Title

Attention to novelty interferes with toddlers emerging memory decision-making.

Journal

Child Development, 95(1)

Authors

Leckey, Sarah
Bhagath, Shefali
Johnson, Elliott
et al.

Publication Date

2024

DOI

10.1111/cdev.13959

Published in final edited form as:
Child Dev. 2024 ; 95(1): 98–113. doi:10.1111/cdev.13959.

Attention to Novelty Interferes with Toddlers' Emerging Memory Decision-Making

Sarah Leckey[1,2], **Shefali Bhagath**[1,2], **Elliott G. Johnson**[1], **Simona Ghetti**[1,2]

[1]Center for Mind and Brain, University of California, Davis

[2]Department of Psychology, University of California, Davis

Abstract

Memory decision-making in 26- to 32-month-olds was investigated using visual-paired comparison paradigms, requiring toddlers to select familiar stimuli (Active condition) or view familiar and novel stimuli (Passive condition). In Experiment 1 (N = 108, 54.6% female, 62% White; replication N = 98), toddlers with higher accuracy in the Active condition showed reduced novelty preference in that condition, but not in the Passive condition ($d = -.11$). In Experiment 2 (N = 78; 52.6% female; 70.5% White), a brief 5% increase in target size boosted gaze transitions across conditions ($d = .50$) and accuracy in the Active condition ($d = .53$). Overall, evidence suggests that better attentional distribution can support decision-making. Research was conducted between 2014 – 2020 in Northern California.

Keywords

memory decision-making; toddlerhood; visual-paired comparison paradigm

If you ask a toddler to find a toy, they may immediately proceed to the correct location if they remember where it is, but they may explore their environment if their memory does not serve them well. Although these behaviors suggest that young children weigh memory evidence in their decisions to act, little is known about how they do so. Indeed, research focusing on early memory functioning has largely capitalized on paradigms that reduce decision demands (e.g., Hayne, 2007), such as visual preferences (Fantz, 1958), conditioned responses (Rovee-Collier & Cuevas, 2009), or imitation (Bauer & Leventon, 2013). The use of these paradigms has proven fruitful to probe the nature and durability of early memory representations (e.g., Bauer & Leventon, 2013; Morgan & Hayne, 2011; Richmond & Nelson, 2007). However, these approaches necessarily limit our ability to understand how young children weigh memory evidence to make their decisions. Without this knowledge, we can only have a limited grasp on how young children translate memory signals into action, such as how they go about responding to a request to find a hidden toy. As verbal interactions and reminiscence become more prominent (Nelson & Fivush, 2004), toddlers' memories are increasingly probed by caregivers' questions or instructions, motivating the question of how young children accumulate and weigh their own memory evidence in

---

Correspondence concerning this article should be addressed to Sarah Leckey, ssleckey@ucdavis.edu or Simona Ghetti, sghetti@ucdavis.edu, Center for Mind and Brain, University of California, Davis, 267 Cousteau Place, Davis, CA 95616.

response to direct memory probes. In the present research, we begin to address this question by examining how two-year-old children make decisions to select a familiar target from an array presenting that item and a novel one, providing the opportunity to gain insight on how young children navigate prioritizing memory accuracy over engaging in exploration of novel objects in response to a verbal instruction to do so.

Emerging Memory Abilities in Infancy and Early Childhood

Although infants and even toddlers cannot readily demonstrate their memory abilities with methods typically used in older children and adults to assess memory (i.e., including many trials, overt choice selection), reliance on alternative methods (Bauer & Leventon, 2013; Fantz, 1958; Rovee-Collier & Cuevas, 2009) has revealed that even very young infants have remarkable memory abilities. For example, newborns show preference for their mothers' voice (DeCasper & Fifer, 1980) and faces (Pascalis et al., 1995). Additionally, older infants compared to younger infants require less time or exposure in order to encode items (Rose et al., 1982; Hayne, 2007), remember more items for longer periods of time (Morgan & Hayne, 2011; Herbert & Hayne, 2000), and can recognize an item even with a change in context (e.g., Robinson & Pascalis, 2004). Finally, infants also become increasingly better at remembering which items were learned together, in what location, and in what order (Barr et al., 1996; Johnson et al., 2020; Newcombe et al., 2014). Despite the remarkable memory abilities documented in these studies, little is known about how young children make their memory decisions, or utilize their memory signals in order to make a decision. This is in part because existing approaches have chiefly focused on reducing response demands thereby limiting our ability to investigate how young children engage their emerging decision processes.

For example, the visual paired-comparison paradigm (VPC; Fantz, 1958), relies on infants' and toddlers' attentional orientation towards novel stimuli to infer that previously viewed stimuli were learned (Sokolov, 1963). This novelty preference response has been shown to be reliable in infants (Robinson & Pascalis, 2004; Rose et al., 1982), children (Morgan & Hayne, 2011) and adults (McKee & Squire, 1993; Richmond et al., 2004). However, because participants are not asked to report what they remember, it is difficult to draw conclusions about how they would use available memory information to support memory decisions, or other goals, following verbal prompts.

Recent evidence has suggested that even young children engage in decision processes during a perceptual task requiring them to determine which of two partially occluded pictures matches the target they are instructed to find (Leckey et al., 2020). However, these findings may not necessarily extend to memory. For example, it has been shown that 3-year-olds can successfully monitor the accuracy of their perceptual decisions but cannot accurately monitor the accuracy of their memory decisions until 4 years of age (Hembacher & Ghetti, 2014), suggesting a difference in the timing between perceptual and memory decisions. Therefore, the availability of successfully formed memories does not necessarily mean that those memories will be used effectively. The use of a modified visual comparison paradigm offers the opportunity to begin to address this question.

The VPC paradigm and Decision-Making

The VPC paradigm is well suited to begin exploring how toddlers weigh memory evidence during decisions for several reasons. First, as indicated earlier, this paradigm has generated vast evidence that infants and toddlers remember familiarized stimuli through exhibiting a novelty preference (e.g., Morgan & Hayne, 2011; Rose et al., 1982), and variables affecting adults' memory behaviors similarly affect novelty preferences. For example, length of exposure has been shown to support adults' memory performance on memory tasks (Musen & Treisman, 1990; Challis & Sidhu, 1993), as does familiarization time with infants. Rose (1983) showed that 6- and 12-month-old infants exhibited stronger novelty preferences after familiarization times of 30 seconds compared to 10 seconds. Therefore, there is strong body of knowledge about the conditions under which there is evidence of memory representations in infants and young children.

