Cognitive Development 55 (2020) 100892

Contents lists available at ScienceDirect

Cognitive Development

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

Coding of featural information in visual working memory in 2.5-year-old toddlers

Chen Cheng[a,b,*], Zsuzsa Kaldy[a], Erik Blaser[a]

[a] University of Massachusetts Boston, Boston, MA, United States
[b] Boston University, Boston, MA, United States

ARTICLE INFO

Keywords:
Visual working memory
Capacity
Representational format
Delayed Match Retrieval
Infants

ABSTRACT

The number of objects that infants can remember in visual working memory (VWM) increases rapidly during the first few years of life (Kaldy & Leslie, 2005; Ross-Sheehy, Oakes, & Luck, 2003). However, less is understood about the *representational format* of VWM: whether storage is determined by fixed-precision memory slots, or the allocation of a limited continuous resource. In the current study, we adapted the Delayed Match Retrieval eye-tracking paradigm (Kaldy, Guillory, & Blaser, 2016), to test 2.5-year-old toddlers' ability to remember three object-location bindings when the to-be-remembered objects were all unique (Experiment 1) versus when they shared features such as color or shape (Experiment 2). 2.5-year-olds succeeded in Experiment 1, but only performed marginally better than chance in Experiment 2. Interestingly, when incorrect, participants in Experiment 2 were no more likely to select a decoy item that shared a feature with the target item. It seems that the increased similarity of to-be-remembered objects did not impair memory for the objects directly, but instead increased the likelihood of catastrophic forgetting.

1. Introduction

While the capacity of Visual Working Memory (VWM), the visual information being maintained in service of an ongoing task, has been shown to be quite limited both in adults and children, there are many questions currently still open about the format of representations in VWM. In the adult VWM literature, there are two main models. One is a 'slot' model that holds that object representations are discrete, and contain featural information about the objects, including their color, spatial orientation, shape, etc. Binding features to representations imposes no additional cost (Cowan, 2001, Anderson, Vogel, & Awh, 2013; Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997; Zhang & Luck, 2008). The other model is a 'resource' model where information is stored in a continuous fashion, and flexibly allocated to increase the precision of stored representations. Here, the number of remembered objects depends on their complexity; when complexity increases, the number of objects that can be accurately retained decreases (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Fougnie, Asplund, & Marois, 2010; Wilken & Ma, 2004).

1.1. The development of VWM capacity

The slots versus resources debate has received little attention from developmental scientists so far (but see Burnett Heyes, Zokaei, van der Staaij, Bays, & Husain (2012), and Zosh and Feigenson (2012), both supporting a resource-based model). Previous

---

[*] Corresponding author at: Boston University, Department of Psychological and Brain Sciences, 64 Cummington Mall, Boston, MA, 02215, United States.

E-mail address: ccheng10@bu.edu (C. Cheng).

https://doi.org/10.1016/j.cogdev.2020.100892
Received 7 September 2019; Received in revised form 4 May 2020; Accepted 9 May 2020

developmental studies have typically started from a slot-based perspective on capacity - *how many items* can infants and children remember? These studies demonstrated significant increases in infancy (Kaldy & Leslie, 2003, 2005; Kibbe & Leslie, 2013a; Oakes, Baumgartner, Barrett, Messenger, & Luck, 2013; Ross-Sheehy et al., 2003) and childhood (Cowan et al., 2005; Pailian, Libertus, Feigenson, & Halberda, 2016; Riggs, McTaggart, Simpson, & Freeman, 2006; Simmering, 2012). Infants as young as 6 months of age can remember one object (Kaldy & Leslie, 2005; Kibbe & Leslie, 2011; Ross-Sheehy et al., 2003) and can hold up to 3 items in VWM by the end of the first year (Feigenson & Carey, 2003; Kibbe & Leslie, 2013b; Ross-Sheehy et al., 2003). In early and middle childhood, VWM capacity also shows an increase from 3 to 10 years of age (Pailian et al., 2016; Riggs et al., 2006; Simmering, 2012). For example, Cowan et al. (2005) tested 7-year-old children and adults in a change detection paradigm and found that VWM capacity increases from approximately 3.5 items (in 7-year olds) to 5.7 items (in adults). However, less is known about the development of the ability to represent the *content* of remembered items (Cowan, 2017; Kibbe, 2015).

By using variants of violation-of-expectation paradigms, researchers investigated whether infants reacted with surprise to a sequence of events when an object has changed its appearance (color and/or shape) after occlusion. For instance, in a paradigm testing the ability to represent object-location bindings (‘*what was where*’), 6.5-month-olds could remember one object’s featural binding with its location (i.e., were surprised when one of two occluded objects was revealed with a different shape), and 9-month-olds could remember two object-location bindings (Kaldy & Leslie, 2003, 2005). While 6-month-olds failed to detect a featural change in the second of two sequentially hidden objects, they did look longer when that object disappeared completely, suggesting infants’ representation of an object can be persistent, and the ability to store featural information in that representation may develop later (Kibbe & Leslie, 2011). This developmental timeline is consistent with the results of an object tracking task that showed that 7.5-month-olds could use different featural information to identify two different objects (features were treated as cues to identify objects), while 5.5-month-olds could only use spatiotemporal information, but not features, to distinguish the two (Wilcox & Schweinle, 2002).

When infants are tested in working memory paradigms with a number of objects that exceeds their capacity limit, instead of remembering a subset, they tend to forget all the objects entirely (Barner, Thalwitz, Wood, Yang, & Carey, 2007; Feigenson & Carey, 2003, 2005). This phenomenon was called ‘catastrophic forgetting’. This term was first used in studies of learning in neural networks (French, 1999). It referred to the phenomenon that in certain situations, forgetting can be total and abrupt, and not gradual. In the developmental literature, researchers have examined the role of array heterogeneity in overcoming this phenomenon, showing that 13-month-olds could track up to three objects when remembering perceptually highly contrastive objects (facilitating discrimination of objects) compared to perceptually similar objects (Zosh & Feigenson, 2015). We will examine whether the different levels of featural discriminability have an effect on toddlers’ memory performance.

1.2. The development of VWM precision

Research in older children has found the precision with which featural differences are encoded and recalled develops over childhood (Burnett Heyes, Zokaei, & Husain, 2016; Burnett Heyes, Zokaei, van der Staaij, Bays, & Husain, 2012; Guillory, Gliga, & Kaldy, 2018; Simmering & Miller, 2016; Simmering & Patterson, 2012). Using a color-discrimination task, Simmering and Patterson (2012) showed a developmental increase in VWM precision, for color, from 4 to 6 years of age. Later, Simmering and Miller (2016) replicated this finding in children from 5 to 8 years of age, and showed a significant correlation between estimates of VWM capacity (as measured by change-detection task) and VWM precision. To quantify VWM precision over development, Burnett Heyes et al. (2012) implemented a delayed estimation task, where children had to reproduce a probed stimulus’ orientation from memory using a rotation dial. They found a linear age-related improvement in the precision of recalling the bar’s orientation from 7 to 13 years of age. Importantly, the improvement in VWM performance has been attributed to a specific decrease in the variability of stored feature representations, rather than a decrease in misbinding or random noise. Converging results were found in a younger population - 4- to 7-year-olds - using a similar delayed estimation task with cues to manipulate attention (Guillory et al., 2018). Together, this can be taken as evidence for resource-based models of VWM in these age groups.

