Research Article

Psychological Science
2022, Vol. 33(7) 1154–1171

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09567976211073131
www.psychologicalscience.org/PS
SAGE

Sources of Interference in Memory Across Development

Hyungwook Yim[1,2], Adam F. Osth[2], Vladimir M. Sloutsky[3], and Simon J. Dennis[2]

[1]Department of Cognitive Sciences, Hanyang University; [2]Melbourne School of Psychological Sciences, The University of Melbourne; and [3]Department of Psychology, The Ohio State University

Abstract

Episodic memory involves remembering not only what happened but also where and when the event happened. This multicomponent nature introduces different sources of interference that stem from previous experience. However, it is unclear how the contributions of these sources change across development and what might cause the changes. To address these questions, we tested 4- to 5-year-olds (n = 103), 7- to 8-year-olds (n = 82), and adults (n = 70) using item- and source-recognition memory tasks with various manipulations (i.e., list length, list strength, word frequency), and we decomposed sources of interference using a computational model. We found that interference stemming from other items on the study list rapidly decreased with development, whereas interference from preexperimental contexts gradually decreased but remained the major source of interference. The model further quantified these changes, indicating that the ability to discriminate items undergoes rapid development, whereas the ability to discriminate contexts undergoes protracted development. These results elucidate fundamental aspects of memory development.

Keywords

memory development, episodic memory, interference, context noise, hierarchical Bayesian model, open data

Received 10/7/20; Revision accepted 11/26/21

Episodic memory involves remembering multiple components of an event such as *what*, *where*, and *when* an event happened (Tulving, 1972). For example, in a courtroom, the testimony becomes valid not when the eyewitness remembers the suspect's face but when they remember the face within the context of a certain place and time. However, our ability to form episodic memories is not perfect and is prone to interference stemming from previous experiences. Therefore, when the witness provides the testimony, inaccuracies (or forgetting) can be driven by other similar faces they have encountered during the event or other similar places, times, and faces that they have experienced before the event.

Identifying different sources of interference and understanding their involvement in memory retrieval has been an important part of building theories of episodic memory (Anderson & Bower, 1972; Criss & Shiffrin, 2004; Dennis & Humphreys, 2001; Osth & Dennis, 2015). For example, suppose that you are trying to recognize whether you met your colleague Adam at the party yesterday (see Fig. 1a). If you actually met him yesterday, there would be a corresponding trace in your memory, which will create a matching signal to the probe (i.e., a *self-match*). However, other nontarget memories will produce interference. First, you may be confused because you met other people at yesterday's party. These memories will match the to-be-recognized cue in terms of context (i.e., yesterday's party) but mismatch the item (i.e., Adam). This form of interference has been called *item noise* (Gillund & Shiffrin, 1984; Shiffrin & Steyvers, 1997). Additionally, you may also be confused because you met Adam at other places in the

Corresponding Authors:
Hyungwook Yim, Hanyang University, Department of Cognitive Sciences
Email: hwyim@hanyang.ac.kr

Simon J. Dennis, The University of Melbourne, Melbourne School of Psychological Sciences
Email: simon.dennis@unimelb.edu.au

past. These memories match the item cue (i.e., Adam) but mismatch the relevant context (i.e., yesterday's party). This form of interference has been called *context noise* (Dennis & Humphreys, 2001). Finally, memories that do not match either the item or the context also produce interference (Murdock & Kahana, 1993a, 1993b; Osth & Dennis, 2015). This form of interference is called *background noise*.

These sources of interference have been studied with different experimental manipulations. For example, item noise has been measured by comparing memory tasks with different numbers of study items by either adding more unique items (i.e., list-length manipulation; Gronlund & Elam, 1994) or repeating some of the items in the list to evaluate their impact on the nonrepeated items (i.e., list-strength manipulation; Ratcliff et al., 1990). Context noise has been examined by manipulating the preexperimental frequency of items. For example, when words are used as stimuli, low-frequency words and high-frequency words are compared under the assumption that less frequent words appear in fewer sentence or story contexts and are thus expected to create less context noise.

Computational models of memory have been used to quantitatively decompose different sources of interference. For example, Osth and Dennis (2015) used a computational model to show that the composition of interference varies across different studies depending on the preexperimental familiarity of the stimulus. For instance, unfamiliar stimuli (i.e., fractal patterns) generate more item noise than familiar stimuli (i.e., words) because unfamiliar stimuli are harder to distinguish from each other than familiar ones. Moreover, the model showed that frequently encountered stimuli (e.g., high-frequency words) create more context noise than those that are less frequently encountered (e.g., low-frequency words) because more frequent stimuli appear in more contexts and generate more interference during retrieval.

An important implication of these findings is that there are both costs and benefits of experience on memory performance. With more experience, more memory traces will be stored, and these traces may become sources of interference. On the other hand, with more experience, the representation of the stimulus becomes more distinct and less confusing, which results in less interference.

In a similar way, the accumulation of experience may also affect memory performance across development. As a child grows, both the number of memory traces and the robustness of representations will tend to increase. Therefore, because young children have less experience than adults, the costs and benefits of experience may be relatively small for young children compared with adults. Moreover, because memory performance typically improves during childhood, the benefit of experience may outweigh the cost at some point in development.

However, it is unclear how the costs and benefits of experience change across development, how different aspects of experience contribute, and what drives these changes. Previous studies show that (a) memory accuracy increases throughout early development (e.g., Ackerman, 1984; Bauer, 2007; Cycowicz et al., 2001; Hayes et al., 2017; Mandler & Robinson, 1978) and (b) the ability to remember different components of an event (e.g., item, context) may have a different developmental trajectory (e.g., Lee et al., 2020; Ngo et al., 2019; Picard et al., 2012). For example, Picard et al. (2012) provided evidence that the ability to remember the content of the event (i.e., item) reaches adultlike performance levels much earlier (around age 7 years) than the ability to remember the context of the event, such as the time and space of the event. However, the consequences of prior experience when testing memory performance were not fully considered in these studies, which makes it hard to evaluate the cost and benefit of experience in memory.

Statement of Relevance

Memory research has shown that memory interference increases with more events to remember. However, even though children accumulate experience throughout development, and the potential for interference may increase, memory performance typically improves. What prevents interference from overwhelming memory as children acquire experience? In this research, we tested memory performance of 4-year-olds, 7-year-olds, and adults. Additionally, we considered the amount of experience one would have accumulated when we analyzed the results. We found that the developmental change is linked to the ability to distinguish memory representations from each other and that this ability develops at a faster rate than experience is accumulated. Moreover, the ability to distinguish specific items from each other emerges early in development, whereas the ability to distinguish different contexts undergoes more protracted development. The study provides a detailed explanation of how children's memory develops and also why even adults are often confused about remembering different contexts.

