran in the two-card version and the other in the three-card version, thereby effectively removing age as a variable and isolating the effect of memory load.

Results

Since the two-card and three-card games had different chance levels (i.e., the expected base rate of success, even with random responses, was 50% and 33.3%, respectively) we applied a "correction for guessing" so that the results could be more directly compared. This was done using the following formula:

$$\text{corrected Proportion Correct} = \frac{\#\text{correct responses} - \dfrac{\#\text{incorrect responses}}{\#\text{choices} - 1}}{\#\text{all responses}}. \tag{1}$$

Correction for guessing is a standard method in psychometrics (see, e.g., Link, 1982); this particular version of it is from Frary (1988). It is helpful to keep in mind that with corrected values, a value of 0 corresponds to chance-level performance.

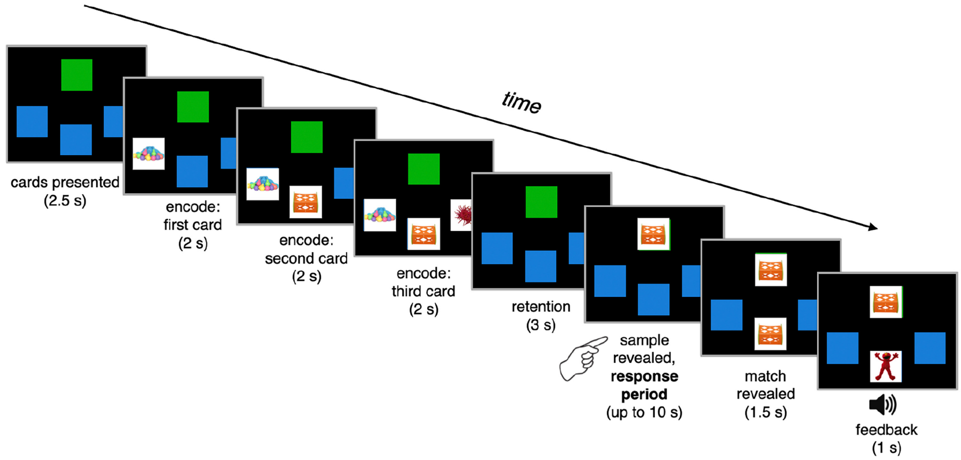
This experiment was designed with an "age overlap group" of children at the same age (4.5- to 5.5-year-olds) with some tested in the two-card version and some in the three-card version. First, we tested whether memory load (two-card vs. three-card version) influenced the PI effect by comparing these two groups of age-matched children. Twenty-three children were tested in the two-card group ($M = 4.93$ years, $SD = 0.26$ years, 14 boys) and 41 in the three-card group ($M = 4.97$ years, $SD = 0.34$ years, 18 boys). An independent-samples t test showed that the PI effects, corrected for guessing, in these two age-matched groups were not significantly different, two-card: $M = 7.97\%$, $SD = 30.29\%$; three-card: $M = 4.14\%$, $SD = 44.30\%$; independent-samples t test, $t(62) = 0.368$, $p = .714$, $d = 0.09$. Figure 5c shows the mean performance in each condition for the overlap and nonoverlap groups separately. In sum, our results from the two groups of 4.5- to 5.5-year-olds using the two different memory load versions demonstrated that while the higher-load, three-card version may have been more challenging (performance in the No_PI condition in these two groups was significantly different, two-card: $M = 69.27\%$, $SD = 31.7\%$, three-card: $M = 47.32\%$, $SD = 32.12\%$; independent-samples t test, $t(62) = 2.64$, $p = .011$, $d = 0.69$, this did not have a significant impact on the PI effect itself.

With this in mind, we ran a set of analyses combining data from the two-card and three-card versions ($N = 172$), and found that performance (corrected for guessing) was significantly higher in the No_PI condition ($M = 56.8\%$, $SD = 34.88\%$) than in the PI condition ($M = 49.97\%$, $SD = 35.45\%$); children between the ages of 2.5 and 7.5 years were sensitive to the effect of interference, paired-samples t test, $t(171) = 2.33$, $p = .021$; $d = 0.18$. The average PI effect (No_PI performance − PI performance, corrected for guessing) was $M = 6.8\%$ ($SD = 38.4\%$). Similar to the first two experiments, the order of conditions (that is, whether children were tested in the PI or No_PI condition first) did not have a significant effect on participants' PI effect, independent-samples t test, $t(170) = 0.784$, $p = .434$, $d = 0.215$. To explore the effect of age on PI resolution, we treated age as a continuous variable. While consistent with an age-related decrease in the PI effect, the resulting trend was not significant, $r^2 = .005$, $F(1, 170) = 0.90$, $p = .34$, see Figure 7a.