However, slot versus resource models have not yet been tested in toddlers between 24 and 36 months of age, a period where most infant paradigms (that are based on looking behavior) become challenging (due to lapses in compliance), yet child paradigms (that are based on verbal instructions) are not yet appropriate, causing a ‘toddler gap’ (Kaldy et al., 2016), or the so-called ‘dark ages’ (Meltzoff, Gopnik, & Repacholi, 1999), but see Alp (1994) for an exception, where the Imitation Sorting Task was used to test working memory capacity in 12- to 36-month-olds). Perhaps due to the different task demands between infants’ and children’s studies, young children’s VWM performance in change-detection tasks was found to be even lower than infants’ (Cowan, AuBuchon, Gilchrist, Ricker, & Saults, 2011; Kaldy & Blaser, 2020; Simmering, 2012), rendering little evidence of VWM capacity development during the toddler gap. Our study is designed to help fill this gap.

1.3. The present study

To test toddlers’ VWM performance, we modified the Delayed Match Retrieval (DMR) paradigm (Kaldy et al., 2016) to test VWM for three object-location bindings. In the experiments, four face-down virtual ‘cards’ were presented. Three of them (one *Match* and two *Non-Match* cards) flipped face-up sequentially, then flipped back face-down. Next, the fourth card (the *Sample*) flipped face-up to reveal a match to one of the previously exposed cards (i.e., the *Match*). Toddlers were expected to make an anticipatory saccade to the (face-down) *Match* card in response to the exposure of the *Sample*. 10- and 13-month-olds could remember two object-location bindings in the DMR task (Cheng, Kaldy, & Blaser, 2019; Kaldy et al., 2016), and by the end of the second year, they could track and update two object-location bindings even if the objects moved during the face-down period (Cheng, Kaldy, & Blaser, 2019).

By adapting this paradigm, the current study aimed to fulfill two goals. The first was methodological: to demonstrate that the DMR paradigm can be scaled up to a three-item memory task and that toddlers between 24 and 38 months of age can succeed in this nonverbal eye-tracking VWM task (Experiment 1). The second goal was to investigate the representational format of toddlers' VWM; whether it is better described as slots or resources. To address this, we conducted an experiment (Experiment 2) where the *Match* card shared a feature (shape or color) with one of the other cards (designated the *MisMatch,* in contrast to the *NonMatch* which had no shared features). We hypothesized, consistent with a resource-based model of VWM, that toddlers, when incorrect, would mistakenly choose the *MisMatch* more often than the *NonMatch*.

In addition to these questions, since we exposed the three to-be-remembered objects serially, we analyzed effects of serial positions on VWM performance. Primacy and recency effects are classic hallmarks of memory encoding in both children and adults (1948, Broadbent & Broadbent, 1981; Ebbinghaus, 1913; Hagen & Kail, 1973; Rose, Feldman, & Jankowski, 2001), and here we will test whether we can observe them in our VWM paradigm.

2. Experiment 1: Delayed Match Retrieval with three unique cards

2.1. Method

2.1.1. Participants

Fifty-three healthy, full-term toddlers (19 girls) were recruited from the Greater Boston area and tested at the University of Massachusetts Boston. Forty-four were included in the final analysis. The average age was 30 months 9 days (SD = 4.14, range: 23 months 23 days - 38 months 9 days). To be included, each participant had to contribute at least 3 valid trials (out of 12), where a trial was considered valid if the participant fixated each of the three to-be-remembered cards during the *encoding phase*, and at least one of the (face-down) cards during the *response phase* (as captured by the Tobii fixation filter, see *Data Analysis*). We used a minimum of 3 trials as an inclusion criterion to err on the side of inclusivity, and to be consistent with the criteria used in our previous studies with the same paradigm (2019b, Cheng et al., 2019b); where this value was chosen as it was 2.5 SD away from the average number of valid trials). Nine additional toddlers were tested, but excluded for not meeting this criteria. Caregivers received a small gift and $20 compensation for their time and travel expenses. All caregivers gave informed consent before the experiment.

2.1.2. Apparatus and stimuli

Caregivers sat in a chair holding their toddlers in their laps in front of a Tobii T120 eye-tracker (Tobii Technology, Stockholm, Sweden) in a dimly lit testing room. Caregivers were asked to wear a visor to cover their eyes and not to interact with their toddlers during the study. The standard Tobii 5-point infant calibration was used. The experimental stimuli were virtual cards that could be shown face up, revealing an abstract, unfamiliar object (Fig. 1a), or face-down, occluding the object. We used a total of four different objects as card faces. To discourage verbal rehearsal, which has been previously shown in 18-month-olds who were capable of silently labeling simple, familiar objects (Mani & Plunkett, 2010), we used abstract, non-nameable geometric objects. On each trial, three different cards were assigned roles as the *Match* and the two *NonMatch* cards, while the *Sample* card had an identical face to the *Match*. Cards subtended 5 × 5 deg and were arranged symmetrically with their centers 5 deg from the center of the screen.

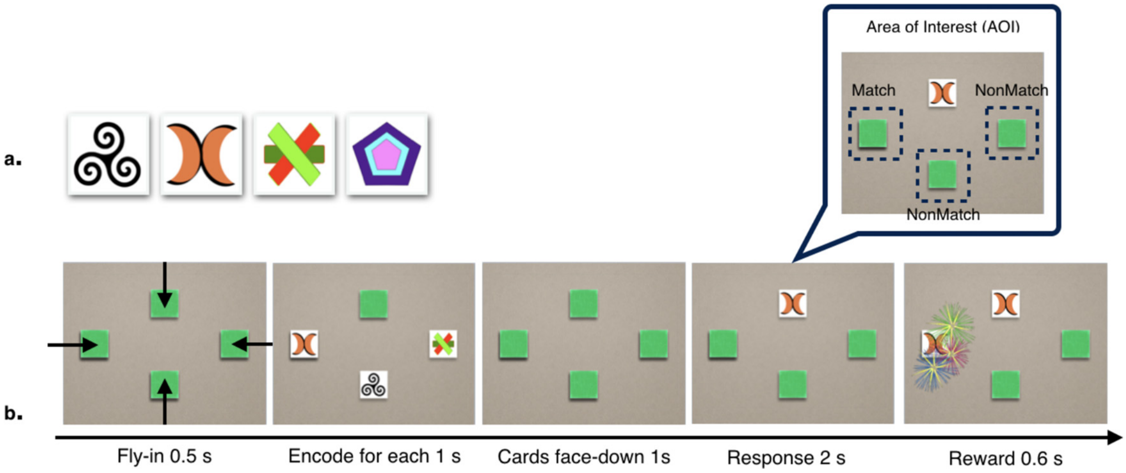

Fig. 1. Panel *a* shows the unique objects used in Experiment 1 and Panel *b* shows an example of the sequence of events in a trial. In the example trial shown in Panel *b*, the orange X-shape was the *Match*, and the black swirl and the green/red crossing bars shape were the two *NonMatch* cards. The inset box pointing toward the *response phase* shows the identities of each of the cards, and the size of the AOI's used for data analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