Second, although novelty preferences are thought to be driven by an automatic and implicit attentional orientation towards novel stimuli (Sokolov, 1963), the memory representations that evoke that orientation have been linked to healthy hippocampal function. The hippocampus is the brain structure that is fundamental to form and retain representations of unique past events (Eichenbaum et al., 1992) and patients with both adult-onset (McKee & Squire, 1993) and developmental amnesia (Munoz et al., 2011) due to hippocampal injury do not show novelty preferences compared to healthy controls after a delay. The same is true for rodents (Mumby et al., 2002) and primates (Pascalis & Bachevalier, 1999) with surgically induced hippocampal lesions. Therefore, we can be reassured that this paradigm captures the functioning of fundamental memory processes pertaining to retention of individual events.

Finally, the paradigm is simple enough that introducing an overt memory goal requiring a decision (i.e., asking to identify the familiar stimulus) does not fundamentally alter its procedure, but affords the additional collection of eye movement data during deliberations as is done in older children (Pathman et al., 2014) and adults (e.g., Richmond et al., 2007). In this way, information surrounding eye movements during the decision process and how these eye movements relate to the final decision can be ascertained in toddlers. This is important because eye movements have been shown to be sensitive to infants' goal representations (Tummeltshammer et al., 2014; Wu & Kirkham, 2010). For example, 8-month-old infants prioritize looking towards a box that a reliable compared to an unreliable speaker looks at (Tummeltshammer et al., 2014), indicating that infants integrate external information when prioritizing how to direct their attention and this is shown in their eye movements. Therefore, by giving toddlers an overt goal through verbal instructions, we can gain valuable insights about how toddlers deliberate and make a selection using visual exploration to assess a familiarized item and a novel one.

To the best of our knowledge, only two studies have examined how toddlers respond to verbal prompts requiring a decision in a VPC paradigm (Hayne et al., 2016; Imuta et al., 2013). Imuta and colleagues (2013) showed that 2- to 4-year-old children exhibited novelty preference for faces after long delays (at least one week) only if they were asked to remember the previous session prior to entering the testing room, suggesting that the instructions successfully cued retrieval. However, as in previous studies employing the VPC paradigm, although this study shows that toddlers can remember stimuli after long delays, it is impossible to establish whether children would have correctly selected the studied targets, by either pointing or touching the selected image, because memory was assessed implicitly through eye movements. In contrast, Hayne and colleagues (2016) asked toddlers to identify old items by pointing; however, the targets in the display were well known (e.g., the child's mother), impeding conclusions on how toddlers respond when asked to identify recently learned items presented along with novel items.

Evidence from the word learning literature suggests that toddlers find it difficult to ignore novel objects when they are asked to match objects to a recently learned label (e.g., Samuelson et al., 2017). Models of visual information sampling during decision-making have highlighted the interplay between goal-directed gaze distributions and salience-driven, bottom-up attentional captures (e.g., Gottlieb, 2018). These models underscore that the attentional capture towards novelty stimuli may interfere with evidence accumulation and evaluation, and lead to the inaccurate selection of that novel stimulus (Gottlieb, 2018; Krajbich, 2019). This is consistent with evidence that toddlers' exploration of novel objects can prevent selection of old objects corresponding to known verbal labels (Horst et al., 2010; Kucker et al, 2018).

Thus, toddlers' evaluation of their own memory evidence is likely influenced not only by how successfully a target cues the corresponding memory during a test, but also by how strongly a distracter captures the toddlers' attention due to its novelty. We propose that to achieve high accuracy, toddlers ought to divert their gaze from the novel stimulus and examine the old stimulus. Therefore, asking toddlers to select the studied item from an array which includes a novel item may challenge toddlers' memory decision processes.

In order to gain insight into these decision processes, and how a preference for novelty may impact toddlers' decisions, we conducted two experiments. In Experiment 1 we examined how toddlers make memory decisions in a VPC task, and Experiment 2 examined how an attentional manipulation impacts their looking behaviors and accuracy in the task. In both experiments we examined novelty preferences and gaze transitions between test stimuli. We examined the latter because they capture evidence evaluation preceding a decision in adults (e.g., Folke et al., 2016; Krajbich, 2019). Moreover, 2-year-old children have shown increased gaze transitions during difficult perceptual decisions (Leckey et al., 2020), further bolstering the case that gaze transitions may be helpful indicators of how toddlers assess and weigh memory evidence.

Experiment 1

The goal of Experiment 1 was to examine how 26- to 32-month-old toddlers make memory decisions. To achieve this goal, we assessed toddlers with both a passive viewing condition, in order to verify that toddlers exhibited the expected robust novelty preferences indicating memory retention, and with an active retrieval condition, in order to examine whether toddlers could avail themselves of these memory representations to guide accurate decisions. We selected this age range because of the paucity of research on 2-year-old children's overt memory decisions. Most research either utilizes infant paradigms that reduce decision demands in this age group or focuses on preschoolers and older children because of their known ability to act on their memory decisions. However, toddlers can respond to instructions (Leckey et al., 2020; Hayne et al., 2016), suggesting that they can direct attention to external stimuli when asked and incorporate instructions into their decision process, which is particularly important for our research question. Additionally, previous studies have shown that some decision processes can be reliably assessed at this age through behaviors such as gaze transitions and response latencies (Leckey et al., 2020).

Toddlers were assessed with a VPC task administered on an eye-tracker (Fig. 1). After familiarization (Fig. 1a), in the Passive condition (Fig. 1b), toddlers viewed displays including one previously presented item and one novel item, and no pointing or touching response was probed. In the Active condition (Fig. 1c), the same toddlers were asked to indicate the old item by pointing to their selection. In addition, toddlers completed a parallel version of the Active condition on a touchscreen computer to record touching response times. In the Passive condition, we expected toddlers to exhibit a novelty preference (Morgan & Hayne, 2011). However, for the Active condition, we expected reduced novelty preference with high accuracy. Specifically, successful evidence evaluation was expected to trigger successful orientation away from the novel item early in the trial, and not merely immediately prior to selecting one of the stimuli presented as response options, given that the extent of looking towards a response option has been found to be predictive of the upcoming choice of that option based on toddlers’ pointing (Hagihara et al., 2020).