Following these analyses on the entire sample, we then performed a set of within-group analyses in order to facilitate a comparison between these results and those from Experiments 1 and 2; here, the group of younger children (tested in the two-card version) had a similar mean age to those tested in Experiments 1 and 2, thereby providing an internal replication and extension. We first ran our central test for the influence of PI on performance. Consistent with the results of Experiments 1 and 2, the group of younger children

Figure 4

Sequence of Events in a Test Trial in Experiment 3 (Three-Card Version)

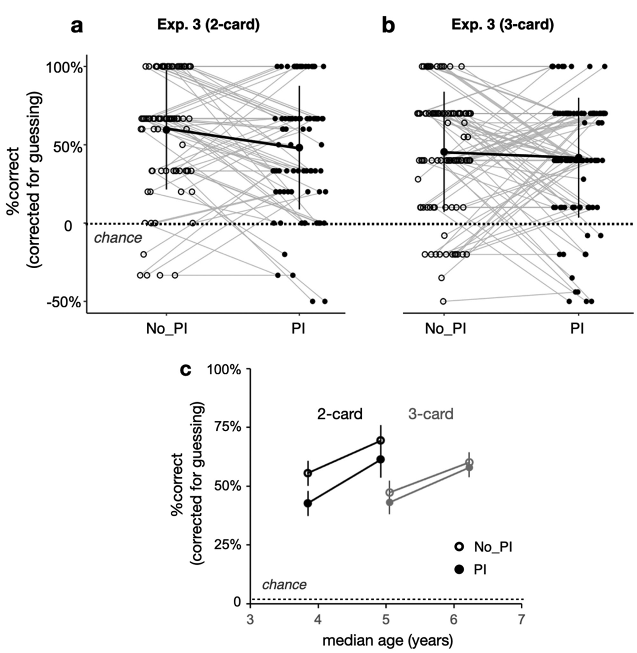


Note. Events in the two-card version were the same as in Experiment 1 (see Figure 1a), except that the two encoding periods were 2 s. See the online article for the color version of this figure.

showed significantly higher performance in the No_PI block ($M = 59.53\%$, $SD = 38.1\%$) than in the PI block, $M = 48.16\%$, $SD = 39.45\%$, $t(77) = 2.66$, $p = .009$, $d = 0.3$, yielding an average PI effect of $M = 11.37\%$ ($SD = 37.71\%$). However, older children (tested in the three-card version), though results were in the expected direction, did not show significantly higher performance in the No_PI block ($M = 54.55\%$, $SD = 32.0\%$) than in the PI block, $M = 51.46\%$, $SD = 31.9\%$; $t(93) = 0.8$, $p = .443$, $d = 0.08$, see Figure 5a and b, with an average PI effect of $M = 3.08\%$ ($SD = 38.8\%$). In a follow-up exploratory (nonpreregistered) analysis, we then compared the PI effect between the two age groups. The independent-samples t test showed that the difference between the groups was small and not statistically significant, $t(170) = 1.411$, $p = .16$, $d = 0.22$.

Figure 5

Average WM Performance (Percent Correct, Corrected for Guessing) in the PI Versus No_PI Conditions in Experiment 3



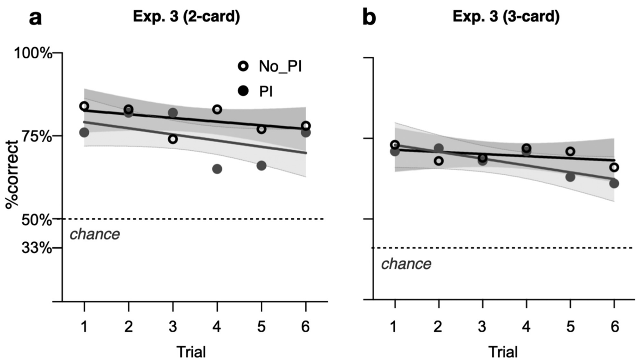
Note. (a) Two-card version, younger children (M_{age}: 4.13 years), (b) three-card version, older children (M_{age}: 5.78 years). Individual data (gray), group means, and 95% CIs (black) are shown. (c) Average WM performance in the overlapping versus nonoverlapping age groups (error bars: ± 1 *SEM*). WM = working memory; PI = proactive interference; CI = confidence interval; Exp. = experiment; *SEM* = standard error of the mean.