2.1.3. Design and procedure

In each test trial, four cards entered the screen, face-down. The three, to-be-remembered cards were then flipped face-up (an animation that took approximately 0.5 s) sequentially. After a card was exposed, 1 s was allowed to elapse, and then another card was exposed, until all three cards were face up. The flipping-over animation of each sequential exposure of a to-be-remembered card tends to draw infants' gaze. In this paradigm, with short exposure times and sequential flip-over card animations, infants' average looking time to each card is about the same (Cheng et al., 2019a). Following this *encoding phase*, all three cards turned face-down simultaneously. After 1 s, the fourth card, the *Sample*, flipped face-up. The face of this card was identical to one of the three previously exposed cards (i.e., the *Match*). The reveal of the *Sample* began a 2 s *response phase*, after which the *Match* card was flipped face-up, accompanied by a brief (0.6 s) reward animation at its location (e.g., a burst of fireworks) (see Fig. 1b). Sounds effects accompanied each event. For movie clips showing sample trials from both Experiment 1 and 2, see *Supplemental Material*.

Twelve test trials were presented, where both the identity and the spatial position of the *Match* card was quasi-randomized between the left, right, and lower middle positions (the *Sample* always occupied the upper middle location). A short attention-getter animation was presented between the test trials.

2.1.4. Data analysis

Each 5 × 5 deg card was surrounded by a 7 × 7 deg Area of Interest (AOI), as larger AOI's are recommended by Tobii to help make gaze measurements robust against modest calibration errors (*Tobii Pro Support*, 2016). Analyses of gaze data were done using custom MATLAB scripts. Fixations were identified using the (default) Tobii Fixation Filter. Gaze position for each eye was collected at 60 Hz, and averaged between the two eyes to reduce noise. Then, missing data is interpolated (if the gap is under 100 ms). Following this, velocity peaks, within a sliding temporal window (of 5 frames @ 60 Hz, i.e. 83.8 ms), are identified. If a peak exceeds a set threshold value (0.42 pixels/ms) it is recorded as a new fixation (but if the distance between two candidate fixations is less than 35 pixels (< 1 deg), they are merged). The duration of a fixation, then, is the elapsed time between peaks, and the position of the fixation is the median of the gaze coordinates during that interval (Tobii Pro Support, 2016).

We measured whether the *Match* or one of the *NonMatch* cards received the first look, or the longer look based on the total accumulated looking time within the 2 s *response phase*. If so, the trial was coded as correct, otherwise incorrect. If the toddler did not look at any of the three face-down cards during the *response phase*, the trial was considered not valid and was excluded from further analyses. For each participant, we then calculated VWM performance as the percent of correct trials for each of these measures. Both of these measures were selected based on previously established anticipatory eye tracking methods in infants (e.g. Addyman and Mareschal (2010); Hochmann, Mody, and Carey (2016)); Kaldy et al. (2016)).

2.2. Results

2.2.1. Overall performance

Participants had an average of 68 % (SD = 23 %) valid responses of all trials (8.2 trials out of 12 trials, SD = 2.76). Participants' average performance based on first looks was 44 % correct, which was significantly above the chance level of 33 % (two-tailed one-sample t-test: M = 44 %, SD = 19 %, t(43) = 3.77, $p = 0.0005$, $d = 1.15$) (Fig. 2a). Similar performance, 43 % correct, was found based on looking time (M = 43 %, SD = 20 %, t(43) = 3.11, $p = 0.0033$, $d = 0.95$) (Table 1). We did not find a significant age-related increase within the age range that we tested ($R^2 = 0.0081$, $p = 0.56$, based on first looks; $R^2 = 0.015$, $p = 0.43$, based on longer looks) (Fig. 2b).

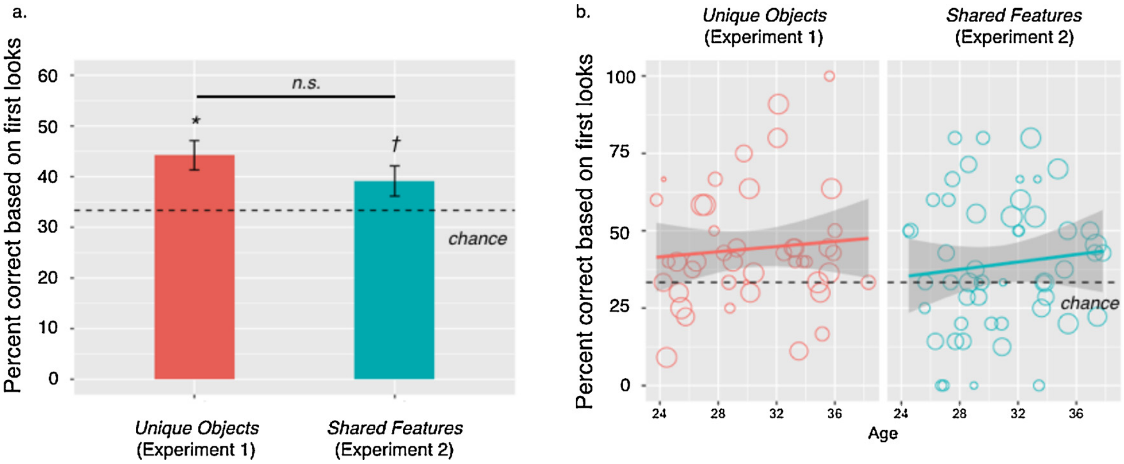

Fig. 2. Panel *a* shows the summary of average VWM performance based on first looks, for Experiments 1 and 2. Panel *b* shows individual participants' performance, based on first looks, as a function of age, for Experiments 1 and 2. The size of each circle corresponds to the number of valid trials for that participant. Solid lines indicate the fit from the linear regression model. Chance level, 33 %, is shown by the dashed black line. Error bars indicate standard error.

Table 1
Sample sizes, percent of valid trials and average percent preferences based on first looks and looking time in Experiments 1 and 2 (SD in parentheses). In reporting results of Experiment 1, NonMatch1 and NonMatch2 labels were coded arbitrarily. Results of one-sample t-tests for percent correct performance based on first look against chance level (33 %) are shown. NonM - NonMatch, MisM - MisMatch.