To gain insight on the evaluation of memory evidence, we also collected gaze transitions between response options. In order to verify that these transitions reflected an assessment of memory evidence, we estimated drift diffusion parameters from touching response latencies on the touchscreen version of the task. Drift diffusion models (Ratcliff et al, 2016) posit that response latencies capture multiple processes yielding three parameters, including drift rate, boundary separation and non-decision (Supplemental Figure S1). The drift rate parameter represents the rate of evidence accumulation and is determined by the quality of the information or evidence gleaned from the presented stimuli. In our paradigm, the drift rate represents the quality of the match between the images on the screen and one’s memory of the image. Images that are encoded better would be expected to show higher drift rates compared to poorly encoded images. For example, in adults, the drift rate associated with the recognition of words studied several times is higher than that associated with the recognition of words only studied once (Ratcliff et al., 2004). The boundary separation parameter represents the decision threshold used, or how much information is required, to make a selection. If stronger evidence for a decision is deemed necessary, the decision boundaries will be expanded (higher separation parameter), allowing for slower and more accurate responses. Thus, higher separation parameter values may indicate that a person is more cautious in their responding, waiting to respond until they have all the evidence needed to make a decision. Finally, the non-decision parameter reflects processes unrelated to decision-making such as time it takes to make a motor response. To the extent that gaze transitions reflect toddlers’ ability to evaluate current evidence (Folke et al., 2016; Leckey et al., 2020), the drift rate and decision boundary parameters estimated from the touchscreen task might be associated with gaze transitions on the eye-tracker Active condition.

Methods

Participants

Participants included 108 26-to 34-month-old toddlers (*M* = 29.02 months, 59 female). Sixty-seven of them were White, 9 were Asian, 2 were African American, 24 were more than one race, and 6 did not report a race. Twenty-five of the parents identified their toddler as Hispanic. Families' household income varied, including less than $15,000 (*n* = 5), $15,000-$25,000 (*n* = 4), $25,000-$40,000 (*n* = 8), $40,000-$60,000 (*n* = 19), $60,000-$90,000 (*n* = 21), more than $90,000 (*n* = 48) and unreported (*n* = 3). Toddlers were recruited between September 2014 and December 2016 from a database of families contacted from birth records of infants born in counties surrounding the greater Sacramento, California area, who had expressed interest in participating in child development studies. We used this recruitment procedure to maximize our chances of enrolling a sample that represented the demographics of our region. None of the toddlers had a history of developmental or speech delays. This experiment was approved by the Institutional Review Board of the University of California, Davis and informed consent was obtained from all parents. All data and analytic code can be found on Open Science Framework (https://osf.io/sr9fq/).

An additional 10 toddlers were tested but were excluded from analyses due to being uncooperative on both tasks (8), or experimenter error in both tasks (2). Of the remaining 108 toddlers, seven toddlers did not complete the Passive condition and 25 did not complete the Active condition, leaving 101 participants in the Passive condition and 83 participants in the Active condition. In addition, 23 toddlers did not complete the Touchscreen task due to failure to cooperate (13), computer error (2), or not returning for the final session (8), leaving 85 participants in the task. Sample size was not determined through an a-priori calculation based on a target power for the study because these participants were recruited as a part of a larger study. However, we determined that the smallest sample size obtained here (N = 83 in the Active condition of the eye-tracker task) was sufficient to detect a small effect size (Cohen's *d* =. 31) at power =. 80 and alpha=.05 for overall novelty preference and memory accuracy.

Materials and Procedure

Toddlers were assessed over three sessions, each separated by about a week and they received a book after each session for their participation. At the beginning of each visit, the experimenter played with the toddler outside of the testing room for about 5 minutes in order to build rapport. Toddlers completed the Passive and Active conditions on the eye-tracker during sessions 1 and 2 and completed the touchscreen task during session 3. The Passive condition was always administered during the first session and the Active condition during the second session in order to prevent toddlers from engaging in active retrieval, responding, or pointing in the Passive condition. Similarly, the Touchscreen task was always administered in the third session, because pilot testing revealed that toddlers were more likely to reach forward to touch the eye-tracker if they had experienced the Touchscreen task first, interfering with eye movement data collection. Although counterbalancing the order of Passive eye-tracker, Active eye-tracker, and Active Touchscreen would be ideal from an experimental design perspective, we prioritized avoiding risks of contamination of the memory decision in the Active conditions on the Passive condition, and the risk of data loss on the eye-tracker if toddlers expected to touch the screen.

The stimuli were drawn from 160 colored line drawings from a widely used database (Rossion & Pourtois, 2004) depicting common objects and animals typically known to 2-year-old children. This was determined by utilizing the MacArthur-Bates Communicative Development Inventories along with age-of-acquisition norms from Morrison and colleagues (1997). Four sets of 40 drawings each were created based on random selection for counterbalancing purposes. Toddlers viewed 20 of them during encoding and the other 20 were used as distractors in the 20-trial retrieval task. Toddlers were assigned to a different set for each version of the task and set use was counterbalanced across participants.

Eye-tracker Task—The Passive and Active conditions were administered on a 17-inch Tobii T-120 eye-tracking system. The stimuli were 10 cm × 10 cm (visual angle 9.53) with 4.45 cm (visual angle 4.24) between them. Toddlers were sat on their parent's lap, approximately 60 cm from the monitor and the experimenter sat on their left. Parents wore dark sunglasses to ensure that their eyes were not recorded and that they could not view the stimuli. Parents were asked to hold their toddler to prevent any excessive movement or leaning forward. They were also asked to not speak to or engage with their toddler during the task.

Before administering each condition, toddlers underwent standard infant calibration procedures (Leckey et al., 2020). The experiment proceeded when toddlers' gaze was captured at all of the five calibration points on the screen. Default Tobii fixation filter settings were used for data reduction (velocity threshold: 35 pixels per sample; distance threshold: 35 pixels).

The encoding phase of the task followed and was identical in the Passive and Active conditions (Fig. 1a). We elected to have the toddlers view all 20 pictures for encoding and then do all 20 retrieval trials because this approach is consistent with how older children and adults are tested (Pathman & Ghetti, 2014; Hembacher & Ghetti, 2014) and allows us to begin bridging together the two literatures. The experimenter introduced the encoding phase by saying, "Now my friend Julia is going to show us some of her drawings.". Then, toddlers saw a video of a female experimenter presented at the center of the eye-tracker screen introducing the pictures as her drawings. Toddlers then viewed the stimuli. There were 20 pictures in the Passive condition and 22 pictures in the Active condition (to account for future practice retrieval trials). Each stimulus was presented individually at the center of the screen for 3 seconds. A white fixation cross, paired with a "ding" sound, was shown in between pictures to maintain the toddlers' attention to the center of the screen.