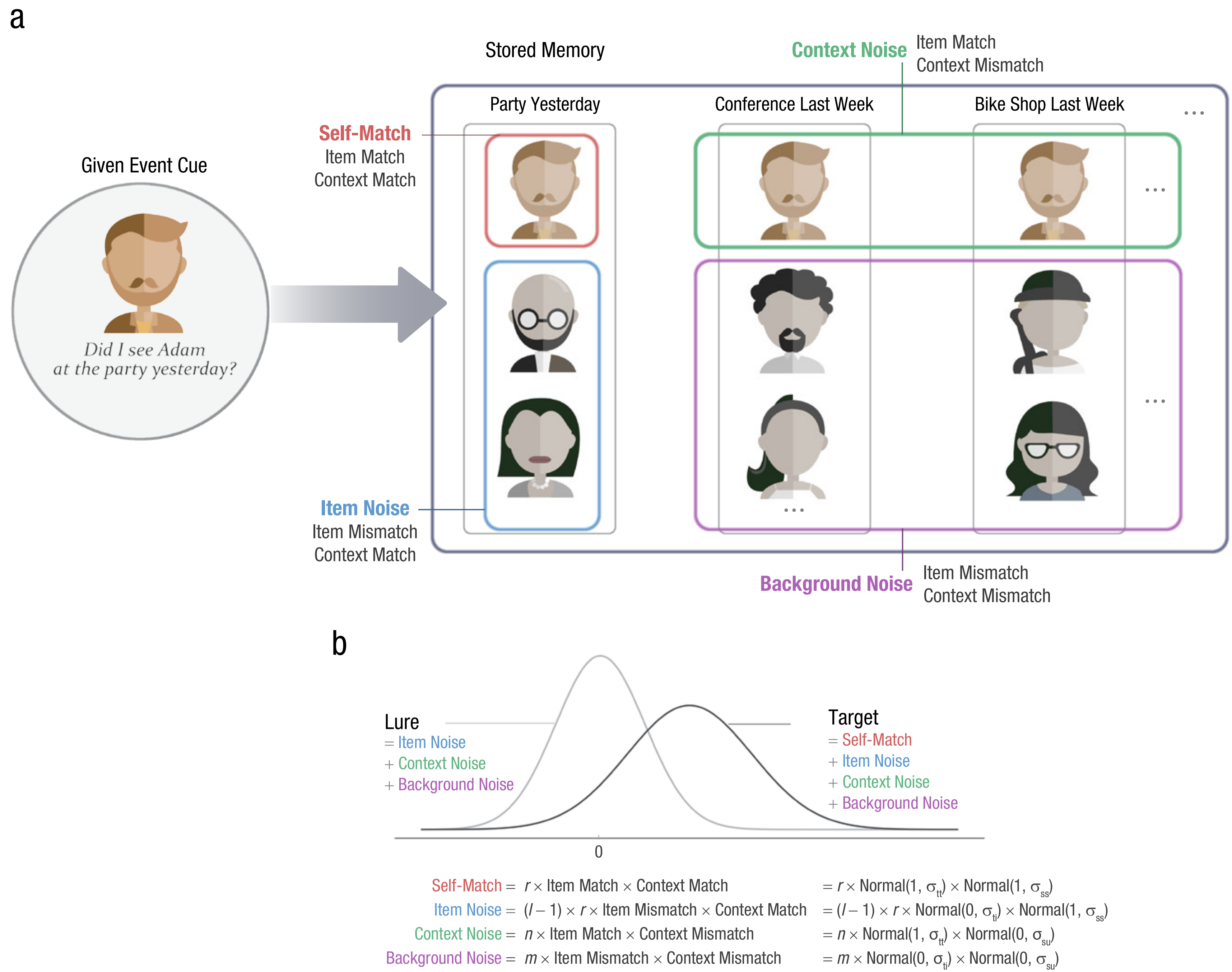

Fig. 1. *(continued on next page)*

Fig. 1. Sources of interference and how they are decomposed by the computational model. Different sources of interference in episodic memory are shown in (a). For example, suppose that you are trying to recognize whether you met your colleague Adam at the party yesterday (see Fig. 1a). If you actually met him yesterday, there would be a corresponding trace in your memory, which will create a matching signal to the probe (i.e., a *self-match*). However, other nontarget memories will produce interference. First, you may be confused because you met other people at yesterday's party. These memories will match the to-be-recognized cue in terms of context (i.e., yesterday's party) but mismatch the item (i.e., Adam). This form of interference has been called *item noise*. Additionally, you may also be confused because you met Adam at other places in the past. These memories match the item cue (i.e., Adam) but mismatch the relevant context (i.e., yesterday's party). This form of interference has been called *context noise*. Finally, memories that do not match either the item or the context also produce interference. This form of interference is called *background noise*. The different components included in the lure and target distribution of the model, and how the model parameterizes these components, are shown in (b). The signal and noise distributions are further parameterized into match and mismatch components. For example, item noise is calculated as the product of the item-mismatch and context-match parameters. Additionally, we multiplied the number of items that were presented in the study list (l) minus 1 in order to incorporate item noise stemming from all mismatching items in the study list. For the components that involve items in the study (self-match and item noise), a learning rate (r_{weak} or r_{strong}, depending on the number of repetitions of the item in the study list) is multiplied and estimated by the model. Context noise is the product of the item-match and context-mismatch parameters multiplied by the number of times participants have encountered the matching item (i.e., word) in their lifetime (n). Finally, background noise is calculated by the product of item-mismatch and context-mismatch parameters multiplied by the number of words encountered in participants' lifetime (m). In the item-recognition task, the model assumes that participants use a generalized source cue that matches the source component in memory storage with a mean value of 1 and a variance of 0. Therefore, for the item-recognition task described above, all terms have an implicit source-match component that is not expressed in the equations. Images are from freepik.com.

To answer these questions, in the current study, we examined how the sources of interference change and affect memory across development. We used item- and source-memory tasks, systematically manipulating different sources of interference in various ways (i.e., list length, repetition of the items, and preexperimental frequency and familiarity of the items) and comparing performance across three different age groups. Then, we investigated the mechanisms that drive these changes by using a computational model that allowed us to decompose different sources of interference in each age group and untangle the cost and benefits of experience's effect on memory performance.

Method

Participants were tested using a memory task that was introduced as a "secret word game." In each game, participants were presented with two speakers (one male and one female) each telling them secret words one at a time. They were asked to remember the secret words (item recognition) and the speaker who said the word (source recognition). There were three different within-subject conditions, which were designed to capture different sources of interference used in previous adult studies (Kinnell & Dennis, 2011, 2012; Osth et al., 2014). Item noise was measured by repeating half of the items (i.e., mixed-strength condition) and by including additional unique items (i.e., long condition); performance in these conditions was compared with a baseline condition. Context noise was measured by using stimuli (i.e., words) of different normative frequencies (Dennis & Humphreys, 2001) because frequency information provides a proxy for how often the word has been experienced preexperimentally. The results were analyzed using a computational model (Osth & Dennis, 2015; Osth, Fox, et al., 2018). The model decomposes the total memory interference into different kinds of interference (item noise, context noise, and background noise) and estimates the ability to discriminate different elements of an event (i.e., benefit of experience) while considering the amount of prior experience the participant received before being tested (i.e., cost of experience).

Participants

One hundred three 4- to 5-year-olds (M = 4.82 years, SD = 0.29, 47 females), eighty-two 7- to 8-year-olds (M = 7.74 years, SD = 1.03, 44 females), and 70 adults (M = 19.16 years, SD = 1.03, 46 females) participated in the experiment. The sample sizes for adults and 7- to 8-year-olds were determined by previous studies that used a similar design (Kinnell & Dennis, 2011; Osth et al., 2014). For 4- to 5-year-olds, we aimed to collect approximately double the sample size of the adult sample because previous memory experiments suggested that the young children's data may be noisier than that of adults (e.g., Yim et al., 2013).

Adults were undergraduate students attending The Ohio State University, who were participating for course credit. Children were recruited from local schools around Columbus, Ohio. The research was approved by The Ohio State University Institutional Review Board. An additional thirty-five 4- to 5-year-olds and one 7- or 8-year-old were excluded from the sample because they had d' scores below zero in more than two test conditions, largely because they were not paying attention to the task.

Materials

From Wordbank (Frank et al., 2017), we selected 150 words that more than 50% of all 30-month-olds in the database knew. There were 75 low- and 75 high-frequency words. For the low-frequency words, the average count per million was 11.4 in the Child Language Data Exchange System (CHILDES) corpus (MacWhinney, 2000) and 4.51 in the Touchstone Applied Science Associates (TASA) corpus (Landauer & Dumais, 1997); for the high-frequency words, the average count per million was 104.68 in the CHILDES corpus and 127.75 in the TASA corpus. The mean age of acquisition for the words was 4.59 years (SD = 1.31; Kuperman et al., 2012). There were significant correlations between the frequencies derived from the CHILDES corpus and the TASA corpus (Spearman's $\rho = .84$, $p < .0001$), between the CHILDES corpus and age-of-acquisition scores ($\rho = -.52$, $p < .0001$), and between the TASA corpus and age-of-acquisition scores ($\rho = -.45$, $p < .0001$). For each word, we created 18 video clips consisting of nine different female and male speakers pronouncing the word. Each speaker appeared on a unique background in order to aid participants to better discriminate between the different speakers (see Fig. 2a).

Procedure

The experiment had four blocks, and each block had a study phase followed by a retention interval and a test phase. In the study phase, participants were told that they would be playing a secret word game in which they would see two speakers each presenting secret words one at a time. Their task would be to remember the secret words and the speaker who said them. Each trial started with a fixation (+++) for 500 ms, followed by a blank screen for 500 ms and then a video clip showing a speaker saying a word for approximately 2,700 ms (see Fig. 2b). In all blocks, one of the speakers was female and the other was male. The video clips were presented on one side of the screen throughout the experiment, depending on the speaker's sex (e.g., female speakers on the left side, male speaker on the right side), and the position was randomized across participants. During the retention interval, participants were presented with two groups of dots on each side of the screen and were told to choose the side that had more dots. After a 500-ms fixation (+++), the adults were presented with the two groups of dots for 250 ms followed by a random-color dot mask, which was presented until a response was made. For children, the dots were presented without the mask until a response was made. The number of dots varied between 10 and 40; we manipulated the ratio of the two numbers. On each trial, the ratio was randomly selected from one of four ratio bins (1.2171–1.4089, 1.4089–1.6308, 1.6965–1.9638, or 2.6260–3.0397 for children and 1.0991–1.1915, 1.1915–1.2917, 1.3302–1.4421, or 2.2906–2.4833 for adults), on the basis of previous studies (Halberda & Feigenson, 2008). In the test phase, participants were auditorily presented with a cue word in a third person's voice (or by the experimenter for the children's groups so that they did not miss the cue word) and were asked whether the word was a secret word (see Fig. 2c). If the participant responded that the word was a secret word, they were then asked who had said the secret word.