	N	Valid trials (%)	Preference based on first looks (%)				Preference based on longer looks (%)			
			Correct (Match)	Incorrect (NonM1)	Incorrect (NonM2)	t-test	Correct (Match)	Incorrect (NonM1)	Incorrect (NonM2)	t-test
Exp. 1 Unique objects	44	68 % (23)	44 % (19)	29 % (19)	27 %(19)	p = .0005 d = 1.15	43 % (20)	28 % (17)	29 % (20)	p = .003 d = 0.95
			Correct (Match)	Incorrect (MisM)	Incorrect (NonM)		Correct (Match)	Incorrect (MisM)	Incorrect (NonM)	
Exp. 2. Shared feature	52	60 % (21)	39 % (21)	31 % (19)	29 % (17)	p = .057 d = 0.55	40 % (21)	31 % (19)	28 % (16)	p = .027 d = 0.64

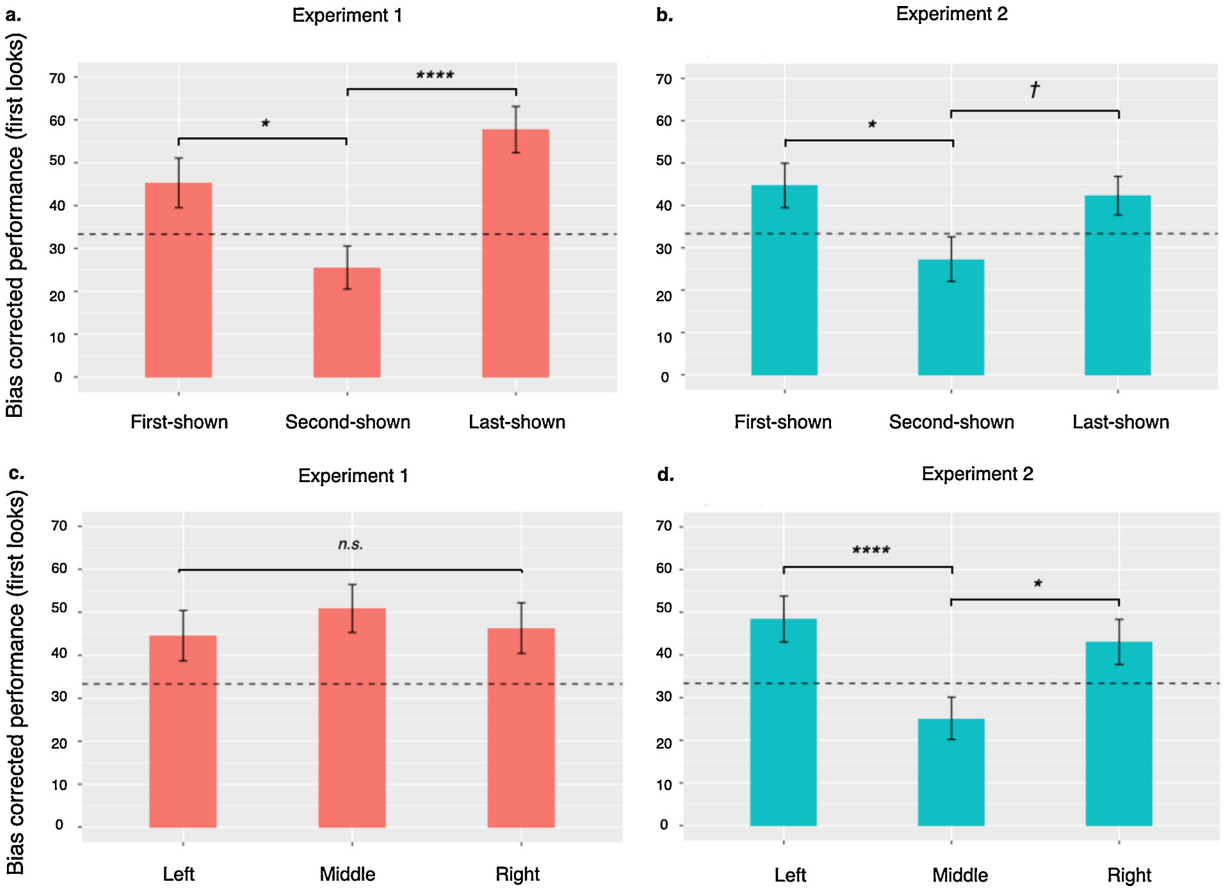

Fig. 3. Average bias-corrected performance as a function of serial position and spatial location. The upper two panels show the average bias-corrected performance when the *Match* was first-shown, second-shown or last-shown in Experiment 1 (Panel *a*) and Experiment 2 (Panel *b*). The lower two panels show the average bias-corrected performance when the *Match* was in the left, middle, or right position in Experiment 1 (Panel *c*) and Experiment 2 (Panel *d*). Chance level, 33 %, is shown by the dashed black line. Error bars indicate standard error. ****$p < .0001$, *$p < .05$, † $p < 0.1$.

2.2.2. *Serial order effect*

Since the *Match* card could have been the first, second, or last card to be exposed, we examined whether the serial order of the presentation of the cards played a role in toddlers' VWM performance. We adopted the "bias-correction" method used in previous studies using a similar task design (Atkinson, Hansen, & Bernnach, 1964; Donaldson & Strang, 1969; Keely, 1971; Siegel, Allik, & Herman, 1976) to account for both the hit rate and the overall response rate at a given position. The bias-corrected performance was calculated as the number of times a position was chosen and was correct, divided by the number of times a position was chosen. Results based on first looks showed that there was a significant effect of serial position on the bias-corrected performance (repeated ANOVA: $F(2,74) = 6.12$, $p = 0.003$, see Fig. 3a). Post hoc t-tests showed a primacy effect, with significantly better performance when the *Match* was first shown (M = 45 %) as opposed to when it was the second shown (M = 27 %, $t(38) = 2.13$, $p = 0.039$, $d = 0.69$). We also found a recency effect, with better performance when the *Match* was last shown (M = 58 %), versus second shown ($t(41) = 4.42$, $p < 0.0001$, $d = 1.38$). (There was no difference between when the *Match* was first shown versus last shown ($t(38) = 1.29$, $p = 0.21$, $d = 0.42$)).

2.2.3. *Spatial location effect*

Next, we examined whether the exposed location (left, lower middle, or right) affected toddlers' VWM performance. We first analyzed participants' response rate to determine whether participants had response biases to one spatial location than the other. We compared the percent of valid responses to each spatial location (left, middle, right), regardless of the responses being correct or incorrect. Results based on first looks showed that there was no significant effect of spatial locations on response rates (repeated ANOVA analysis: $F(2, 43) = 0.62$, $p = 0.54$). We then compared participants' task performance in trials where the *Match* was shown in the left, middle, or right location using the same bias-correction as above. Results showed that participants' performance was not different across the three locations (repeated ANOVA: $F(2, 54) = 0.32$, $p = 0.73$) (see Fig. 3c).