As in Experiments 1 and 2, trial-by-trial analyses using generalized estimating equations were used to see if we could detect the accumulation of PI as a function of trial number. This analysis was performed separately for the two groups because the two data sets could not be combined in one logistic regression, as 1 (success) means different performance in the two- versus the three-card versions (for results, see Figure 6a and b). In younger children (tested in the two-card version), the main effect of condition was significant, Wald $\chi^2(1) = 7.182$, $p = .007$, but trial number was not, Wald $\chi^2(5) = 6.492$, $p = .261$. Accumulation is evidenced by a relatively larger drop in PI performance, versus No_PI, as a function of trial number, that is, an interaction. Here, the interaction of condition and trial number was not far from the traditional significance threshold: Wald $\chi^2(5) = 10.513$, $p = .062$. In older children (tested in the three-card version), neither of the main effects, condition: Wald $\chi^2(1) = 0.595$, $p = .440$; trial number: Wald $\chi^2(5) = 4.543$, $p = .474$, nor the interaction between condition and trial number, Wald $\chi^2(5) = 1.503$, $p = .913$ were significant.

Finally, in an exploratory analysis, we tested whether WM performance (as measured by performance in our No_PI condition) increased with age. These analyses were not preregistered, but are valuable to show a ground-truth relationship, here, of age-related improvement in VWM performance in our task. We ran linear regressions looking at the relationship between performance in the No_PI condition and age. Since No_PI performance was significantly different between the two age-overlap groups (see above), we did not combine data from the two-card and the three-card

Figure 6

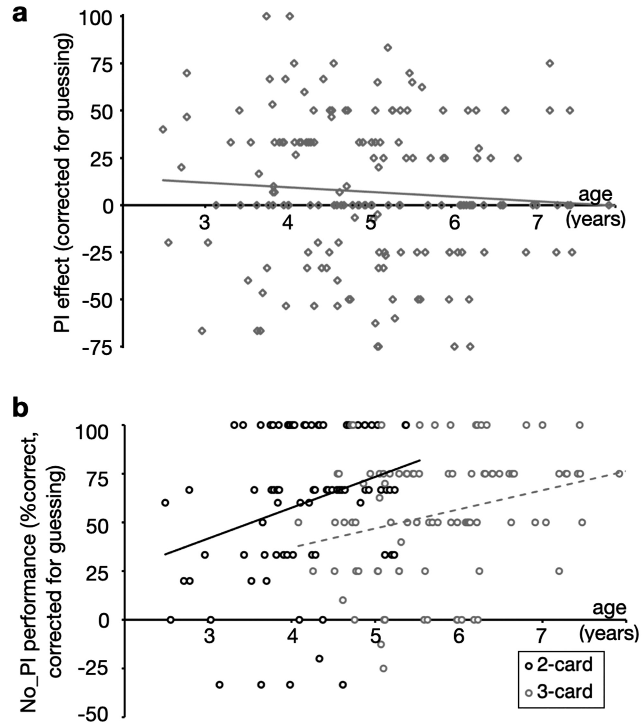
Average WM Performance as a Function of Trial Number (Percent Correct) in the PI Versus No_PI Conditions in Experiment 3



Note. (a) Two-card version, younger children (M_{age}: 4.13 years), (b) three-card version, older children (M_{age}: 5.78 years). Please note that these data are not corrected for guessing, as the logistic regressions were conducted on the binary (1: correct, 0: incorrect) data. (Thus, results can be directly compared to those in Figure 3.) Highlighted regions indicate the 95% confidence bands for the regression fits. WM = working memory; PI = proactive interference; Exp. = experiment.