The four blocks had three different conditions (baseline, mixed strength, and long); the mixed-strength condition was repeated twice in order to match the total number of target items in other conditions (see Fig. 2d). These three different conditions have been extensively used in the adult literature (e.g., Dennis et al., 2008; Kinnell & Dennis, 2012; Osth, Jansson, et al., 2018). In the baseline condition, eight different words were presented in the study phase, half delivered by a female speaker and the other half by a male speaker. Half of the words for each speaker were high-frequency words, and the other half were low-frequency words. Ten trials were presented during the test phase: Eight were from the study phase (i.e., old items), and two were new words (i.e., lures). In the mixed-strength condition, eight different words were presented in the study phase. Then half of them were repeated two more times in a blocked fashion (i.e., strong items). The repeated words are called "strong" items because they are strengthened with repetition, whereas the nonrepeated words are called "weak" items. During the test phase, the four weak items and four strong items that were presented during the study phase were randomly presented with two lures. In the long condition, 16 words were presented in the study phase. During the test phase, the first half of the items (eight) during the study phase were presented with two lures, and then the second half of the items (eight) were tested with another two lures. As in the baseline condition, the words in the mixed-strength and long conditions were balanced by frequency and the speaker's sex. The length of the retention interval was 60 s for the mixed-strength and long conditions and 89.6 s for the baseline condition. The additional 29.6 s in the baseline condition corresponds to the additional eight trials that the mixed-strength and long conditions have during the study phase; it was added to equate the total length of the retention interval (Kinnell & Dennis, 2011).

Before the main experiment, there was a practice session in which two words were presented during the study phase. One was presented by a female speaker and another by a male speaker. Then during the test phase, three words were presented: two from the study phase (old items) and a new word (lure item). The

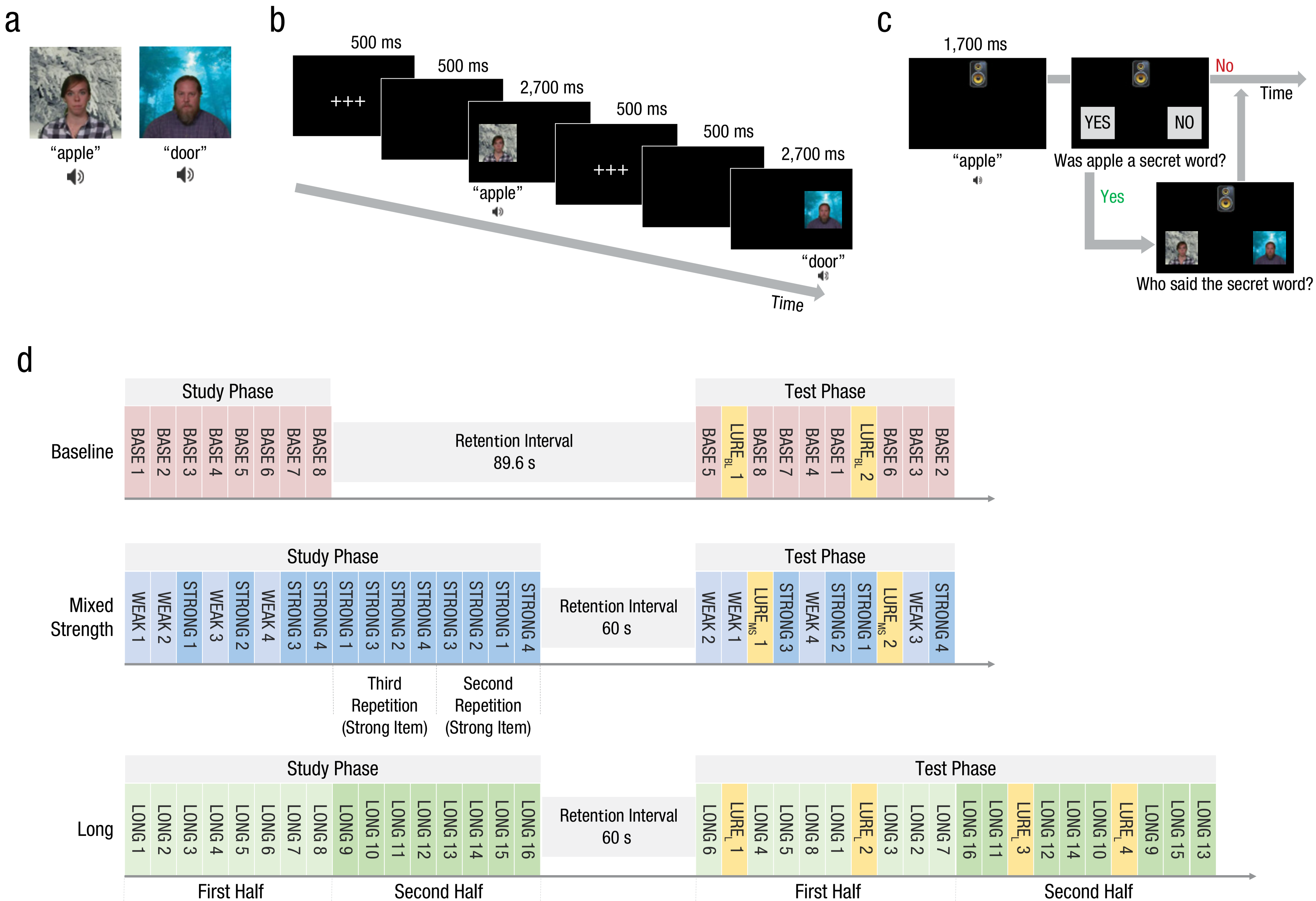

Fig. 2. Stimulus and experimental design. In each video clip used in the experiment, examples of which are shown in (a), one of 18 speakers presented a single word. In the study phase (b), participants saw each clip one at a time and were asked to remember the words and which speaker said them. In the test phase (c), participants first completed an auditory item-recognition test in which they had to provide a yes/no response indicating whether they had heard the word before. If they said "yes," this was followed by a source-recognition test, in which they had to identify the word's speaker. The figure depicts the procedure for adults. The design of the three different conditions is shown in (d). Participants completed three conditions in a random order. In the baseline condition, eight words were presented in the study phase, which was followed by a test phase that additionally included two words that were not presented in the study phase (i.e., lures). The mixed-strength condition differed from the baseline condition in that half of the items were repeated once (weak items), whereas the other half were repeated three times (or strengthened; strong items). The long condition differed from the baseline condition in that there were additional unique words presented, thus making the list longer.

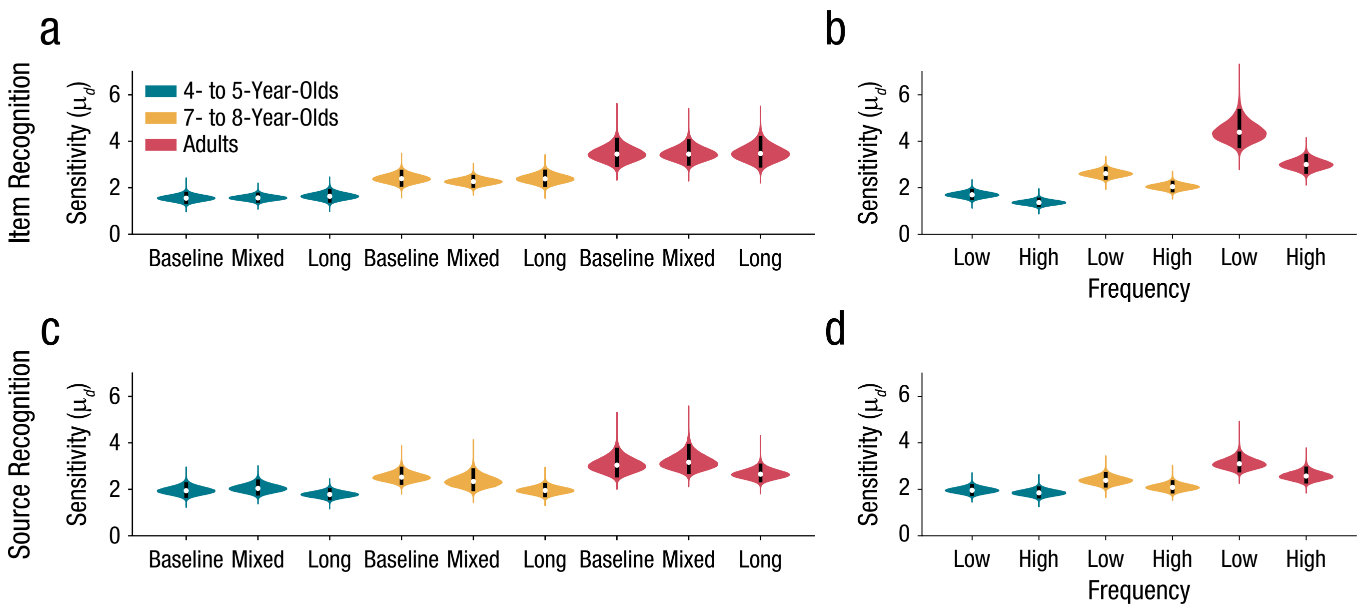

Fig. 3. Behavioral results for item recognition and source recognition. Sensitivity (d') is shown as a function of condition and age for (a) item recognition and (c) source recognition. (b) Sensitivity (d') is shown as a function of word frequency and age for (b) item recognition and (d) source recognition. In all graphs, sensitivity is estimated as a group-level hyperparameter (μ_d). In each violin plot, the white dot and the black line represent, respectively, the median and 95% credible interval of the posterior distribution; the width of each plot indicates the density of the data.

children's groups repeated the identical practice session up to three times if they did not understand the task. All children understood the task within three practice sessions. The speakers and the words used in the practice session were not used in the main experiment.