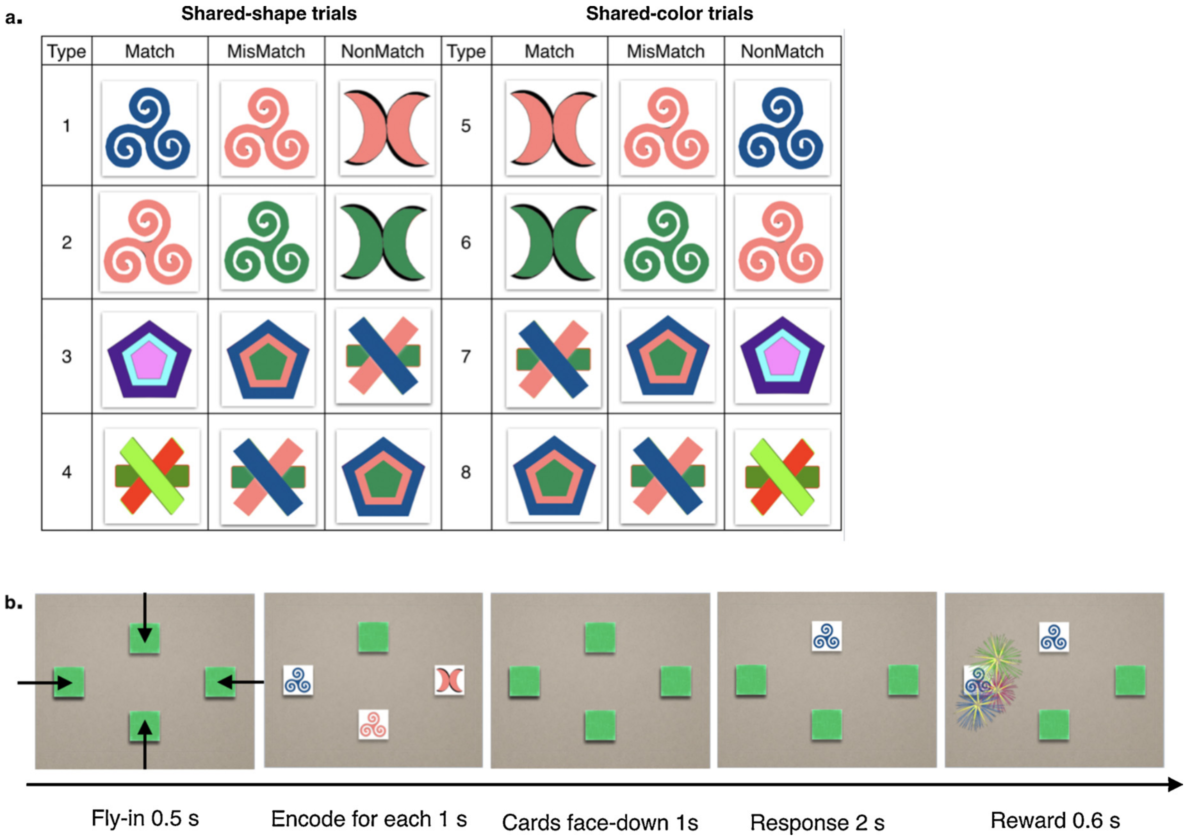

Fig. 4. Panel *a* shows the sets of stimuli used in Experiment 2 in the different trial types. The three objects shared features in the following way: the *MisMatch* shared either the same shape (Trial-type 1-4) or color (Trial-type 5-8) with the *Match*, while the *NonMatch* shared either shape or color with the *MisMatch*, but not with the *Match*. Panel *b* shows an example of a test trial in Experiment 2 (with Trial-type 1 stimuli).

3. Experiment 2: Delayed Match Retrieval with three shared-features cards

The results of Experiment 1 demonstrated above-chance performance in 2.5-year-old toddlers when tested with three object-location bindings with unique features. In the next experiment, we tested the precision of VWM representations by examining whether toddlers could remember objects that shared a visual feature (either shape or color).

3.1. Method

3.1.1. Participants

A new group of fifty-nine healthy, full-term toddlers (22 girls) were recruited from the Greater Boston area and tested at the University of Massachusetts Boston. Fifty-two were included in the final analysis. The average age of the group was 30 months 21 days (SD = 3.57, age range: 24 months 15 days - 37 months 24 days). Seven toddlers were tested but were excluded due to an insufficient number of valid trials (fewer than three, trial inclusion criteria as defined in Experiment 1). Participants were recruited in the same way as in Experiment 1.

3.1.2. Apparatus and stimuli

The same apparatus was used as in Experiment 1. The stimuli used here were similar to Experiment 1, except for the two non-matching cards, which were designed in a different way. One of the two was designated the *MisMatch*, and shared either shape or color (depending on the trial) with the *Match,* while the other was designated the *NonMatch*, and was distinct from the *Match* (see Fig. 4a).

3.1.3. Design and procedure

The design and procedure were the same as in Experiment 1 (see Fig. 4b). All trial types were presented in a mixed block and in a pseudo-randomized order.

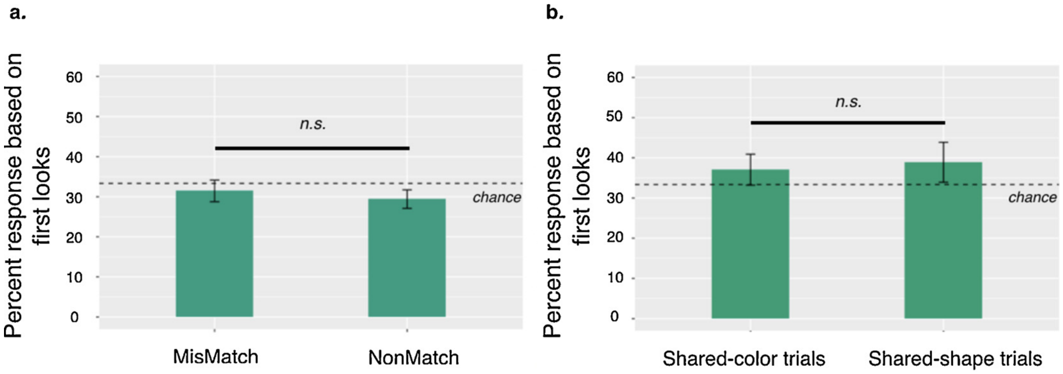

Fig. 5. Panel *a* shows an analysis of errors (incorrect trials only) in Experiment 2. The difference between the average percent responses to the *MisMatch* vs. *NonMatch* based on first looks was not significant (Section 3.2.2). Panel *b* shows that the average percent correct performance (all trials, first looks at the *Match)* in shared-color vs. shared-shape trials was not significantly different (Section 3.2.3). Chance level, 33 %, is shown by the dashed black line.

3.2. Results

3.2.1. Overall performance

The final analysis included 11 test trials. One trial contained incorrect stimuli due to a design error and had to be discarded. As there were redundant examples of each trial type within a block, this would not affect performance or analyses. Participants, on average, had valid responses in 60 % (SD = 21 %) of all trials (6.6 trials out of 11 trials, SD = 2.31). There was no significant difference between the percent of valid trials in Experiment 1 and Experiment 2 (two-sample t-test: t(94) = 1.79, $p = 0.077$).

Participants' average performance was 39 % correct (SD = 21 %) based on first looks, which was only marginally significant different from chance, however, the difference showed a medium effect size (two-tailed one-sample t-test: t(51) = 1.95, $p = 0.057$, $d = 0.55$) (Fig. 2a). Performance based on longer looks was significantly above chance (M = 40 %, SD = 21 %, t(51) = 2.28, $p = 0.027$, $d = 0.64$). Similarly to Experiment 1, we did not find a significant age-related increase within the age range that we tested ($R^2 = 0.0104$, $p = 0.47$ based on first looks; $R^2 = 0.0171$, $p = 0.36$ based on longer looks) (see Fig. 2b).