A new calibration procedure preceded the retrieval phase. The retrieval phase included 20 trials, each including one old and one new picture presented side by side. In the Passive condition (Fig. 1b), the experimenter stated that more drawings were going to be shown (i.e., "Now we are going to see some more drawings."). No other instructions were given. Before each trial, a brief video clip was shown of the same experimenter who presented the encoding trials saying, “Hey, look.”. Then, the stimuli were presented for 10 seconds. In the Active condition (Fig. 1c), toddlers were instructed to find the experimenter’s drawings (i.e., “Now we are going to help Julia find her drawings.”) and to indicate their answer by pointing. Then, they completed two practice trials to verify that toddlers responded to the instructions by pointing. If the toddlers pointed to the correct picture, they were told that they were correct, and if they pointed to the wrong picture, they were corrected and reminded that their job was to point to the old picture. Once the toddlers completed the two practice trials, they moved on to the 20 test trials, during which no feedback was provided. For each trial, the experimenter asked the toddler to indicate the previously seen picture (i.e., “Which picture did Julia show you before? Only one is Julia’s!”) on a screen with a fixation cross and then pressed the space bar to present the trial. This specific wording was chosen in order to focus on the action of the experimenter and reduce any confusion that would be potentially result from making reference to memory states (e.g., asking toddlers to “point to the one you remember”) or the status of the item (e.g., asking toddlers to “point to the old item”). The question was posed before the images appeared due to toddlers’ tendency to turn towards the experimenter talking; we wanted to ensure that the toddlers were oriented back to the screen for the images. As soon as the toddlers chose a picture, the experimenter keyed in the pointing response which brought up the next screen with a fixation cross. If toddlers refused to choose a picture, the experimenter pressed the space bar button and that trial was removed from analyses.

Touchscreen Task—The Touchscreen task was administered on a 17-inch Planar PT1701MU LCD touchscreen monitor with 1280 × 1024 resolution. The stimuli were the same size and distance apart as the eye-tracker task. This task was identical to the Active condition administered on the eye-tracker except that the child sat alone about 60 cm in front of the touchscreen monitor on a child-sized chair and for each retrieval trial the toddler was instructed to respond by touching the screen. As soon as the child touched a side of the screen, the task advanced to a blank screen before starting the next trial. If a child refused to respond, the experimenter keyed in a separate code to remove that trial from analysis and moved on to the next trial.

Additional Cognitive Assessments—We administered additional measures to examine differences between toddlers on cognitive skills such as language, inhibitory control and working memory. In order to conserve space, these tasks are described in Supplemental Results 1.

Data Processing

We used Tobii Studio software to create areas of interest (AOIs) for our intended analyses. These AOIs encompassed separate square images for the old and new stimuli, so that each trial had an old AOI and a new AOI. Our overall novelty preference variable was calculated by dividing the amount of time the toddler spent looking at the new AOI during the entire trial by the amount of time spent looking at both the old and new AOIs during the trial. In addition, we were interested in the toddlers’ looking patterns during the decision process and prior to committing to an answer; since the toddlers had an average pointing response latency in the Active condition of 5.03 seconds ($SD = 2.67$), we examined the first 4.5

seconds (1.69 standard deviations) of the trial for the cluster-based permutation tests to ensure that we had observations from as many trials and toddlers as possible prior to them committing to a decision.

Our gaze transitions variable corresponded to the number of times toddlers' gaze transitioned from one stimulus to the other during a trial. We defined a transition as a fixation on an AOI that was preceded by a fixation to the other AOI, including instances in which there were fixations on other (non-AOI) areas in between fixations to AOIs. The first fixation to an AOI in a trial was not counted as a transition, so the minimum number of transitions in a trial was zero.

Before data analysis on looking behaviors, we removed retrieval trials in the Active condition for which toddlers did not provide a pointing response. This resulted in 41 trials (2.46%) across 14 participants being eliminated from analyses. Next, we removed retrieval trials in both the Passive and Active conditions for which the toddlers had not looked at the picture during the encoding phase, indicated by no fixations during the 3-second period of encoding phase. This criterion resulted in the exclusion of 85 trials (5.12%) across 41 toddlers for the Active condition, and 113 trials (5.59%) across 44 toddlers for the Passive condition. Finally, we also removed trials for which the eye-tracker did not measure any look time to either AOI, old or new, during retrieval. This criterion resulted in the exclusion of 179 trials (10.78%) across 41 toddlers for the Active condition and 141 trials (6.98%) across 41 toddlers for the Passive condition. Overall, on average, we retained data from 17.49 trials (*SD* = 3.58, range 3–20 trials) per participant in the Passive condition and 16.33 trials (*SD* = 4.29, range 3–20 trials) per participant in the Active condition.

For the Touchscreen task, we first removed trials for which the toddlers did not offer a response by touching the screen. This resulted in 16 trials (.93%) across 7 participants being removed from analysis. Next, we removed any trials with touching response latencies less than 700 ms in duration. These touches were likely produced before processing the stimuli or trials in which the toddler was inattentive. We followed previous research in toddlers which has used 700 ms as a touching response latency response cutoff (Leckey et al., 2020). This criterion resulted in 8 trials (.47%) across 7 participants being removed from analysis. We also removed trials where the z-scored touching response latencies across each individual participant were +\- 3 standard deviations. This resulted in the exclusion of 41 trials (2.39%) across 41 participants. Additionally, since drift diffusion model estimations require relatively quick response latencies (Ratcliff et al., 2016), we removed trials with touching response latencies greater than 15 seconds. This cutoff was based on research indicating that drift diffusion parameters can be reliably estimated for response latencies up to 15 seconds (Lerche et al., 2017). This resulted in removing 37 trials (2.16%) across 18 participants. Overall, on average, we retained data from 18.54 trials (*SD* = 2.53, range 4–20 trials) per participant in this condition.

Analytical Approach

Multilevel Models.—We tested our hypotheses on overall novelty preference using multilevel models on the trial level data. We tested our multilevel models using the RStudio (Version 3.3.1, 2016) package nlme (Pinheiro et al., 2018). By utilizing a multilevel model, we were able to account for different participants contributing different numbers of trials and to avoid case-wise deletions inherent to ANOVA designs that would have occurred if participants did not have data in each of the eye-tracker conditions. In order to estimate the significance of our models we used a chi-square difference test, testing whether the model was different from a baseline model that included only the intercept.

Cluster-based Permutation Tests.—In order to assess the change in toddlers' novelty preference across the duration of the trial, we utilized cluster-based permutation tests. This analytical technique is frequently used in EEG research (Maris & Oostenveld, 2007), and has also been used for eye-tracking data in infants (Beckner et al., 2020; Oakes et al., 2017). A cluster-based permutation test is a nonparametric analytic approach, which allows us to utilize the high-temporal resolution of eye-tracking data and compare looking patterns of participants without the problem of multiple comparisons. This technique involves several steps. First, we performed uncorrected t tests at each individual time stamp of the eye tracker. For our dataset, this meant that we computed t tests for every 16.7 ms of the trial. Then, all adjacent time bins that had a significant t-value ($p < .05$) were grouped into a cluster. We summed all the t-values of this cluster which produced a cluster mass. In order to determine if the cluster mass was greater than what would be expected by chance, we created a null distribution. In order to create this null distribution, the original data set was randomly shuffled, t tests were computed for the shuffled data set, significant clusters were again formed, and the largest cluster mass observed was saved. This step was completed 1000 times, in order to create the distribution of cluster mass sizes based on our data. Finally, our original clusters were compared to this distribution. Significance of the clusters was computed by taking the percentage of the distribution that was bigger than the cluster, with statistically significant clusters falling within the bottom or top 2.5 percent of clusters from our null distribution. Our cluster-based permutation tests were tested using the RStudio package eyetrackingR (Dink & Ferguson, 2015).