Figure 7

Linear Regression Analyses



Note. (a) PI effect as a function of age (across the whole sample, $N = 172$), (b) WM performance in the No_PI condition as a function of age in the two-card version, with younger children (M_{age}: 4.13 years, $N = 78$, black circles), and the three-card version with older children (M_{age}: 5.78 years, $N = 94$, gray circles). WM = working memory; PI = proactive interference.

versions for these analyses. Linear regressions showed that No_PI performance increased significantly with age both in the two-card version, $r^2 = .081$, $F(1, 76) = 6.71$, $p = .011$, and in the three-card version, $r^2 = .074$, $F(1, 92) = 7.35$, $p = .008$, see Figure 7b.

Discussion

Experiment 3 was conducted on a large ($N = 172$) cross-sectional sample of 2.5- to 7.5-year-old children. This experiment was similar to Experiment 1, with adaptations to the procedure for the children's museum setting and to children's age (a two-card memory load version for younger children, and a three-card version for older ones). Overall, as in Experiments 1 and 2, we found a weak, but significant main effect of PI ($p = .021$; $d = 0.18$). We also found that in the age-matched groups that were tested in both versions (4.5–5.5 years of age), the PI effects did not differ significantly, suggesting that the PI effect (after correction for guessing) in the two versions of the task can be compared. When age was treated as a continuous variable, while consistent with an age-related decrease in the PI effect, the trend toward smaller PI effects as a function of age was not significant ($r^2 = .005$, $p = .34$).

Broken down by age group, the results with younger (average age: 4.13 years, range: 2.5- to 5.5-year-olds) children showed the same pattern as the 3-year-olds in Experiment 1: a significant overall PI effect in WM ($p = .009$, $d = 0.3$) that showed some evidence of accumulation over trials ($p = .062$). This represents an internal replication of the PI effect we found in Experiment 1 with the 3-year-old sample. However, there was no significant PI effect in the older group (average age: 5.78 years, range: 4.5- to 7.5-year-olds; $p = .443$, $d = 0.08$). Effect sizes showed a small-to-medium PI effect in the younger ($d = 0.3$) and a very small effect ($d = 0.08$) in the older group; however, a direct comparison of the two age groups showed that the difference was not significant ($p = .16$, $d = 0.22$). Overall, the evidence for age-related change in the PI effect presented here is weak, as it only shows up in the group-level comparison, and not with age as a continuous measure, which is considered to be the statistically more appropriate method (MacCallum et al., 2002). We can speculate why we did not find a robust developmental effect. Unfortunately, the museum setting, with its multiple distractions, meant that WM performance in our task (as estimated by the No_PI condition) was not particularly high (two-card: 59.5%, three-card: 54.5%; both values are corrected for guessing). As well, some of the other changes that we needed to implement to keep children engaged (face-down training trials, fewer trials in each block) may have also diminished the PI effect. We suspect that by changing some of these task parameters, the observed developmental change in the PI effect would be more robust (similar to what was found by Kail, 2002 and by Loosli et al., 2014 in 8- to 14-year-olds).

General Discussion

PI occurs when previous, now-irrelevant memories intrude on current, task-relevant ones. Interference is promoted, then, by reusing to-be-remembered stimuli from trial-to-trial (as in our PI condition) or, alternatively, minimized by employing novel stimuli on each trial (as in our No_PI condition). Our delayed match retrieval VWM paradigm required toddlers to solve a "what was where" task. In the repeated (PI) condition, having seen a particular object at a particular location on a previous trial interfered with maintaining its location in the current trial. Through a series of three preregistered[1] experiments, we tested the effect of PI on VWM in children between 2.5 and 7.5 years of age, including the youngest participants yet tested. The relatively large behavioral effect in the 3-year-old children (in Experiments 1 and 2, and confirmed in the younger, 2.5- to 5.5-year-old group of Experiment 3) opens up the field to future studies investigating the development of neural mechanisms of interference resolution, which have been studied extensively in adults (Jonides & Nee, 2006; Kliegl & Bäuml, 2021; Öztekin et al., 2009).

Experiment 1 showed that 3-year-old children are highly susceptible to the detrimental effects of PI, with significantly lower performance in the PI compared to the No_PI conditions. Indeed, by the end of the eight-trial block, WM performance in the PI condition was no better than chance (Figure 3a).