3.2.2. The effect of similarity on VWM performance

In Experiment 2, given the similarity of the to-be-encoded objects, if participants did not have a precise representation, when incorrect, the 'non-match' card that was more similar to the *Match* card (i.e., the *MisMatch*) may be mistakenly selected more often than the *NonMatch*. We investigated the pattern of incorrect responses by comparing participants' first looks to *MisMatch* and *NonMatch* cards during the *response phase* on incorrect trials (when they did not look at the *Match* card first). Results showed that participants' percentage of responses to the *MisMatch* (M = 31 %, SD = 19 %) versus the *NonMatch* (M = 29 %, SD = 17 %) were not significantly different from each other (paired t-test: t(51) = 0.50, $p = 0.62$, $d = 0.14$) (Fig. 5a). Participants' responses based on longer looks to the *MisMatch* (M = 31 %, SD = 19 %) and the *NonMatch* (M = 28 %, SD = 16 %) showed similar results (t (51) = 0.72, $p = 0.48$, $d = 0.20$). In short, the *MisMatch*, in spite of sharing a feature with the Match, was not the more attractive decoy.

3.2.3. The effect of features (shape versus color) on VWM performance

Given that objects in Experiment 2 are colored shapes, we next examined whether participants had better memory for one feature dimension (i.e., shape or color) over the other. Note that in half of the trials, the *MisMatch* shared the same shape with the *Match* (see Fig. 3a, Trial types 1–4) (shared-shape trials), while in the other half of the trials, the *MisMatch* shared the same color with the *Match* (shared-color trials; see Fig. 3a, Trial types 5–8). For instance, if participants remembered the shape of each to-be-remembered card but not their color, then they would have a better performance in shared-color trials, in which shape was the diagnostic feature to discriminate the *Match* from the *MisMatch* and *NonMatch* cards. Alternatively, if participants were better at remembering the color of each to-be-remembered card, rather than their shape, they would perform better in shared-shape trials, where color was the diagnostic feature to discriminate the *Match* from the *MisMatch* and *NonMatch*. We compared performance in shared-shape versus shared-color trials across all participants. Results showed that performance neither in Shape trials (M = 39 % (SD = 0.36), t(51) = 1.12, $p = 0.27$, $d = 0.31$ based on first looks; M = 41 % (SD = 0.38), t(51) = 1.19, $p = 0.24$, $d = 0.33$ based on longer looks), nor Color trials (M = 38 % (SD = 0.28), t(50) = 1.11, $p = 0.27$, $d = 0.31$ based on first look; M = 38 % (SD = 0.30), t(50) = 1.55, $p = 0.13$, $d = 0.44$ based on longer looks) was significantly above chance, and performance did not differ significantly between the two trial types (paired t-test: t(50) = 0.29, $p = 0.77$, $d = 0.03$ based on first looks; t(50) = 0.50, $p = 0.62$, $d = 0.14$ based on longer looks) (see Fig. 5b).

3.2.4. Serial order effect

We next examined the effect of the serial order (first, second, and last) of the *Match* card during the *encoding phase* and compared participants' bias-corrected performance using the same method as in Experiment 1. The trend across three serial positions was similar to Experiment 1, but was not significant (repeated ANOVA: F(2, 64) = 2.84, $p =$ 0.066) (see Fig. 3b). As in Experiment 1, we performed post hoc t-tests. These showed that participants performed significantly better when the *Match* was first shown (M = 42 %) than when it was second shown (M = 24 %) (t(35) = 2.38, $p =$ 0.023, $d =$ 0.80), and slightly better (though not significantly) when the *Match* was last shown (M = 43 %) versus when it was the second shown (t(38) = 1.88, $p =$ 0.068, $d =$ 0.61). (As in Experiment 1, there was no difference between when the *Match* was first shown versus last shown (t(41) = 0.60, $p =$ 0.55, $d =$ 0.19).

3.2.5. Spatial location effect

We also examined the effect of the spatial location (left, middle, and right) of the *Match* card during the *encoding phase* and compared participants' bias-corrected performance using the same method as in Experiment 1. The analysis showed that there were no significant differences in response rates across the three spatial locations (repeated ANOVA: F(2,51) = 0.77, $p =$ 0.47). Results of bias-corrected performance in each spatial location showed a different pattern than in Experiment 1: participants' performance was significantly different across the three spatial locations (repeated ANOVA: F(2,46) = 11.27, $p =$ 0.0001) (see Fig. 3d). Post-hoc t-tests showed participants performed significantly better when the *Match* was on the left (M = 53 %) than when it was in the middle (M = 24 %) (t(32) = 3.89, $p =$ 0.0005, $d =$ 1.38), and significantly better when the *Match* was on the right (M = 38 %) than when it was in the middle (t(34) = 2.33, $p =$ 0.026, $d =$ 0.80), but not different between the trials when the *Match* was on the left versus right (t(28) = 0.155, $p = 0.88$, $d =$ 0.06). Overall, these results show that participants performed better when the *Match* was at the end positions of the configuration (left or right) compared to the middle location.

4. General discussion

The current study investigated the representational format of Visual Working Memory (VWM) during the third year of life. We tested toddlers' VWM when to-be-remembered objects were unique (Experiment 1), and when the objects shared features (either shape or color) (Experiment 2). In our Delayed Match Retrieval (DMR) task, toddlers must remember visual information about each object *and* bind it to the respective location. Previous studies using the DMR task have studied infants' VWM development in the first year (8- and 10-month-olds, Kaldy et al. (2016)), and in the second year (13-month-olds (Cheng et al., 2019a), 20- and 25-month-olds (Cheng et al., 2019b)) of life. Our current study shows that the paradigm can be used without major modifications in 2.5-year-olds as well, and to study the development of VWM throughout the first three years of life, from infancy all the way into the 'toddler gap'. This feature is especially useful for future longitudinal studies that are aimed at examining individual differences in VWM capacity development in the early years, and the effects of other cognitive factors on VWM (attention, language, etc.).

In Experiment 1, we found above-chance performance in 2.5-year-old toddlers when tested with three object-location bindings with unique stimuli in a rapid dynamic VWM task: toddlers made their first look, during the response phase, to the *Match* card significantly more often than the other two non-matching cards, and looked at it significantly longer. In Experiment 2, one of the non-matching cards (the *MisMatch*) shared either shape or color with the *Match*. Here, results were mixed. Toddlers showed a pattern of success, but this was only significant when based on the dependent variable of looking time, not first looks. The observed effect sizes in the two studies were markedly different: in Experiment 1, $d =$ 1.15 and 0.95, based on first looks and longer looks, respectively, while in Experiment 2, $d =$ 0.55 and 0.64, based on first looks and longer looks, respectively. This suggests that the introduction of shared features made the task more difficult for our 2.5-year-old participants. Further manipulating the perceptual discriminability of to-be-remembered items would be a rich area for future studies.