Presentation of all stimuli was controlled using MATLAB (The MathWorks, Natick, MA) with Psychophysics Toolbox 3 (Kleiner et al., 2007). Video stimuli were presented on a 22-in. monitor, and audio stimuli were presented via headphones. Participants responded by clicking the corresponding image on the screen using a computer mouse (adults) or by touching the touch screen (children). The order of conditions and the combination of the word pairs and speakers were randomized across participants.

Results

Behavioral-data analysis

Behavioral data were analyzed using a signal detection theory (SDT) model to account for the decision biases that are usually reported in studies involving young children. We specifically used a hierarchical Bayesian version of the model (Lee & Wagenmakers, 2014) in order to avoid edge effects (i.e., statistical biases that occur for correcting hit rates and false-alarm rates of 1 or 0 when calculating d'). Moreover, the Bayesian analysis produces Bayes factors (BFs), which unlike the p values used in frequentist analysis, allowed us to evaluate both the null hypothesis and the alternative hypothesis. We used the interpretation of BF_{10} values by Jeffreys (1961), where a BF_{10} value between 1 and 3 is considered weak evidence for the alternative hypothesis, between 3 and 10 is considered substantial evidence, and above 10 is considered strong evidence. On the other hand, a BF_{10} value between 1 and 1/3 is considered weak evidence, between 1/3 and 1/10 is considered substantial evidence, and below 1/10 is considered strong evidence for the null hypothesis. All models used for data analysis were implemented using *STAN* (Version 2.21.0; Carpenter et al., 2017) with 12 chains, each with 60,000 iterations (20,000 warm-up). The convergence of the chains was checked visually, and convergence statistic $\hat{R}$ values were all below 1.01. The full details of the hierarchical Bayesian SDT model and the results are reported in section A of the Supplemental Material available online, which includes all mean values, Bayesian 95% highest-density intervals (HDIs), BFs for individual analysis, and analyses on hit rates and false-alarm rates. Here, we present the main findings for conciseness. Data from this study can be found on our OSF page (https://osf.io/2fg5w/).

For the item-recognition task, overall accuracy increased across age groups, as shown by the group d' parameter (μ_d) of the SDT model (see Fig. 3a; BF_{10}s > 1,800). Item noise was measured in two ways. First, we compared the performance of weak items in the mixed-strength condition with the baseline condition (i.e., the list-strength effect). If there is a significant amount of item noise, weak items (i.e., items that were presented only once in the study list) will be affected by interference caused by strong items (i.e., items that were repeated three times in the study list). We also compared the performance of the items in the first half of the list in the long condition with the baseline condition (i.e., list-length effect). Only the items in the first half were used because they were comparable with the items in the baseline conditions in terms of retention interval, and if there is a significant amount of item noise, items in the first half will be affected by interference caused by items in the second half. Results showed weak support for no list-length effects (see Fig. 3a; 4- to 5-year-olds: $BF_{10} = 0.47$, 7- to 8-year-olds: $BF_{10} = 0.46$, adults: $BF_{10} = 0.36$) and no list-strength effects (4- to 5-year-olds: $BF_{10} = 0.39$, 7- to 8-year-olds: $BF_{10} = 0.45$, adults: $BF_{10} = 0.44$).

If context noise exists, performance should be better for low-frequency words than for high-frequency words. This is because low-frequency words would have been encountered in fewer contexts than high-frequency words, and, therefore, create less interference.[1] Results showed an advantage for low-frequency words (i.e., frequency effect) for the 7-year-olds and adults and a tendency toward such an advantage for 4-year-olds, which replicates findings of previous adult studies and provides evidence for context noise (see Fig. 3b; 4- to 5-year-olds: $BF_{10} = 2.14$, 7- to 8-year-olds: $BF_{10} = 9.28$, adults: $BF_{10} = 13.77$).

Source-recognition performance was also analyzed using a hierarchical Bayesian SDT model, in which we analyzed only the trials on which items were correctly recognized. Results showed an increase in performance across age groups (see Fig. 3c; 4- to 5-year-olds vs. adults: $BF_{10} > 197{,}000{,}000{,}000$, 4- to 5-year-olds vs. 7- to 8-year-olds: $BF_{10} = 1.47$, 7- to 8-year-olds vs. adults: $BF_{10} = 31.09$). We found only a numerical effect for frequency in all age groups (see Fig. 3d; 4- to 5-year-olds: $BF_{10} = 0.34$, 7- to 8-year-olds: $BF_{10} = 0.43$, adults: $BF_{10} = 1.23$). There was also weak evidence for null list-length effects (4- to 5-year-olds: $BF_{10} = 0.34$, 7- to 8-year-olds: $BF_{10} = 3.47$, adults: $BF_{10} = 0.43$) and list-strength effects (4- to 5-year-olds: $BF_{10} = 0.50$, 7- to 8-year-olds: $BF_{10} = 0.45$, adults: $BF_{10} = 0.94$).

In sum, the results from the item-recognition task showed a similar pattern compared with previous studies, in which memory accuracy increased across development (Ghetti & Angelini, 2008; Hayes et al., 2017), and accuracy for low-frequency words was better than for high-frequency words (Dennis & Humphreys, 2001; Glanzer & Bowles, 1976; Glanzer et al., 1993). Also, consistent with previous adult studies that used words as stimuli, our results revealed no evidence for strong list-length or list-strength effects (Kinnell & Dennis, 2011; Osth et al., 2014). Similarly, results from the source-recognition task showed a developmental increase in accuracy as in previous studies (e.g., Drummey & Newcombe, 2002), with a numerical advantage for low-frequency words and no list-strength effect.

Although the results from the behavioral data provide some evidence for the existence of different sources of interference, it is not easy to directly compare the contribution of these sources across different age groups. Therefore, we further analyzed the data using a computational model that can quantitatively decompose the number of different sources of interference. The results from the computational model allowed us to compare the amount of interference across age groups more directly and estimate specific mechanistic changes that underlie these performance differences.

Computational modeling

We modified a computational model by Osth and Dennis (2015), which is implemented in a hierarchical Bayesian framework (Kruschke, 2014). The model has been successful in capturing a number of different episodic-recognition phenomena and is also capable of decomposing the total interference into different sources quantitatively in a way behavioral data on its own does not (Fox et al., 2020; Osth, Fox, et al., 2018; Osth, Jansson, et al., 2018).

Two main components of the model are worth highlighting before we describe the model in greater detail. One is the discriminability of different representations. In the current version, the model represents item, source, and context components as a three-way binding (e.g., Yim et al., 2018), and for each kind of representation, there is a parameter responsible for how well the representations can be distinguished from each other. Higher discriminability between representations reduces interference and improves recognition. Another main component of the model is the number of memory traces experienced. Both the total number of memories and the number of memories corresponding to a particular word determine interference (e.g., Fig. 1a); as the number of memories is increased, there is more interference. These two features, along with the interaction of other components (e.g., learning rate, decision bias), allow for the estimation of the noise that affects recognition performance. Additionally, they allow us to estimate (a) the cost of experience in memory performance by considering the number of items that one has experienced prior to the experiment and (b) the benefit of experience by parameterizing the discriminability of items and contexts.

Here, we describe the model on the conceptual level (the mathematical details can be found in section C in the Supplemental Material). Consistent with other episodic-memory models that explain recognition memory, the current model assumes that when a cue is presented at test, the cue is compared with all stored memories (i.e., global matching; Clark & Gronlund, 1996; Osth & Dennis, 2020). Then the summed similarity (i.e., memory strength) between the cue and all contents of memory is calculated. More specifically, the probe-memory similarity is calculated for each cue dimension (item, context, and source), and the product between them is calculated. As described in the introduction, even if the cue was experienced in the past, the calculated similarity will be a sum of a matching signal (i.e., self-match) along with the different sources of noise from memories that do not match the presented cue (e.g., item noise, context noise, background noise) to produce strength for targets. If the cue was not experienced, only noise is produced, resulting in the strength for lures. Strengths of targets and lures are combined with their expected strengths to produce a log-likelihood ratio that the cue is a target, as depicted in Figure 1b (Glanzer et al., 2009; Osth et al., 2017). When the log-likelihood ratio is above a criterion, the model elicits a "yes" response.