Drift Diffusion Modeling.—To estimate drift diffusion parameters, we utilized a hierarchical drift diffusion model (HDDM; Vandekerckhove et al., 2011). HDDM uses a Bayesian estimation approach (Vandekerckhove et al., 2011; Voss et al., 2013) which allows for simultaneous estimation of group and subject parameters (Wiecki et al., 2013). This method takes into account the group variability for each subjects' parameter, which then allows for a smaller number of trials per participant compared to typical drift diffusion estimation procedures. We fit our HDDM model in python 3.6 on touching response times and accuracy data generated from the Touchscreen task. Response thresholds represented accuracy (accurate versus inaccurate). We estimated three parameters; *drift* parameter (v), *boundary separation* parameter (a), and *non-decision* parameter (t_0). Since there were no theoretical reasons for differences in evidence accumulation towards accurate versus inaccurate touching responses, the start point parameter, z_0, was fixed to 0.5. Drift and boundary separation parameters were then entered in a multiple regression analysis along with age to predict gaze transitions in the eye-tracker conditions.

Results and Discussion

Preliminary analyses verified that there were no significant differences between the toddlers who completed both conditions compared to the toddlers who did not complete the Active condition in age, novelty preference in the Passive condition, or in vocabulary, inhibitory control, and working memory (ps ≥ .06). Therefore, we decided to include all participants in the following analyses, regardless of their completion of only one task. Additionally, preliminary analyses verified that the effect of trial order did not matter (ps ≥ .31; See Supplemental Figure S2) and when trial order was included in the multilevel models the results stayed the same. Therefore, we did not include trial order as an additional factor in our analyses.

Novelty Preference and Accuracy Levels

First, we conducted a confirmatory analysis to verify that toddlers exhibited the expected novelty preference by examining whether their proportion looking time to the novel picture exceeded .50. In the Passive condition, we found a significant novelty preference, (M = .57, SD = .07), t(100) = 9.14, $p < .001$, d = .91. Similarly, we found a significant novelty preference for the Active condition, (M = .56, SD = .10), t(82) = 4.93, $p < .001$, d = .54. This indicates that we were able to obtain a reliable novelty preference using this paradigm in both conditions. Second, we explored toddlers' accuracy in the Active condition and found that overall as a sample, it was not different from chance, (M = .48, SD = .19, range: .10–1.00), t(82) = −1.10, p = .275, d = .12. Thus, although toddlers' novelty preference suggests that they had successfully learned the pictures, these memories did not appear to guide their decisions. However, we noted substantial individual variability in memory accuracy and thus conducted an exploratory analysis and created two groups of toddlers, those who performed above chance (>50%; N=35, M = .65, SD = .12) and those who performed at or below chance (≤ 50%; N=48, M = .35, SD = .12) and used this grouping to guide subsequent analyses (See Supplemental Results 2 for a complementary analysis of novelty preference in which accuracy is used as a continuous variable).

Individual Differences in Accuracy and Novelty Preference

To investigate whether individual differences in accuracy levels were associated with differences in novelty preferences across the Passive and Active conditions, we conducted a mixed level model, with fixed effects of retrieval condition (Passive and Active) and accuracy group (Low and High, based on whether they performed above chance) and a random effect of subject. We found significant main effects of retrieval condition, b = .09, $t(2611)$ = 5.42, $p < .001$, d = .10, and accuracy group, b = .12, $t(81)$ = 7.81, $p < .001$, d = .86. These main effects were qualified by a significant retrieval condition by accuracy group interaction, b = −.12, $t(2611)$ = −5.80, $p < .001$, d = −.11. This model was significantly different than the baseline model, X^2 (3) = 59.61, $p < .001$. Post-hoc analyses revealed that in the Passive condition, there was no significant difference in novelty preference between toddlers exhibiting high accuracy (M = .57, SD = .09), or low accuracy (M = .58, SD = .07), t(74) = −.39, p = .696, d = −.09. However, in the Active condition, the high accuracy group demonstrated significantly lower novelty preference (M = .49, SD = .09) compared to the low accuracy group (M = .61, SD = .08), t(81) = −6.08, p <.001, d = −1.35. Thus, regardless of their accuracy level in the Active condition, toddlers showed robust novelty preference in the Passive condition indicative of similar retention across the two accuracy groups. However, those who responded accurately were more likely to divert their gaze from the novel item to the target due to the instruction, resulting in lower levels of novelty preference (Fig. 2a).

To investigate when during retrieval trials looking patterns between toddlers who exhibited high or low accuracy begun to diverge, we examined the time course of their novelty preferences across the duration of the trials and compared differences in the proportion of looking time towards the novel item as a function of accuracy group, using cluster-based permutation tests. In the Active condition, we found 11 different clusters. After comparing the clusters to the distribution created by the random shuffling, we found that 2 clusters were significantly different. Toddlers who exhibited high accuracy showed significantly reduced novelty preference between 1.65 to 2.55 seconds after stimulus onset ($p < .001$) and again from 2.77 to 3.07 seconds after stimulus onset ($p = .050$; Fig. 2b). These changes in looking pattern occurred well before toddlers committed to a decision which was on average 5.47 seconds ($SD = 3.55$) for the high accuracy group and 4.71 seconds ($SD = 1.75$) for low accuracy group. In the Passive condition, we found 4 different clusters where looking time was seemingly different between low and high accuracy toddlers. After comparing these clusters to the distribution created by the random shuffling, we found that none of the clusters were significant ($ps \geq .490$; Fig. 2c). Therefore, although toddlers in the low and high accuracy groups showed distinct looking patterns in the Active condition, they were comparable in the Passive condition. We did not initially intend to take an individual difference approach for the comparison between Active and Passive condition and analyses of memory accuracy. Thus, we conducted a successful replication experiment to confirm this pattern of results (Supplemental Experiment 1).

In order to investigate factors that might explain differences between high and low-accuracy groups in the Active condition, a supplementary analysis examining differences between language skills, working memory, and inhibitory control was conducted (See Supplemental Results 1 & Supplemental Table 1). No significant differences between the groups were found in any of the measures.