4.1. Serial order and spatial biases

In both experiments, we presented our to-be-remembered stimuli in a serial order and we found primacy and recency effects in both of them. These effects are classic hallmarks of serial recall performance in both children and adults (Broadbent & Broadbent, 1981; Ebbinghaus, 1913; Hagen & Kail, 1973; Rose et al., 2001), and the fact that we found it here shows that information encoding follows the same general principles in our paradigm. We also analyzed the effect of spatial position of the *Match* on performance. Within our 3-card configuration, the end positions have the advantage of having only one neighbor, while the middle position has two. Participants' bias-corrected performance in Experiment 1 was not different across the three locations, while in Experiment 2, toddlers performed better when the *Match* was presented on the left or the right compared to the middle location. The same effect was found by Siegel et al. (1976): 5−7-year-old children remembered the target better when it was at either end of a linear configuration of 8 items, where they only have other stimuli in one side, providing a unique spatial cue for remembering these items, compared to the middle locations.

4.2. The organization of VWM representations

Why would it be harder for toddlers to remember three object-location bindings when the stimuli are similar than when they are unique? We will examine two possibilities. A fixed slot model proposes that once an object is encoded into VWM, all the information (shape, color, size, orientation, etc.) that constitutes the object is saved (and forgotten) together (Luck & Vogel, 1997). A resource model, on the other hand, argues that VWM is best conceptualized as a flexible resource - as objects get more complex, or as memory degrades, the precision the representation decreases (Alvarez & Cavanagh, 2004).

While our results do not definitively support either model, the smaller effect size of the performance in Experiment 2, where shared-feature objects were introduced, as compared to larger effect in Experiment 1 with unique objects are more consistent with a resource model. That said, our results do not fully support the predictions of this model. A resource model predicts that the *MisMatch* should be chosen more often than the *NonMatch* when toddlers are incorrect (i.e., any residual memory for color (or shape) should render the *Mismatch* a more compelling decoy, since it shares a feature with the *Match*) but we did not see a difference between the responses to the *MisMatch* and the *NonMatch*. On the other hand, a slot model may also predict the results in Experiment 2. Note that not only the *Match* and *MisMatch* cards shared a perceptual feature, the *MisMatch* and *NonMatch* cards also shared a feature along the other dimension. As shown in Fig. 4a, the *MisMatch* shared a feature (e.g., color) with the *Match*, and the *NonMatch* shared a different feature (e.g., shape) with the *MisMatch*. This was designed to prevent participants from only developing a strategy of only encoding a subset of the objects during the encoding period. Having a perceptually 'unique' *NonMatch* (with no shared feature) could result in biased encoding such that 1) participants might only pay attention to the *NonMatch* (the odd one out); or 2) participants might only encode the *Match* and the *MisMatch* (as a result of only getting the reward feedback from the two similar cards). It is likely that participants encoded the objects as discrete items. However, the overall similarity of the three objects with shared features may increase the background noise, resulting in decreased memory performance. Indeed, a dynamic model has been recently proposed by Johnson, Simmering, and Buss (2014)), where the same computational resource can be distributed to process both discrete and continuous representations. Empirical evidence supports the view that estimates of VWM capacity (as measured by a change detection task) and resolution (as measured by a color discrimination task) were correlated with each other and increase over development (Simmering & Miller, 2016).

It seems, especially in Experiment 2, that when participants were incorrect, they were more likely to be *catastrophically* incorrect, which may explain both the slight reduction in overall performance and the lack of bias between *MisMatch* versus *NonMatch*. "Catastrophic forgetting" emerged from the learning literature (see e.g. French, 1999), referring to the phenomenon that in certain situations, forgetting seems not gradual, but abrupt. In the infant literature, a number of studies have demonstrated this phenomenon (Barner et al., 2007; Feigenson & Carey, 2003, 2005): 12 – 20-month-old infants could remember objects up to their working memory capacity of 3, but when faced with tasks involving 4 objects, they responded as if they had forgotten all of them. (In similar situations, adults typically would still be able to recall a subset of the items, so with increasing set sizes, their performance declines in a gradual way.) Crucially, in those infant studies that showed catastrophic forgetting, the objects looked identical (e.g. four red balls). Zosh and Feigenson showed that in the same paradigm, if instead they used 4 objects that were perceptually highly discriminable ("array heterogeneity"), 13-month-olds could overcome catastrophic forgetting, and be successful in memory tasks with up to 4 objects (Zosh & Feigenson, 2015).

In our study, the objects we used in Experiment 2 were less perceptually discriminable than in Experiment 1, which meant that catastrophic forgetting could have occurred more often. This catastrophic forgetting phenomenon may also explain the lack of a difference in performance between Shape and Color trials of Experiment 2, despite earlier studies demonstrating a shape bias in infant memory (remembering the shape of an object better than some other perceptual features, e.g. Kaldy & Blaser, 2009; Vlach, 2016).

In sum, we have demonstrated that 2.5-year-olds performed above chance when remembering three object-location bindings when the stimuli were unique objects, and their memory performance was similar, but with a smaller effect size when the to-be-remembered objects shared features. Future studies are needed to investigate the role of feature similarity in toddlers' catastrophic forgetting that we observed in Experiment 2, and its implications for VWM representations in early development.

Author note

This research was supported by National Institutes of Health Grant R15HD086658 awarded to ZK and EB and a Dissertation Grant from the University of Massachusetts Boston to CC. Preliminary results from this study were presented at the Biennial Meeting of the Cognitive Development Society in October 2017 (Cheng, Kaldy, & Blaser, 2017). We would like to thank the caregivers and their children who participated in our studies. We would also like to thank Sangya Dhungana and other members of the UMass Boston Baby Lab for their help with data collection.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version (please see: https://osf.io/yzxqw/), at doi:https://doi.org/10.1016/j.cogdev.2020.100892.

References

Addyman, C., & Mareschal, D. (2010). The perceptual origins of the abstract same/different concept in human infants. *Animal Cognition, 13*(6), 817–833. https://doi.org/10.1007/s10071-010-0330-0.

Alp, I. E. (1994). Measuring the size of working memory in very young children: The imitation sorting task. *International Journal of Behavioral Development, 17*(1), 125–141. https://doi.org/10.1177/016502549401700108.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x.

Anderson, D. E., Vogel, E. K., & Awh, E. (2013). A common discrete resource for visual working memory and visual search. *Psychological Science, 24*(6), 929–938. https://doi.org/10.1177/0956797612464380.

Atkinson, R. C., Hansen, D. N., & Bernnach, H. A. (1964). Short-term memory with young children. *Psychonomic Science, 1*(1), 255–256. https://doi.org/10.3758/BF03342896.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*(7), 622–628.

https://doi.org/10.1111/j.1467-9280.2007.01949.x.

Barner, D., Thalwitz, D., Wood, J., Yang, S.-J., & Carey, S. (2007). On the relation between the acquisition of singular-plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science, 10*(3), 365–373. https://doi.org/10.1111/j.1467-7687.2007.00591.x.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851–854. https://doi.org/10.1126/science.1158023.

Broadbent, D. E., & Broadbent, M. H. P. (1981). Recency effects in visual memory. *The Quarterly Journal of Experimental Psychology Section A, 33*(1), 1–15. https://doi.org/10.1080/14640748108400762.

Burnett Heyes, S., Zokaei, N., & Husain, M. (2016). Longitudinal development of visual working memory precision in childhood and early adolescence. *Cognitive Development, 39*, 36–44. https://doi.org/10.1016/j.cogdev.2016.03.004.