Experiment 2 was designed to determine whether PI can stem from reactivated, LTMs. To accomplish this, we had to remove the potential interference from "residual" information lingering in VWM from the preceding trial. We did this by employing an interleaved design where No_PI trials alternated with PI trials. Like a palate cleanser, information in VWM from each PI trial was "flushed" by the demands of the intervening No_PI trial. The fact that 3-year-old toddlers' WM performance was still significantly lower in PI trials than in No_PI trials (Figures 2b and 3b)—that is, that a significant PI effect remained—means interference can stem from recently activated LTMs alone.

In Experiment 3, we sought to characterize how the ability to inhibit previously relevant, but now irrelevant information develops between 2.5 and 7.5 years of age. Thirty years ago, Dempster suggested that maturation of the ability to resolve interference—to better distinguish relevant and irrelevant information at the moment of retrieval—may at least partially underlie the developmental increase in effective WM capacity (Dempster, 1993). While we found, as predicted by numerous previous studies (Ahmed et al., 2022; Gathercole et al., 2004; Simmering, 2012), that WM performance, overall, increased as a function of age, our results provided only weak evidence for age-related increases in interference resolution in our sample. When age was treated as a continuous variable, the trend toward smaller PI effects as a function of age was not significant. While at the group level, we found that the effect of PI was more robust in our younger (2.5- to 5.5-year-olds; tested in the two-card version) versus older (4.5- to 7.5-year-olds; tested in the three-card version) children ($d = 0.3$ vs. 0.08); a direct comparison of the two age groups did not show a significant difference in the magnitude of the PI effect.

Understanding how children resolve interference is important to understanding how they learn best, particularly in school settings. Indeed, children seem to learn better when interference-inducing repetition of similar concepts is avoided and instead, concepts from one topic are interleaved with dissimilar concepts from another topic (for a review, see Rohrer, 2012). For example, learning the difference between transduction, translation, transcription, and transformation in biology is particularly difficult for students as they are similar in spelling and meaning. Learning is more optimal when these topics are separated by discussions of different concepts in order to reduce confusion between similar terms (Rohrer, 2012; Sana & Yan, 2022). In another example, multiplication problems that share more digits in common with other multiplication problems are harder to solve because the shared digits increase interference, even in adults (De Visscher & Noël, 2014b). As well, some researchers have even suggested that difficulty in resolving interference may be one of the mechanisms underlying dyscalculia, a learning disability characterized by difficulty with learning math facts (De Visscher & Noël, 2014a).

Finally, accounting for PI in WM development has important methodological implications. Current methods used to measure WM in young children typically use multiple trials with repeated stimuli (Kaldy & Leslie, 2003; Pailian et al., 2016; Simmering, 2012) thereby inadvertently promoting PI (Hamilton et al., 2022). This issue was first discussed by primate researchers who reported that macaque monkeys' WM performance drastically decreased across trials when stimuli were repeated (Wright et al., 1986). Neuropsychologists pointed out the same confounding effect in assessments of WM in patients (Brophy et al., 2009). Thus, some characterizations of children's WM may be underestimated, not representing the true capabilities or developmental trajectories of children. The present study provides an illustrative example of this. Let us say a group of researchers, unfamiliar with the role of PI in memory, had conducted a test of 3-year-old children's VWM capacity using a setup similar to our Experiment 1. Given our results, if they chose to repeat to-be-remembered stimuli over trials (unwittingly inducing PI), they would obtain a VWM capacity estimate nearly ½ item lower than if they had opted to use novel stimuli in each trial. (Once corrected for guessing, capacity can be estimated simply from $\text{correctedProportionCorrect} \times \text{memorySetSize}$ (Luck & Vogel, 2013). So, for the PI version, that means a capacity estimate of 0.48 (0.24×2) and, for the No_PI version, a capacity estimate of 0.96 (0.48×2)—a difference of $0.96 - 0.48 = 0.48$ item). This is a dramatic underestimate, especially in a developmental context where even best-case capacity estimates are already well under the adult three- to four-item limit (Cowan, 2001).

Limits on the Generalizability of Our Findings

Our convenience sampling yielded child participants representing the racial and ethnic distribution of the Northeastern United States, largely from middle-income families. These participants are likely familiar with tablet-based games (though prior touchscreen experience or experience with the game memory did not seem to affect WM performance, see the online supplemental materials). Planned future studies using an app-based, further gamified version of our paradigm will reach a geographically wider and globally more representative sample.