Gaze Transitions, Evidence Accumulation, and Decision Boundaries

To explore how participants accumulated and weighed memory evidence, we asked whether parameters corresponding to evidence accumulation (indicated by drift rate parameter) and amount of evidence needed to endorse an option (indicated by boundary separation parameter) from the drift diffusion model, estimated with response latencies from the Touchscreen task were related to toddlers' gaze transitions. We used the response latencies from the Touchscreen Task because toddlers directly provided them, but these response latencies were significantly associated with those entered by experimenters in response to pointing in the Eye tracker task (See Supplemental Figure S3 for correlation between Touchscreen task response latencies and response latencies in the Active condition of the Eye-tracker task).

Overall transitions between response options in the Active condition (M = 1.96, SD = 1.03; High Accuracy M = 1.90, SD = 1.12; Low Accuracy M = 2.00, SD = .97) were entered as dependent measure in a multiple regression analysis in which we entered the drift rate (M = .01, SD = .25, range: −.47 - .75; Supplemental Figure S4a) and boundary separation parameters (M = 2.64, SD = .60, range: 1.48 – 4.33; Supplemental Figure S4b) simultaneously as predictors. Toddlers' age was also included in the multiple regression. The model was significant (R^2=.08, F(3,66) = 3.04, p=.035) and the boundary separation parameter significantly predicted average gaze transitions (b = .47, p=.023, d = .28; Supplemental Figure S5a). However, neither the drift rate parameter (b = .28, p=.584, d = .07; Supplemental Figure S5b) nor age (b = −.13, p=.085, d = −.21) predicted average gaze transitions.

A control multiple regression model utilizing gaze transitions measured during the Passive condition (M = 2.42, SD = .64; High Accuracy M = 2.20, SD = .75; Low Accuracy M = 2.62, SD = .55) was not significant (R^2=.01, F(3,74) = 1.36, p=.261) and neither the boundary separation parameter (b = −.13, p=.345, d = −.11), nor drift rate parameter (b = −.12, p=.687, d = −.05), and age (b = .07, p=.083, d = .20) predicted average gaze transitions.

Overall, our results revealed individual differences in memory accuracy, which were associated with reduced novelty preference early in the Active condition trial. Importantly, our cluster-based permutation analysis revealed that regardless of accuracy level, all toddlers began the task with similar novelty preferences. The groups then slowly diverged with the Low accuracy group continuing to show a preference for the novel image. The fact that toddlers in the High accuracy group does not merely favor the old item from the beginning, but rather start out by first examining the novel image and then transition to examining both images more equally suggests that they are more readily able to inhibit their tendency to favor the novel image when their given goal is to identify the old image. Additionally, analyses of individual differences showed that toddlers who switched gaze more frequently on the eye-tracking task were more likely to exhibit a wider decision boundary suggesting that they require more evidence before making their selection on the touchscreen task, and more generally that there may be a functional relation between alternating visual exploration of response options and weighing evidence for decision making. Thus, these findings converge in suggesting that accuracy depends on toddlers' ability to exert attentional control and divert their attention away from the novel image and towards the target stimulus; toddlers seemingly exhibit individual differences in the extent to which they can re-orient their attentional mechanisms based on their current goals. If this is the case, then an attentional manipulation aimed at briefly directing attention toward the target stimulus could help toddlers overcome their difficulties.

Experiment 2

The goal of Experiment 2 was to examine whether an attentional manipulation designed to induce better gaze distribution between response options, and thus mimic the behavior shown in high-performing toddlers, would increase memory accuracy. Past research in infants has shown that manipulating one's gaze patterns during familiarization with an attentional manipulation in a VPC procedure increases novelty preference at test (Jankowski et al., 2001). Here, we were interested in whether manipulating gaze patterns during retrieval impacts toddlers' memory performance. Our hypothesis was that the Active Condition in Experiment 1 posed attentional inhibition demands that exceeded those of most participants. If this is the case, helping toddlers distribute their attention should benefit accuracy. Therefore, in Experiment 2, we asked whether a brief flicker of the studied item early during each retrieval trial (Flicker condition, Fig. 3a) compared to no flicker (Control condition, Fig. 3b) would not only promote immediate viewing of the target, but also disrupt attentional capture towards the novel item throughout the trial supporting decision making. Prior research has shown that cues such as a brief movement of the target, a flashing square around the target, or a brief cue on the same side of the target before it appears, helps infants direct their attention towards that target and remember placement of that target (Ross-Sheehy et al., 2011; Wu & Kirkham, 2010; Johnson & Tucker, 1996). Thus, we expected that the brief flicker would facilitate toddlers' accumulation of evidence from the target and better evaluation of this evidence resulting in lower novelty preference, greater number of gaze transitions and, ultimately, higher accuracy in the Flicker compared to the Control condition. As in Experiment 1, all toddlers completed a Passive and an Active version of the task to establish viewing behaviors in the absence of an explicit memory demand. We note that a control experiment ruled out the possibility that the presentation of the flicker in and of itself, without memory retrieval, increased selection of the old item (Supplemental Experiment 2).

Methods

Participants

Participants included 78 toddlers aged 25–34 months (*M* = 28.48 months, 41 female). Fifty-five of the toddlers were White, 2 were Asian American, 15 were more than one race, and 6 did not report a race. Fourteen of the toddlers identified as Hispanic. Families' household incomes were $15,000-$25,000 (*n* = 1), $25,000-$40,000 (*n* = 3), $40,000-$60,000 (*n* = 3), $60,000-$90,000 (*n* = 19), more than $90,000 (*n* = 49), and 3 declined to answer. Toddlers were recruited between July 2018 and March 2020 using the same sampling procedures and exclusion criteria as Experiment 1. We initially set our sample size as 52, with 26 toddlers in each attention condition who complete both the Passive and Active conditions, based on a power analysis for a between-subject design to achieve .80 statistical power and alpha smaller than .05 to detect a large effect size based on Experiment 1. A large effect size was expected based on the large difference between high performing and low performing toddlers in Experiment 1 and its replication. However, after data collection for these 52 toddlers with data in both conditions was complete, we decided to utilize multilevel models for our analyses which allowed us to include all toddlers who contributed any data, even if they were missing data from either the Passive or Active condition, in line with the approach adopted for Experiment 1. Some toddlers only completed one of the conditions due to inattention, computer error, or experimenter error in the other: 5 toddlers did not complete the Passive condition and 15 did not complete the Active condition, resulting in 73 participants in the Passive condition (40 of whom were in the Flicker condition), and 63 participants in the Active condition (34 of whom were in the Flicker condition). This experiment was approved by the Institutional Review Board of the University of California, Davis and informed consent was obtained from all parents.