Burnett Heyes, S., Zokaei, N., van der Staaij, I., Bays, P. M., & Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science, 15*(4), 528–539. https://doi.org/10.1111/j.1467-7687.2012.01148.x.

Cheng, C., Kaldy, Z., & Blaser, E. (2017). *The role of similarity in toddlers' Visual working Memory representations.* Poster presented at the Biennial Meeting of the Cognitive Development Society, Portland, OR.

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*, 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*, 104649. https://doi.org/10.1016/j.jecp.2019.06.002.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24*(1), 87–114. discussion 114–185 https://www.ncbi.nlm.nih.gov/pubmed/11515286.

Cowan, N. (2017). Mental objects in working memory: Development of basic capacity or of cognitive completion? In J. B. Benson (Vol. Ed.), *Advances in child development and behavior: Vol. 52*, (pp. 81–104). JAI. https://doi.org/10.1016/bs.acdb.2016.12.001.

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science, 14*(5), 1066–1074. https://doi.org/10.1111/j.1467-7687.2011.01060.x.

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001.

Donaldson, M., & Strang, H. (1969). Primacy effect in short-term memory in young children. *Psychonomic Science, 16*(2), 59–60. https://doi.org/10.3758/BF03336616.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* Teachers College Presshttps://doi.org/10.1037/10011-000.

Ebbinghaus, H. (1948). Concerning memory, 1885. In W. Dennis (Ed.). *Readings in the history of psychology* (pp. 304–313). Appleton-Century-Crofts. https://doi.org/10.1037/11304-034.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*(5), 568–584. https://doi.org/10.1111/1467-7687.00313/full.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(3), 295–313. https://doi.org/10.1016/j.cognition.2004.09.010.

Fougnie, D., Asplund, C. L., & Marois, R. (2010). Visual working memory capacity can be assessed independent of comparison errors. *Journal of Vision, 9*(8), https://doi.org/10.1167/9.8.591 591–591.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences, 3*(4), 128–135. https://doi.org/10.1016/s1364-6613(99)01294-2.

Guillory, S. B., Gliga, T., & Kaldy, Z. (2018). Quantifying attentional effects on the fidelity and biases of visual working memory in young children. *Journal of Experimental Child Psychology, 167*, 146–161. https://doi.org/10.1016/j.jecp.2017.10.005.

Hagen, J. W., & Kail, R. V. (1973). Facilitation and distraction in short-term memory. *Child Development, 44*(4), 831–836. https://doi.org/10.2307/1127733.

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.

Johnson, J. S., Simmering, V. R., & Buss, A. T. (2014). Beyond slots and resources: Grounding cognitive concepts in neural dynamics. *Attention, Perception & Psychophysics, 76*(6), 1630–1654. https://doi.org/10.3758/s13414-013-0596-9.

Kaldy, Z., & Blaser, E. (2009). How to compare apples and oranges: Infants' object identification tested with equally salient shape, luminance and color changes. *Infancy, 14*(2), 222–243. https://doi.org/10.1080/15250000802707088.

Kaldy, Z., & Blaser, E. (2020). Putting effort into infant cognition. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721420903015 0963721420903015.

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.

Kaldy, Z., Guillory, S. B., & Blaser, E. (2016). Delayed Match Retrieval: A novel anticipation-based visual working memory paradigm. *Developmental Science*. http://onlinelibrary.wiley.com/doi/10.1111/desc.12335/pdf.

Keely, K. (1971). Age and task effects in short-term memory of children. *Perception & Psychophysics, 9*(6), 480–482. https://doi.org/10.3758/BF03208958.

Kibbe, M. M. (2015). Varieties of visual working memory representation in infancy and beyond. *Current Directions in Psychological Science, 24*(6), 433–439. https://doi.org/10.1177/0963721415605831.

Kibbe, M. M., & Leslie, A. M. (2011). What do infants remember when they forget? Location and identity in 6-month-olds' memory for objects. *Psychological Science, 22*(12), 1500–1505. https://doi.org/10.1177/0956797611420165.

Kibbe, M. M., & Leslie, A. M. (2013a). What's the object of object working memory in infancy? Unraveling "what"and "how many.". *Cognitive Psychology, 66*(4), 380–404. https://www.sciencedirect.com/science/article/pii/S0010028513000224.

Kibbe, M. M., & Leslie, A. M. (2013b). 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.

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

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science, 21*(7), 908–913. https://doi.org/10.1177/0956797610373371.

Meltzoff, A. N., Gopnik, A., & Repacholi, B. M. (1999). *Toddlers' understanding of intentions, desires and emotions: Explorations of the dark ages.* https://psycnet.apa.org/record/1999-02439-001.

Oakes, L. M., Baumgartner, H. A., Barrett, F. S., Messenger, I. M., & Luck, S. J. (2013). Developmental changes in visual short-term memory in infancy: Evidence from eye-tracking. *Frontiers in Psychology, 4*, 697. https://doi.org/10.3389/fpsyg.2013.00697.

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.

Riggs, K. J., McTaggart, J., Simpson, A., & Freeman, R. P. J. (2006). Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology, 95*(1), 18–26. https://doi.org/10.1016/j.jecp.2006.03.009.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology, 37*(4), 539–549. https://doi.org/10.1037//0012-1649.37.4.539.

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/full.

Siegel, A. W., Allik, J. P., & Herman, J. F. (1976). The Primacy Effect in Young Children: Verbal Fact or Spatial Artifact? *Child Development, 47*(1), 242–247. https://doi.

org/10.2307/1128306.

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.

Simmering, V. R., & Miller, H. E. (2016). Developmental improvements in the resolution and capacity of visual working memory share a common source. *Attention, Perception & Psychophysics, 78*(6), 1538–1555. https://doi.org/10.3758/s13414-016-1163-y.

Simmering, V. R., & Patterson, R. (2012). Models provide specificity: Testing a proposed mechanism of visual working memory capacity development. *Cognitive Development, 27*(4), 419–439. https://doi.org/10.1016/j.cogdev.2012.08.001.

Tobii Pro Support. (2016). https://www.tobiipro.com/learn-and-support/support/.

Vlach, H. A. (2016). How we categorize objects is related to how we remember them: The shape bias as a memory bias. *Journal of Experimental Child Psychology, 152*, 12–30. https://doi.org/10.1016/j.jecp.2016.06.013.

Wilcox, T., & Schweinle, A. (2002). Object individuation and event mapping: Developmental changes in infants' use of featural information. *Developmental Science, 5*(1), 132–150. https://doi.org/10.1111/1467-7687.00217.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12), 1120–1135. https://doi.org/10.1167/4.12.11.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235. https://doi.org/10.1038/nature06860.

Zosh, J. M., & Feigenson, L. (2012). Memory load affects object individuation in 18-month-old infants. *Journal of Experimental Child Psychology, 113*(3), 322–336. https://doi.org/10.1016/j.jecp.2012.07.005.

Zosh, J. M., & Feigenson, L. (2015). Array heterogeneity prevents catastrophic forgetting in infants. *Cognition, 136*, 365–380. https://doi.org/10.1016/j.cognition.2014.11.042.