Summary

In a series of preregistered experiments, we found that 3-year-old toddlers are highly sensitive to the effects of PI that stemmed from the multiple repetitions of items in a VWM task. Their VWM capacity dropped to chance level by the end of a series of eight trials, and PI was present even when trials with repeated items were interleaved with nonrepeated items. When we tested a large cross-sectional

[1] It is worth noting that while preregistration has become a more common practice in experimental psychology in the past decade, it is still not the norm in studies with young children. A recent meta-analysis found that the median effect size of 93 preregistered studies across different subdisciplines of psychology was 0.16 (Schäfer & Schwarz, 2019). We believe this provides an important context for the effect sizes of the current results.

sample of children between 2.5 and 7.5 years of age and found a similar PI effect in the younger portion of our sample, and no PI effect in the older ones. However, when the effect of age was examined as a continuous variable across the entire sample, it was not significant. These findings are a first step in understanding the early development of a core cognitive control mechanism (interference resolution, or in a broader sense, memory updating), and open an avenue to study its neural mechanisms in young children. Our findings also provide evidence that previous studies that repeat to-be-remembered information across trials may unwittingly induce PI, thereby underestimating young children's WM capacity.

References

Ahmed, S. F., Ellis, A., Ward, K. P., Chaku, N., & Davis-Kean, P. (2022). Working memory development from early childhood to adolescence using two nationally representative samples. *Developmental Psychology*, *58*(10), 1962–1973. https://doi.org/10.1037/dev0001396

Anderson, M. C., & Neely, J. H. (1996). Chapter 8—Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 237–313). Academic Press. https://doi.org/10.1016/B978-012102570-0/50010-0

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Badre, D., & Wagner, A. D. (2005). Frontal lobe mechanisms that resolve proactive interference. *Cerebral Cortex*, *15*(12), 2003–2012. https://doi.org/10.1093/cercor/bhi075

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901. https://doi.org/10.1016/j.neuropsychologia.2007.06.015

Balaban, H., Fukuda, K., & Luria, R. (2019). What can half a million change detection trials tell us about visual working memory? *Cognition*, *191*, Article 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

Brophy, L. M., Jackson, M., & Crowe, S. F. (2009). Interference effects on commonly used memory tasks. *Archives of Clinical Neuropsychology*, *24*(1), 105–112. https://doi.org/10.1093/arclin/acp013

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

Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology*, *17*(2), 243–250. https://doi.org/10.1016/j.conb.2007.02.005

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 183–196. https://doi.org/10.1037/0278-7393.32.2.183

Carvalho, P. F., & Goldstone, R. L. (2019). When does interleaving practice improve learning? In J. Dunlosky (Ed.), *The Cambridge handbook of cognition and education* (Vol. 729, xviii, pp. 411–436). Cambridge University Press. https://doi.org/10.1017/9781108235631.017

Cheng, C., Kaldy, Z., & Blaser, E. (2019a). Focused attention predicts visual working memory performance in 13-month-old infants: A pupillometric study. *Developmental Cognitive Neuroscience*, *36*, Article 100616. https://doi.org/10.1016/j.dcn.2019.100616

Cheng, C., Kaldy, Z., & Blaser, E. (2019b). Two-year-olds succeed at MIT: Multiple identity tracking in 20- and 25-month-old infants. *Journal of Experimental Child Psychology*, *187*, Article 104649. https://doi.org/10.1016/j.jecp.2019.06.002

Cheng, C., Kaldy, Z., & Blaser, E. (2020). Coding of featural information in infant Visual Working Memory. *Cognitive Development*, *55*, Article 100892. https://doi.org/10.1016/j.cogdev.2020.100892

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114; discussion 114–185. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2022). Working memory development: A 50-year assessment of research and underlying theories. *Cognition*, *224*, Article 105075. https://doi.org/10.1016/j.cognition.2022.105075

Crowder, R. G. (1976). *Principles of learning and memory*. Lawrence Erlbaum Associates.