Procedure

As in Experiment 1, toddlers completed the eye-tracker task in the Passive and Active conditions, but toddlers were assessed after being familiarized with only 10 drawings. During retrieval, they were assigned to either a Control or a Flicker condition. In the Control condition (Fig. 3a), the retrieval procedure was the same as in Experiment 1. In the Flicker condition (Fig. 3b), the procedure was identical except that the old image increased in size by 5%, 1,000 msecs into the trial and remained large for 50 msecs before returning to its original size. This created a brief flicker of the image. Additionally, both conditions had the same practice trials as Experiment 1, in which they received feedback on their pointing responses for two static image trials. Finally, toddlers completed the same additional cognitive tasks described in Experiment 1 and Supplemental Results 1. The results of these tasks are reported in Supplemental Table 1c.

Data Processing

Data were processed as in Experiment 1. Toddlers' average pointing response latency was 5.45 seconds, so we examined the first 5 seconds of the trial for analyses examining look time across trials. Thirteen trials (2.06%) across 5 participants were eliminated due to toddlers not providing a pointing response. There were 34 excluded trials (5.40%) across 15 toddlers for the Active condition and 55 trials (7.53%) across 25 toddlers for the Passive condition for toddlers not looking at the old picture during the encoding phase. Finally, we removed 39 trials (6.19%) across 20 toddlers in the Active condition and 32 trials (4.38%) across 20 toddlers in the Passive condition due to no looking time measured for either AOI in the retrieval phase. Overall, we retained an average of 8.81 trials ($SD = 1.73$, range 1–10 trials) per participant in the Passive condition and 8.64 trials ($SD = 1.92$, range 3–10 trials) per participant in the Active condition.

Analytical Approach

As in Experiment 1, we used multilevel models on the trial level data to test our hypotheses on overall novelty preferences. For assessing differences in novelty preferences across time between the Control and Flicker conditions, we utilized cluster-based permutation tests.

Results and Discussion

Similar to Experiment 1, preliminary analyses revealed that there were no significant differences in age or novelty preference in the Passive condition between the toddlers who completed both conditions compared to the toddlers who did not complete the Active condition (ps ≥ .24). Therefore, we decided to include all participants, regardless of their completion of both tasks. Additionally, preliminary analyses verified that the effect of trial order did not matter for novelty preferences (ps ≥ .33) and when trial order was included in the multilevel models the results stayed the same. Therefore, trial order was not included as an additional factor.

Novelty Preference

We first conducted a confirmatory analysis to establish whether the proportion of looking time to the novel picture exceeded .50 in the Flicker and Control attention conditions. In the Control condition ($M = .55$, $SD = .12$), we found that proportion looking times were overall significantly different from .5 ($t(61) = 3.13$, $p = .003$, $d = .40$), replicating the results of Experiment 1. However, in the Flicker condition ($M = .49$, $SD = .13$), the proportion looking times were not overall significantly different from .5 ($t(73) = -.52$, $p = .607$, $d = .06$). We next determined the effect of our attention manipulation on the overall novelty preference in the Passive and Active conditions. We conducted a confirmatory multilevel model with fixed effects of attention condition (Flicker versus Control) and retrieval condition (Passive versus Active) and a random effect of subject. We found a significant main effect of the attention condition, $b = -.05$, $t(1107) = -2.39$, $p = .017$, $d = -.07$, such that in the Flicker condition, overall novelty preference was reduced ($M = .50$, $SD = .26$) compared to the Control condition ($M = .55$, $SD = .30$). The main effect of retrieval condition was not significant, $b = -.02$, $t(1107) = -1.38$, $p = .169$, $d = -.04$, as the flicker diverted gaze in both conditions. This model was significantly different than the baseline model, X^2 (2) = 7.28, p =.026. Thus, the attention manipulation effectively impacted looking behaviors in both conditions (Fig. 4a).

This overall decrease in novelty preference in the Flicker condition indicates that the manipulation successfully diverted gaze away from the novel stimulus and towards the target. However, it does not tell us yet whether this change in eye movements influenced evidence accumulation and accuracy. We further investigated novelty preference throughout the duration of the trial (instead of overall novelty preference) to establish the extent to which toddlers' looking patterns in the Active Flicker versus Control condition mimicked the looking patterns of the High and Low accuracy toddlers in Experiment 1. We examined the differences in the proportion of looking time towards the novel item as a function of flicker condition in the Active condition by conducting a cluster-based permutation test. We found 3 different clusters. After comparing the clusters to the distribution created by the random shuffling, we found that 1 cluster was significant. Looking patterns for the Control and Flicker conditions differed from 1.23 to 2.03 seconds after stimulus onset ($p = .001$; Fig. 4b). In the Passive condition, only a briefer significant cluster was found from 1.42 to 1.97 seconds around the time of the presentation of the flicker ($p = .008$; Fig. 4c).

Gaze Transitions

We next examined the effect of our manipulation on the amount of gaze transitions. We conducted a confirmatory Poisson multilevel model with fixed effects of attention condition (Control and Flicker) and retrieval condition (Passive and Active) and a random effect of subject. This model was significantly different from the baseline model, X^2 (3) = 34.39, $p <$ .001. We found a significant main effect of attention condition, $b = .40$, z = 4.43, $p < .001$, $d = .50$, with the toddlers in the Flicker condition showing more gaze transitions ($M = 2.59$, $SD = 1.93$) than those in the Control condition ($M = 1.90$, $SD = 1.70$). The main effect of retrieval condition was not significant, $b = .08$, $z = 1.58$, $p = .115$, $d = .13$, but there was a significant interaction between attention and retrieval conditions, $b = .22$, z = 2.62, $p = .01$, $d = .38$. Although post-hoc analyses revealed that for both Passive ($b = .22$, z = 2.96, $p =$

.003, Bonferroni-corrected p (p_{bonf}) = .006, d = .32) and Active (b = .43, z = 2.72, p = .006, p_{bonf} = .012, d = .51) conditions, there were more gaze transitions for the Flicker condition compared to the Control condition, the difference between the Flicker and Control condition was greater in the Active compared to the Passive condition according to effect sizes.

Memory Accuracy

Next, we examined toddlers' accuracy in the Active condition. Similar to Experiment 1, we found that overall, as a sample, toddlers' accuracy was not different from chance (M = .51, SD = .21, range: 0.00–1.00), t(62) = .51, p = .611, d = .06. Finally, we investigated the effect of our manipulation on accuracy. We conducted a confirmatory logistic multilevel model with trial accuracy (0 or 1) as the dependent measure and a fixed effect of attention condition (Control versus Flicker) and random effect of subject. We found a significant main effect, b = .43, z = 2.04, p = .041, d = .53, such that Flicker condition (M = .56, SD = .24) resulted in significantly higher accuracy compared to the Control condition (M = .46, SD = .16). This model was significantly different from the baseline model, X^2 (1) = 4.12, p =.043.