The crux of the model is that the signal and noise distributions are further parameterized into match and mismatch components (see Fig. 1b, bottom). For example, item noise is calculated as the product of the item-mismatch and context-match parameters. Additionally, we multiplied the number of items that were presented in the study list (l) minus 1 in order to incorporate item noise stemming from all mismatching items in the study list. For the components that involve items in the study (self-match and item noise), a learning rate (r_{weak} or r_{strong}, depending on the number of repetitions of the item in the study list) is multiplied and estimated by the model. Context noise is the product of the item-match and context-mismatch parameters multiplied by the number of times participants have encountered the matching item (i.e., word) in their lifetime (n). Finally, background noise is calculated by the product of item-mismatch and context-mismatch parameters multiplied by the number of words encountered in participants' lifetime (m).

The constants n and m are important because they provide information about one's prior experience to the model and incorporate the cost of experience in memory. The two constants were derived from natural language corpora and empirical studies providing better constraints to capture the developmental differences compared with previous versions of the model. We used the empirical estimation that each individual experienced 10,950,000 words each year (Hart & Risley, 1995), and therefore each participant's m was derived by multiplying the individual's exact age by 10,950,000. The number of times a word was experienced in the past (n) was approximated for each individual as well. We derived per-million counts for each word using the CHILDES corpus for the children's data (MacWhinney, 2000) and the TASA corpus for the adults' data (Landauer & Dumais, 1997). Then raw frequency was calculated on the basis of the estimated total number of words experienced (m) for each individual. Each age group was modeled separately and was fitted using the differential-evolution Markov chain Monte Carlo method (Turner et al., 2013) with 21 chains of 10,000 samples (25,000 burn-in, 20 thin) each. The convergences of the chains were checked visually, and all chains showed $\hat{R}$ values below 1.01.

The match and mismatch components are the main parameters of the model, and each follows a normal distribution (see section E in the Supplemental Material for a description of all the parameters that were used in the model). To fix the scale of the model, we set the means of the distributions to 1.0 if they matched the cue and 0.0 if they mismatched the cue, and we estimated the variance of the mismatch parameters. Importantly, the variance of the mismatch parameters indexes the discriminability of the representations and was interpreted as a proxy for the benefit of experience. A higher value of variance of the mismatch parameter indicates lower discriminability. Modeling code from this study can be found on our OSF page (https://osf.io/2fg5w/).

For an illustrative example of how the item noise and context noise affect recognition performance in the model, see the model simulations in Figure 4. We fixed all other parameters but changed the components of the item noise or the components of the context noise. In the model shown in Figure 4a, we changed the length of the study list and the variance of the item-mismatch parameter because item noise is calculated by multiplying these two elements. The simulation shows that performance decreases as both the length of the study list and the variance of the item-mismatch parameter increases, where the performance decrease is manifest in both a decrease in hit rates (solid lines) and an increase in false-alarm rates (dashed lines). In the model shown in Figure 4b, we manipulated word frequency and the variance of the context-mismatch parameter because context noise is calculated by multiplying these two elements. The simulation shows that performance decreases as both the frequency and variance of the context-mismatch parameter increases. See section G in the Supplemental

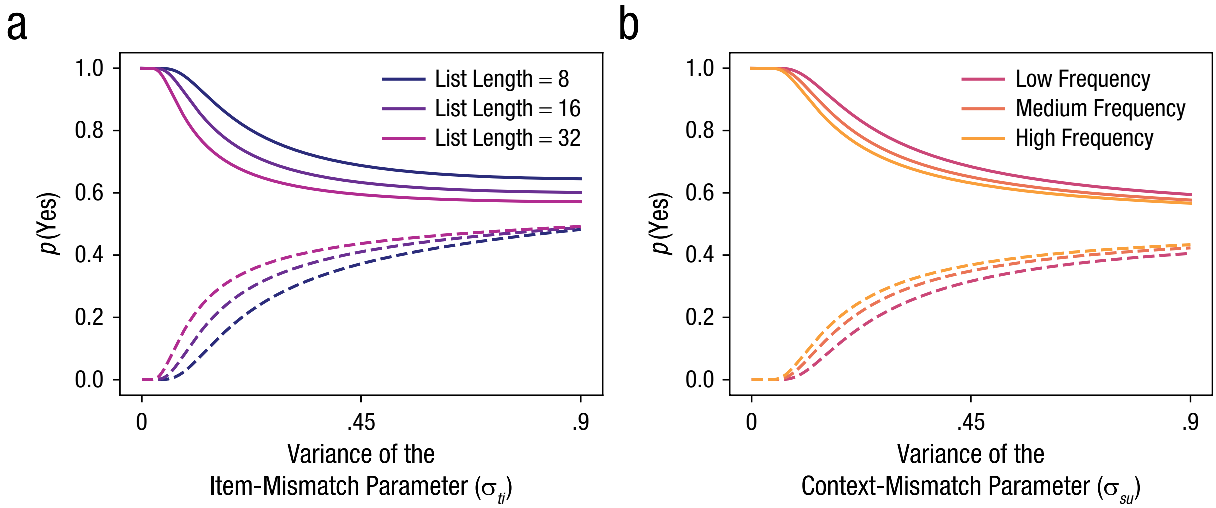

Fig. 4. Model simulations showing recognition performance when item noise and context noise were changed in the model. The probability of a "yes" response is shown as function of (a) variance of the item-mismatch parameter and list length and (b) variance of the context-mismatch parameter and item frequency. Solid lines are hit rates, and dashed lines are false-alarm rates. Because item noise is calculated by multiplying the length of the study list and the variance of the item-mismatch parameter, we changed each component separately. Because context noise is calculated by multiplying the frequency of the item (i.e., word) and the variance of the context-mismatch parameter, we changed each component separately. See section G in the Supplemental Material for additional simulations, in which decision bias is also changed in addition to item- and context-noise components.

Material for additional simulations, including how the model behaves when the decision bias changes along with item and context noise.

We used the model to estimate the different sources of interference across age and condition in the item-recognition task (see Fig. 5 for model fits). As can be seen in Figure 6, the overall amount of interference decreased across age (see section D in the Supplemental Material for detailed values). Most importantly, item noise in all three conditions drastically dropped between the ages of 4 to 5 years and 7 to 8 years, whereas context noise continued to decrease between 7 to 8 years and adulthood and remains the primary source of interference in adults. At the same time, the overall contribution of background noise was negligible throughout development.

Interference in the source-recognition task showed a similar pattern as in item recognition: Item noise rapidly decreased early in development, and background noise was not a major source of interference (see Fig. 7). However, unlike in item recognition, context noise did not change across development. The constant context-noise estimates across development could be explained by the fact that the source-recognition task was administered only when the participants responded that they recognized the item in the item-recognition task, which means that the participants had already discriminated the experimental context from previous contexts. Therefore, the procedure may have filtered out the age difference of context noise.

The modeling results can appear puzzling, as the model is showing that 4- to 5-year-olds have higher item noise, whereas the behavioral results show a null list-length and -strength effect. However, it is worth noting that the model is measuring item noise through both the list-length effect (i.e., performance difference between the baseline condition and the long condition) and the list-strength effect (i.e., performance difference between the baseline condition and the mixed-strength effect) in both memory tasks. Moreover, the model separates other factors that may affect recognition-memory performance (e.g., context noise) from the behavioral data when estimating item noise. Therefore, the model is more sensitive in capturing item noise than merely comparing the behavioral results (e.g., d') between different conditions. A way to examine the connection between the behavioral data and the model parameter is by looking into (a) each participant's list-length and list-strength effect and (b) the proportion of item noise estimated by the model. When conducting a correlation analysis between these two components, there was only a statistically significant correlation between the 4-year-olds' proportion of item noise and list-strength effect score (Spearman's $\rho = .755$, $p < .001$), whereas other relations did not show a statistically significant effect for 4-year-olds (list length: $\rho = .012$,