Darby, K. P., & Sloutsky, V. M. (2015). The cost of learning: Interference effects in memory development. *Journal of Experimental Psychology: General*, *144*(2), 410–431. https://doi.org/10.1037/xge0000051

Dempster, F. N. (1993). Resistance to interference: Developmental changes in a basic processing mechanism. In M. L. Howe & R. Pasnak (Eds.), *Emerging themes in cognitive development: Volume I: Foundations* (pp. 3–27). Springer. https://doi.org/10.1007/978-1-4613-9220-0_1

De Visscher, A., & Noël, M.-P. (2014a). Arithmetic facts storage deficit: The hypersensitivity-to-interference in memory hypothesis. *Developmental Science*, *17*(3), 434–442. https://doi.org/10.1111/desc.12135

De Visscher, A., & Noël, M.-P. (2014b). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *Journal of Experimental Psychology: General*, *143*(6), 2380–2400. https://doi.org/10.1037/xge0000029

Donenfeld, J., Blaser, E., & Kaldy, Z. (under review). The resolution of proactive interference in a novel visual working memory task: A behavioral and pupillometric study. *Attention, Perception & Psychophysics*. https://doi.org/10.31234/osf.io/yfajp

Elliott, E. M., Morey, C. C., AuBuchon, A. M., Cowan, N., Jarrold, C., Adams, E. J., Attwood, M., Bayram, B., Beeler-Duden, S., Blakstvedt, T. Y., Büttner, G., Castelain, T., Cave, S., Crepaldi, D., Fredriksen, E., Glass, B. A., Graves, A. J., Guitard, D., Hoehl, S., … Voracek, M. (2021). Multilab direct replication of Flavell, Beach, and Chinsky (1966): Spontaneous verbal rehearsal in a memory task as a function of age. *Advances in Methods and Practices in Psychological Science*, *4*(2), Article 25152459211018187. https://doi.org/10.1177/25152459211018187

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., & Szabó, S. (2017). Interference and memory capacity limitations. *Psychological Review*, *124*(5), 551–571. https://doi.org/10.1037/rev0000071

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fitch, A., Thaker, N., & Kaldy, Z. (2021). Object labels do not boost preverbal infants' visual working memory. *Infant Behavior and Development*, *64*, Article 101617. https://doi.org/10.1016/j.infbeh.2021.101617

Frary, R. B. (1988). Formula scoring of multiple-choice tests (correction for guessing). *Educational Measurement: Issues and Practice*, *7*(2), 33–38. https://doi.org/10.1111/j.1745-3992.1988.tb00434.x

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 3–15. https://doi.org/10.1037/0278-7393.12.1.3

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322. https://doi.org/10.1038/nn1014

Hamilton, M., Blaser, E., & Kaldy, Z. (2022a). *Proactive interference in 3-year-olds' visual working memory*. https://osf.io/sfxg5/

Hamilton, M., Blaser, E., & Kaldy, Z. (2022b). *What is the source of proactive interference in young children's visual working memory?* https://osf.io/5bdxg/

Hamilton, M., Blaser, E., & Kaldy, Z. (2023). *The resolution of proactive interference in young children*. https://osf.io/39w7u/

Hamilton, M., Ross, A., Blaser, E., & Kaldy, Z. (2022). Proactive interference and the development of working memory. *WIRES Cognitive Science*, *13*(3), Article e1593. https://doi.org/10.1002/wcs.1593

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

Hochmann, J.-R., Mody, S., & Carey, S. (2016). Infants' representations of same and different in match- and non-match-to-sample. *Cognitive Psychology*, *86*, 87–111. https://doi.org/10.1016/j.cogpsych.2016.01.005

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*(1), 181–193. https://doi.org/10.1016/j.neuroscience.2005.06.042

Kail, R. (2002). Developmental change in proactive interference. *Child Development*, *73*(6), 1703–1714. https://doi.org/10.1111/1467-8624.00500

Kaldy, Z., Guillory, S. B., & Blaser, E. (2016). Delayed Match Retrieval: A novel anticipation-based visual working memory paradigm. *Developmental Science*, *19*(6), 892–900. https://doi.org/10.1111/desc.12335

Kaldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: Integrating "what" and "where" information. *Developmental Science*, *6*(3), 360–373. https://doi.org/10.1111/1467-7687.00290

Kaldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, *97*(2), 153–177. https://doi.org/10.1016/j.cognition.2004.09.009

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 336–358. https://doi.org/10.1037/0278-7393.26.2.336

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*(3), 153–161. https://doi.org/10.1016/S0022-5371(62)80023-1

Kibbe, M. M., & Leslie, A. M. (2013). What's the object of object working memory in infancy? Unraveling "what" and "how many". *Cognitive Psychology*, *66*(4), 380–404. https://doi.org/10.1016/j.cogpsych.2013.05.001

Kliegl, O., & Bäuml, K.-H. T. (2021). Buildup and release from proactive interference—Cognitive and neural mechanisms. *Neuroscience & Biobehavioral Reviews*, *120*, 264–278. https://doi.org/10.1016/j.neubiorev.2020.10.028

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *The Journal of Neuroscience*, *28*(35), 8765–8771. https://doi.org/10.1523/JNEUROSCI.1953-08.2008

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lin, H.-Y., & Oberauer, K. (2022). An interference model for visual working memory: Applications to the change detection task. *Cognitive Psychology*, *133*, Article 101463. https://doi.org/10.1016/j.cogpsych.2022.101463

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*, Article 42. https://doi.org/10.3389/fpsyg.2012.00042

Link, S. W. (1982). Correcting response measures for guessing and partial information. *Psychological Bulletin*, *92*(2), 469–486. https://doi.org/10.1037/0033-2909.92.2.469

Loosli, S. V., Rahm, B., Unterrainer, J. M., Weiller, C., & Kaller, C. P. (2014). Developmental change in proactive interference across the life span: Evidence from two working memory tasks. *Developmental Psychology*, *50*(4), 1060–1072. https://doi.org/10.1037/a0035231

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. https://doi.org/10.1037/1082-989X.7.1.19

Makovski, T. (2016). Does proactive interference play a significant role in visual working memory tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1664–1672. https://doi.org/10.1037/xlm0000262

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

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

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

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

Overkott, C., Souza, A. S., & Morey, C. C. (2023). The developing impact of verbal labels on visual memories in children. *Journal of Experimental Psychology: General*, *152*(3), 825–838. https://doi.org/10.1037/xge0001305

Öztekin, I., Curtis, C. E., & McElree, B. (2009). The medial temporal lobe and the left inferior prefrontal cortex jointly support interference resolution in verbal working memory. *Journal of Cognitive Neuroscience*, *21*(10), 1967–1979. https://doi.org/10.1162/jocn.2008.21146

Pailian, H., Libertus, M. E., Feigenson, L., & Halberda, J. (2016). Visual working memory capacity increases between ages 3 and 8 years, controlling for gains in attention, perception, and executive control. *Attention, Perception, & Psychophysics*, *78*(6), 1556–1573. https://doi.org/10.3758/s13414-016-1140-5

Poloczek, S., Henry, L. A., Messer, D. J., & Büttner, G. (2019). Do children use different forms of verbal rehearsal in serial picture recall tasks? A multi-method study. *Memory*, *27*(6), 758–771. https://doi.org/10.1080/09658211.2018.1563615

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

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, *107*(3), 900–908. https://doi.org/10.1037/edu0000001

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development*, *74*(6), 1807–1822. https://doi.org/10.1046/j.1467-8624.2003.00639.x

Sana, F., & Yan, V. X. (2022). Interleaving retrieval practice promotes science learning. *Psychological Science*, *33*(5), 782–788. https://doi.org/10.1177/09567976211057507

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*, Article 813. https://doi.org/10.3389/fpsyg.2019.00813

Simmering, V. R. (2012). The development of visual working memory capacity during early childhood. *Journal of Experimental Child Psychology*, *111*(4), 695–707. https://doi.org/10.1016/j.jecp.2011.10.007

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

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 108–124. https://doi.org/10.1080/14640749508401379

Unsworth, N., Heitz, R. P., & Parks, N. A. (2008). The importance of temporal distinctiveness for forgetting over the short term. *Psychological Science*, *19*(11), 1078–1081. https://doi.org/10.1111/j.1467-9280.2008.02203.x

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

Wright, A. A., Urcuioli, P. J., & Sands, S. F. (1986). Proactive interference in animal memory. In D. F. Kendrick, M. E. Rilling, M. R. Denny (Eds.), *Theories of animal memory* (pp. 101–125). Lawrence Erlbaum Associates.

Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The development of episodic memory. *Psychological Science*, *24*(11), 2163–2172. https://doi.org/10.1177/0956797613487385

Received August 2, 2022
Revision received September 3, 2023
Accepted November 6, 2023 ■