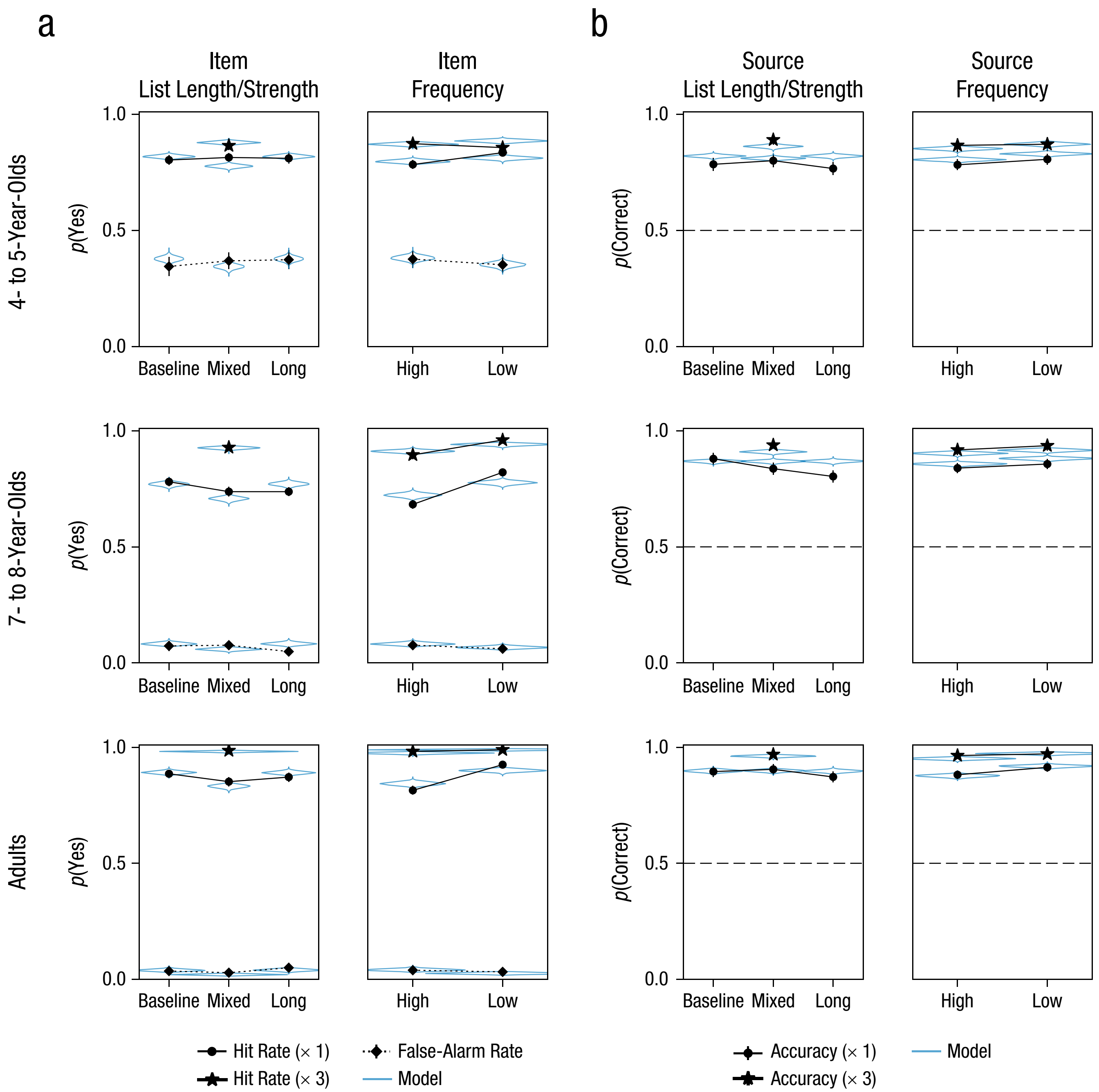

Fig. 5. Model fits for the (a) item-recognition task and (b) source-recognition task across age groups and conditions. In (a), the first column shows hit rates for items presented once (circles), hit rates for items presented three times (stars; only in the mixed-strength condition), and false-alarm rates (diamonds) for the baseline, mixed-strength, and long conditions, which were used to examine list-length and list-strength effects. The second column shows the same data as the first column but arranged in terms of high and low frequencies. In (b), the first column shows accuracy for items presented once (circles) and items presented three times (stars) in each condition. The second column shows the same data as the first column but arranged in terms of high and low frequencies. The dashed lines indicate chance performance (.5). In both panels, the blue violin plots represent model predictions (i.e., posterior predictions), whereas the black markers represent results from the experiment. Error bars represent standard errors of the mean.

p = .91), 7-year-olds (list length: ρ = −.075, p = .54, list-strength: ρ = .205, p = .07) or adults (list length: ρ = −.071, p = .56, list strength: ρ = .087, p = .48). The results imply that the high item noise in the 4-year-olds' data stems from the list-strength effect (see section H in the Supplemental Material for details).

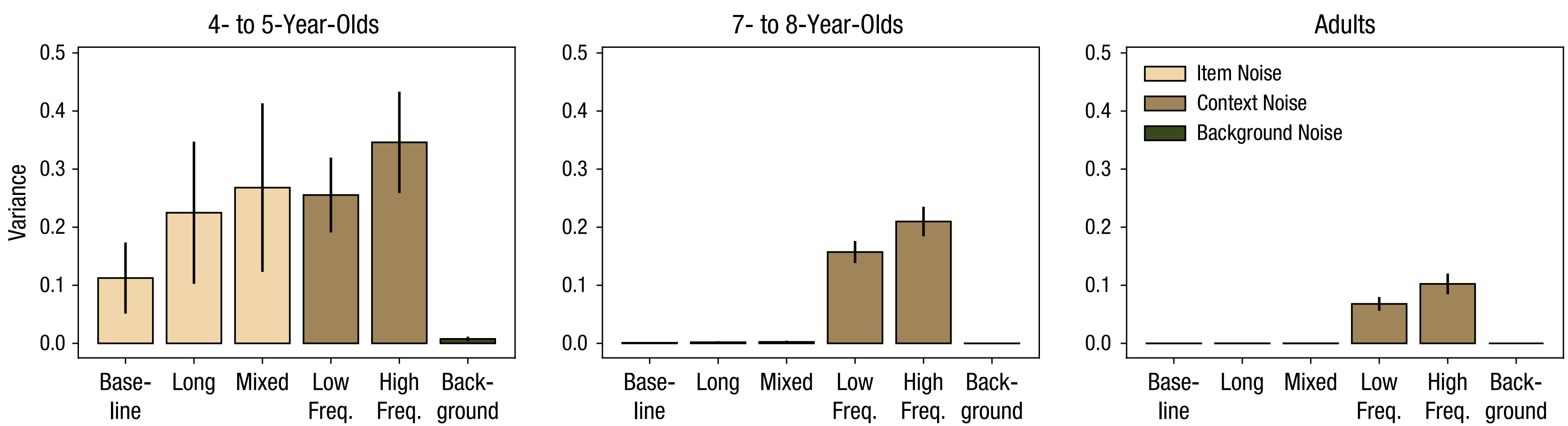

Fig. 6. Sources of interference (noise) estimated by the model in the item-recognition task, separately for each age group. Item-noise variance estimates are shown for each condition (i.e., baseline, long, and mixed strength), and context-noise variance estimates are shown for each frequency (along with the background-noise estimate for comparison). Error bars represent 95% highest-density intervals.

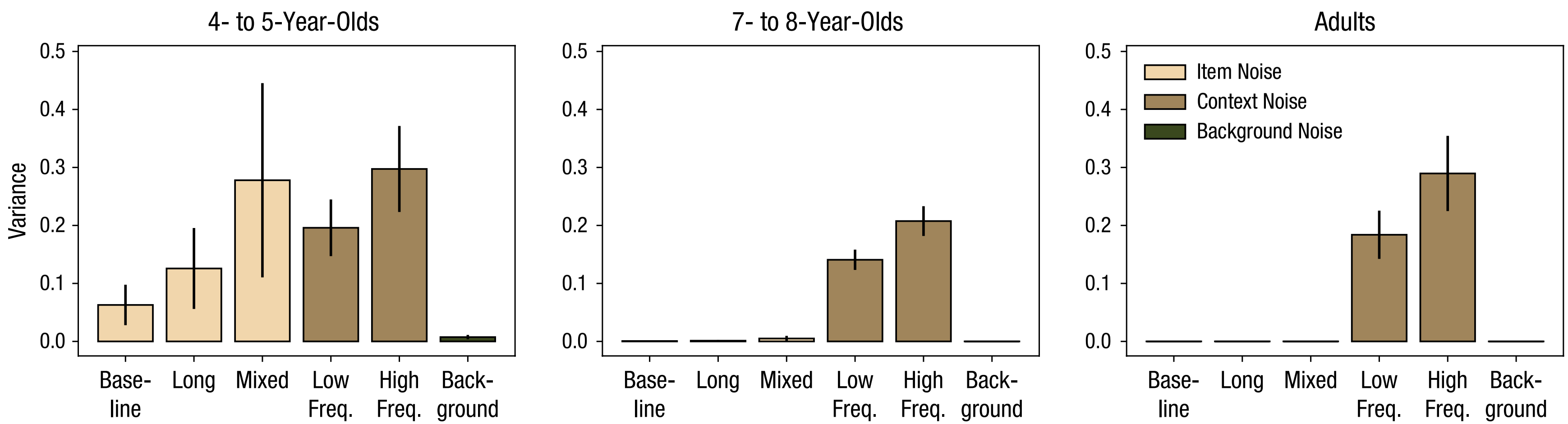

Fig. 7. Sources of interference (noise) estimated by the model in the source-recognition task, separately for each age group. Item-noise variance estimates are shown for each condition (i.e., baseline, long, and mixed strength), and context-noise variance estimates are shown for each frequency (along with the background-noise estimate for comparison). Error bars represent 95% highest-density intervals.

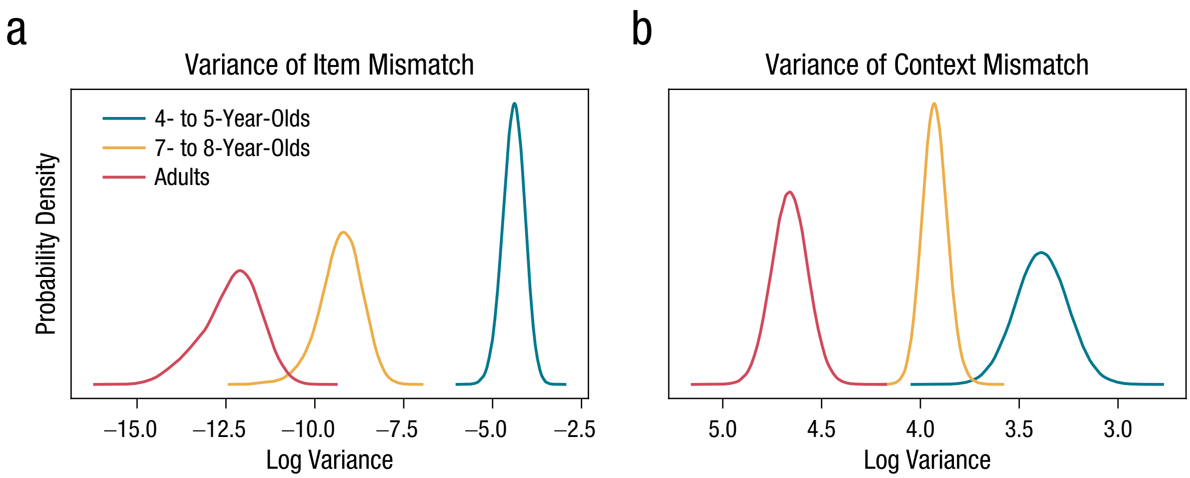

Fig. 8. Item-mismatch and context-mismatch parameters across age groups. Log variance of the item-mismatch parameter σ_{ti} is shown in (a); log variance of the context-mismatch parameter σ_{su} is shown in (b). Note that the parameters are log transformed for easier visual comparison.

Overall, interference decreased with age. The decrease occurred despite the fact that the number of memory traces across development increased in the model (i.e., the *n* and *m* constants), which will contribute to increasing memory interference (i.e., the cost of experience in memory). To understand the underlying mechanism of this change, we investigated the mismatch parameters that are responsible for the discrimination of different representations and can serve as a proxy for the benefit of experience in memory. Figure 8 shows the posterior distribution of the group means of two parameters—variance of the item-mismatch (σ_{ti}) and context-mismatch (σ_{su}) parameters (posterior distributions of the other model parameters are presented in section F in the Supplemental Material). We evaluated the difference between parameters by examining whether the two 95% HDI overlaps. A lower value of the mismatch parameters indicates better discrimination (the parameters in Fig. 8 are log transformed for easier visual comparison). First, we can see that the variance of the item-mismatch parameter, $\log(\sigma_{ti})$, decreases with development—4- to 5-year-olds: $M = -4.41$, 95% HDI = [−5.02, −3.81]; 7- to 8-year-olds: $M = -9.27$, 95% HDI = [−10.48, −8.14]; adults: $M = -12.36$, 95% HDI = [−14.09, −10.90] (see Fig. 8a). This indicates that items become more distinct (i.e., less confusable) with development, despite the fact that vocabulary size increases. The variance of the context mismatch, $\log(\sigma_{su})$, also decreases with development, indicating that different contexts also become more distinct (i.e., less confusable) with development—4- to 5-year-olds: $M = -3.39$, 95% HDI = [−3.65, −3.13]; 7- to 8-year-olds: $M = -3.93$, 95% HDI = [−4.05, −3.80]; adults: $M = -4.66$, 95% HDI = [−4.84, −4.48] (see Fig. 8b). Regarding the overall decrease in interference across development, the results imply that even though memory traces accumulate with age, the ability to distinguish different representations develops at a faster rate and overcomes the interference that these accumulated traces generate.

It is also notable that the developmental pattern of the context-mismatch parameter differs from that of the item-mismatch parameter. For the item-mismatch parameter, the value decreases more rapidly early in development, whereas for the context-mismatch parameter the decrease is more gradual and protracted across age. This asynchronous pattern, to an extent, underlies the drastic decrease in item noise between 4- to 5-year-olds and 7- to 8-year-olds, while having a constantly decreasing context noise across development (shown in Figs. 5 and 6).

General Discussion

The current study examined how different sources of interference in memory change across development and how prior experiences may contribute to these changes. We tested three different age groups using item- and source-recognition tasks with manipulations of list length, list strength, and word frequency, and we decomposed different sources of interference using a computational model. Results showed that interference stemming from other items during study (i.e., item noise) rapidly decreases between the ages of 4 to 5 years and 7 to 8 years, whereas interference stemming from other contexts outside of the experiment (i.e., context noise) gradually decreases through to adulthood and remains as the major source of interference in recognition memory. The model accounts for this change through an early development in the ability to discriminate items and a more gradual development in the ability to discriminate between different contexts.

In accordance with the fact that memory performance typically increases through development, our results show that the benefit of experience overrides the cost. The rapid decrease in item noise from the ages of 4 to 5 years to the ages of 7 to 8 years aligns well with previous literature suggesting that the representations of concepts are refined with learning and development (Keil, 1979; Mandler et al., 1991). Because words are used as items in the current study, better discrimination among different words can be explained by the extra exposure to language and formal education that the 7- to 8-year-olds start to receive (Hoff, 2014). The argument is additionally supported by the fact that the mean age of acquisition for the words used in the experiment was 4.59 years ($SD = 1.31$; Kuperman et al., 2012), which is slightly lower than the average age of the 4- to 5-year-old group (i.e., 4.82 years).

On the other hand, context noise, which is closely linked to the ability to discriminate between different contexts, decreases gradually over the course of development and remains a major source of interference in adults. This accords with other studies that have found that the ability to discriminate among different contexts has a protracted development (Ghetti & Lee, 2011). However, previous studies mainly examined how well different contexts can be discriminated in the study without considering the effects of prior experience that the participants bring to the experiment (Drummey & Newcombe, 2002; Lloyd et al., 2009; Yim et al., 2013). By contrast, the current study distinguished these two components by approximating the effect of prior experience and then using a computational model to estimate the ability to discriminate between different contexts.

The increase in the ability to discriminate among different representations may be driven by at least two mechanisms. First, evidence shows that discrimination may be enhanced by refining the representations through learning (Keil, 1979), which we term the *benefit of experience*. For example, previous studies show that children's memories are better than adults' when child-friendly items are used (Gross et al., 2016) or when children are tested in a domain in which they show expertise (Gobbo & Chi, 1986). However, it is also possible that the change can be driven by maturation. For example, *pattern separation*, or the ability to dissociate highly similar memory representations, has been linked to the maturation of the hippocampus (Keresztes et al., 2018), and the hippocampus develops through to adulthood (De Master et al., 2014). The two mechanisms are not mutually exclusive, and the division of labor between the two requires further investigation.

To capture how different sources of interference change across development, we used a computational model in the current study. Previous studies have been able to explain developmental changes with increases in discriminability (Hayes et al., 2017). However, studies are agnostic about the cognitive sources of development, such as whether changes are due to representations or to retrieval mechanisms. Our model decomposed performance changes to measure how latent parameters (i.e., discriminability of different memory representations) developed, while considering the increase in experience that comes with age. The results showing an asynchronous development of item and context discrimination in recognition memory suggest that there are at least two representational changes involved in recognition-memory development. Additionally, the current model explains memory development and the changes in representations that underlie development through a retrieval process based on familiarity, which suggests that familiarity may be an important construct in understanding human memory (cf. Geurten et al., 2021; Ghetti & Angelini, 2008; Hayes et al., 2017).

Finally, the current study provides a basis for interesting future investigations. First, although we proposed a computational model that assumes representational change across development, we did not describe how it changes. Because our results show that the process of making representations more distinguishable from each other is a crucial component in memory development, incorporating a mechanistic account of how different representations are refined in the model will provide a more complete understanding of how memory develops (e.g., Saxe et al., 2019). Second, the current study used a cross-sectional design in which the developmental changes cannot be directly measured but only inferred. We therefore are planning to confirm the current findings using a longitudinal design. Embedded in the longitudinal design, we also plan to measure other cognitive abilities (e.g., executive function, vocabulary knowledge) that may aid us to better understand the mechanisms involved in memory development. Future studies should also investigate other age groups to fully understand the developmental trajectory of human memory, as our current study examined only three age groups. Finally, improvements can be made to the measurement of previous exposure of an item rather than using normative frequency (e.g., TASA or CHILDES). For example, using experience-sampling methods, an individualized corpus can be constructed, and individualized frequency will provide better measures of previous exposure for each participant. Moreover, the recency of the item can be controlled, as it will also affect memory performance (cf. Yim et al., 2020).

In conclusion, the current study found evidence that different sources of interference change across development at a different rate and that this difference stems from the fact that the ability to discriminate between different kinds of memory representations (i.e., item and context) develop asynchronously. Whereas item-based interference (i.e., item noise) decreases rapidly between the ages of 5 and 7, context-based interference (i.e., context noise) decreases gradually and continues to affect memory in adults. These findings advance our understanding of memory development by identifying its sources. Moreover, the novel computational model provides a way to decompose the effects stemming from experience and from maturation, which can be further applied in cognitive development research more broadly.

Transparency

Action Editor: Thanujeni Pathman
Editor: Patricia J. Bauer
Author Contributions
All the authors designed the study. H. Yim implemented the experiment, managed data collection, and analyzed the data. A. F. Osth implemented the model. All the authors discussed the behavioral and model results, prepared the manuscript, and approved the final manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This research was supported by the research fund of Hanyang University (Grant No. HY-202100000001012), by a National Research Foundation of Korea grant funded by the Korean government to H. Yim (Ministry of Science and ICT Grant No. 2018R1A5A7059549), by the Australian Research Council's funding scheme to S. J. Dennis and V. M. Sloutsky (Grant No. DP150100272), by two National Institutes of Health grants to V. M. Sloutsky (Grant No. R01HD078545 and Grant No. P01HD080679), and by an ARC Award to A. F. Osth (Grant No. DE170100106).
Open Practices
All data and modeling code have been made publicly available via OSF and can be accessed at https://osf.io/2fg5w/. The design and analysis plans for the study were not preregistered. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

Hyungwook Yim https://orcid.org/0000-0002-9276-3544
Vladimir M. Sloutsky https://orcid.org/0000-0003-4054-3360

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211073131

Note

1. However, the magnitude of the frequency effect is not directly connected to the magnitude of the context noise. Because context noise is calculated by multiplying the variance of the context-mismatch parameter and the actual frequency count, the frequency difference between the high- and low-frequency words is the main driver of the magnitude of the frequency effect when the context noise is greater than zero.

References

Ackerman, B. P. (1984). Item-specific and relational encoding effects in children's recall and recognition memory for words. *Journal of Experimental Child Psychology*, *37*(3), 426–450.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123.

Bauer, P. J. (2007). *Remembering the times of our lives: Memory in infancy and beyond*. Erlbaum.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*(1), 37–60.

Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, *111*(3), 800–807.

Cycowicz, Y. M., Friedman, D., Snodgrass, J. G., & Duff, M. (2001). Recognition and source memory for pictures in children and adults. *Neuropsychologia*, *39*(3), 255–267.

De Master, D., Pathman, T., Lee, J. K., & Ghetti, S. (2014). Structural development of the hippocampus and episodic memory: Developmental differences along the anterior/posterior axis. *Cerebral Cortex*, *24*(11), 3036–3045.

Dennis, S. J., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478.

Dennis, S. J., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language*, *59*(3), 361–376.

Drummey, A. B., & Newcombe, N. S. (2002). Developmental changes in source memory. *Developmental Science*, *5*(4), 502–513.

Fox, J., Dennis, S., & Osth, A. F. (2020). Accounting for the build-up of proactive interference across lists in a list length paradigm reveals a dominance of item-noise in recognition memory. *Journal of Memory and Language*, *110*, Article 104065. https://doi.org/10.1016/j.jml.2019.104065

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694.

Geurten, M., Willems, S., & Lloyd, M. (2021). Too much familiarity! The developmental path of the fluency heuristic in children. *Child Development*, *92*(3), 919–936.

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development*, *79*(2), 339–358.

Ghetti, S., & Lee, J. (2011). Children's episodic memory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 365–373.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*(3), 546–567.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory*, *2*(1), 21–31.

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three implied regularities. *Psychonomic Bulletin & Review*, *16*(3), 431–455.

Gobbo, C., & Chi, M. T. H. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development*, *1*, 221–237.

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1355–1369.

Gross, J., Gardiner, B., & Hayne, H. (2016). Developmental reversals in recognition memory in children and adults. *Developmental Psychobiology*, *58*(1), 52–59. https://doi.org/10.1002/dev.21344

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes.

Hayes, B. K., Dunn, J. C., Joubert, A., & Taylor, R. (2017). Comparing single- and dual-process models of memory development. *Developmental Science*, *20*(6), Article e12469. https://doi.org/10.1111/desc.12469

Hoff, E. (2014). *Language development* (5th ed.). Wadsworth Cengage Learning.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Harvard University Press.

Keresztes, A., Ngo, C. T., Lindenberger, U., Werkle-Bergner, M., & Newcombe, N. S. (2018). Hippocampal maturation drives memory from generalization to specificity. *Trends in Cognitive Sciences*, *22*(8), 676–686.

Kinnell, A., & Dennis, S. (2011). The list length effect in recognition memory: An analysis of potential confounds. *Memory & Cognition*, *39*(2), 348–363.

Kinnell, A., & Dennis, S. J. (2012). The role of stimulus type in list length effects in recognition memory. *Memory & Cognition*, *40*(3), 311–325.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Suppl.).

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lee, J. K., Fandakova, Y., Johnson, E. G., Cohen, N. J., Bunge, S. A., & Ghetti, S. (2020). Changes in anterior and posterior hippocampus differentially predict item-space, item-time, and item-item memory improvement. *Developmental Cognitive Neuroscience*, *41*, Article 100741. https://doi.org/10.1016/j.dcn.2019.100741

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lloyd, M. E., Doydum, A. O., & Newcombe, N. S. (2009). Memory binding in early childhood: Evidence for a retrieval deficit. *Child Development*, *80*(5), 1321–1328.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Erlbaum.

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, *23*(2), 263–298.

Mandler, J. M., & Robinson, C. A. (1978). Developmental changes in picture recognition. *Journal of Experimental Child Psychology*, *26*(1), 122–136.

Murdock, B. B., & Kahana, M. J. (1993a). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 689–697.

Murdock, B. B., & Kahana, M. J. (1993b). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and Nobel (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1450–1453. https://doi.org/10.1037/0278-7393.19.6.1450

Ngo, C. T., Lin, Y., Newcombe, N. S., & Olson, I. R. (2019). Building up and wearing down episodic memory: Mnemonic discrimination and relational binding. *Journal of Experimental Psychology: General*, *148*(9), 1463–1479.

Osth, A. F., & Dennis, S. J. (2015). Sources of interference in item and associative recognition memory. *Psychological Review*, *122*(2), 260–311.

Osth, A. F., & Dennis, S. J. (2020). *Global matching models of recognition memory*. PsyArXiv. https://doi.org/10.31234/osf.io/mja6c

Osth, A. F., Dennis, S. J., & Heathcote, A. (2017). Likelihood ratio sequential sampling models of recognition memory. *Cognitive Psychology*, *92*, 101–126.

Osth, A. F., Dennis, S. J., & Kinnell, A. (2014). Stimulus type and the list strength paradigm. *Quarterly Journal of Experimental Psychology*, *67*(9), 1826–1841.

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. J. (2018). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language*, *103*, 91–113.

Osth, A. F., Jansson, A., Dennis, S., & Heathcote, A. (2018). Modeling the dynamics of recognition memory testing with a combined model of retrieval and decision making. *Cognitive Psychology*, *104*, 106–142. https://doi.org/10.1016/j.cogpsych.2018.04.002

Picard, L., Cousin, S., Guillery-Girard, B., Eustache, F., & Piolino, P. (2012). How do the different components of episodic memory develop? Role of executive functions and short-term feature-binding abilities. *Child Development*, *83*(3), 1037–1050.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 163–178.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2019). A mathematical theory of semantic development in deep neural networks. *Proceedings of the National Academy of Sciences, USA*, *166*(23), 11537–11546. https://doi.org/10.1073/pnas.1820226116

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368–384.

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

Yim, H., O'Brien, C. R., Stone, B., Osth, A., & Dennis, S. J. (2020). Using emails to quantify the impact of prior exposure on word recognition memory. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 185–190). Cognitive Science Society.

Yim, H., Osth, A. F., Sloutsky, V. M., & Dennis, S. J. (2018). Evidence for the use of three-way binding structures in associative and source recognition. *Journal of Memory and Language*, *100*, 89–97.