Overall, the manipulation effectively increased accuracy by supporting toddlers' gaze distribution and visual comparison between the response options. Critically, we showed that the use of an attentional capture alone without a memory retrieval demand did not result in increased gaze transitions and selection of that target (Supplemental Experiment 2). Instead of merely causing participants to look at the studied item and choose that item, the flicker seems to have interfered with the attentional capture towards novelty paving the way for toddlers being more likely to engage in a process of evaluation of both items resulting in the increased gaze transitions. These results underscore the importance of redirecting attention away from the novel item and towards the studied items to support memory decision processes.

General Discussion

Children's memory abilities develop rapidly in the first few years of life (e.g., Hayne, 2007). Despite the tremendous progress in this field of research, little is known about early memory decision processes and how these decisions may come to play when toddlers are tasked with identifying and previously seen item in the presence of a distracter. In this context, one could make contrasting predictions. On the one hand, high levels of memory accuracy may be expected under conditions that elicit strong levels of spontaneous novelty preference, consistent with the idea that strong memories will be most likely to yield higher levels of accurate memory decisions. On the other hand, toddlers' tendency to look at novel stimuli may interfere with the evaluation of memory evidence, leading to inaccurate decisions (Gottlieb, 2018). The current research began to examine these possibilities. By assessing toddlers with a passive viewing condition, we verified that toddlers exhibited the expected robust novelty preferences indicating memory retention, and by including active retrieval conditions, we examined whether toddlers could avail themselves of these memory representations to guide accurate decisions.

In Experiment 1, and its replication (Supplemental Experiment 1), toddlers overall achieved low levels of accuracy in the Active condition, consistent with the hypothesis that attentional orientation toward novelty may interfere with their memory decisions. However, there was substantial individual variability in accuracy. Toddlers whose accuracy was above chance diverted their gaze away from the novel picture early in the trial, around a second and a half after stimulus onset, and eventually selected the old picture a few seconds later. In the Passive condition, these toddlers exhibited similar levels of novelty preference to those who exhibited low accuracy in the Active condition, indicating that general differences in memory retention could not account for these findings. Infants attend to novelty in order to learn about their environment (i.e., properties of a new toy; Oudeyer & Smith, 2016). However, the presence of novel objects has also been shown to interfere with infants' and toddlers' ability to successfully select an object corresponding to a learned word label (e.g., Mather & Plunkett, 2012; Samuelson et al., 2017). This seems similar to what occurred in our study and we asked what factors might explain how a subset of toddlers were able to respond accurately.

One explanation has to do with toddlers' ability to reorient their attention from the novel stimulus to the old stimulus based on current goals. Attentional processes, including alerting, orienting, and executive attention emerge during infancy (e.g., Amso & Sceriff, 2015) and their early development affects exploration and decision making (Blanco & Sloutsky, 2020). Experiment 2 examined the possibility that orienting attention away from the novel stimulus and towards the old stimulus may facilitate consideration of memory evidence. Consistent with our hypothesis, toddlers in the Flicker condition showed lower novelty preference, more gaze transitions and higher accuracy compared to the Control condition, suggesting that in Experiment 1 the toddlers' preference to look at novel items may have led them to choose incorrectly.

Another, similar, explanation that may explain how some toddlers were able to respond accurately in the face of novelty preferences has to do with toddlers' capacity to inhibit their tendency to examine novel objects in the first place. Inhibitory control, which is the ability to stop a dominant response (e.g., examining a novel image) in order to achieve a goal (e.g., choosing the previously seen image), emerges around the first year of life and continues to develop across childhood (Davidson et al., 2006). The toddlers who were successful in selecting the correct image may have better inhibitory control which allowed them to disengage from the novel object in order to follow the instructions given to them. We did not find any differences in cognitive inhibitory control utilizing typical stroop-like methods (Supplemental Results 1); however, these methods do not assess the ability to inhibit visual preferences and do not pit novelty against the desired response. Future research utilizing a task focusing in attentional inhibition is needed to fully assess this explanation.

We also examined how gaze transitions between response options were related to toddlers' decision processes. Although looking times and gaze transitions are related, proportion looking times can be obtained with varying amounts of gaze transitions (e.g., the same proportion of looking time towards a novel item can be obtained from toddlers who exhibited one or many transitions). Thus, toddlers could exhibit identical novelty preference, but different gaze transitions, which could translate into differing decision processes.

Our examination of the relation between gaze transitions and drift diffusion parameters (Vandekerckhove et al., 2011) in Experiment 1 revealed that the boundary separation parameter, which measures the amount of evidence needed in order to render a response, positively predicted the average amount of gaze transitions toddlers made in the Active condition. This finding is consistent with the possibility that toddlers who were more cautious, or experienced the options as carrying more distinct information, exhibited more gaze transitions, submitting a pointing response only after thoroughly examining the stimuli. In Experiment 2, the presentation of a brief flicker, resulting from a momentary and small change in size of the target image, increased gaze transitions in addition to increased accuracy, suggesting that reorienting attention affected how much evidence toddlers collected before committing to a decision. Research with adults has reported connections between eye movements towards upcoming targets and markers of decision making (Parker et al., 2020), underscoring the importance of further characterizing these behaviors when overt memory decision making emerges in childhood.

A potential concern in Experiment 2 is that toddlers' accuracy gains were merely due to the flicker inducing overall more looking at the target as opposed to engaging more in evidence evaluation. Increased gaze transitions in the Flicker condition suggest instead that the manipulation induced exploration of both options by altering attentional capture to novelty. Additionally, the apparent flicker is very brief and very early in the trial, on average 4 seconds before the toddlers indicate their choice. Therefore, toddlers did not simply choose what they are currently being encouraged to look at, because there is considerable time in between the appearance of the flicker and their actual decision time. Moreover, we conducted a control experiment (Supplemental Experiment 2), in which a new group of toddlers completed only the retrieval portion of the Active condition as a guessing game. Thus, they had no memory to rely on when asked which image was the experimenters. Results showed that toddlers in that Flicker condition did not switch gaze between response options or choose the flickering image more than toddlers who completed the Control condition. Thus, the Flicker condition in the Active condition of Experiment 2 reduced the inaccurate selection of the novel image, supporting toddlers' more effective decision making by disengaging them from the novel image, allowing them to visually explore response options and be more likely to utilize their memories. We recognize, however, that it would be inappropriate to consider increased gaze transitions as a marker of decision accuracy. Indeed, in our previous research on perceptual decisions, increased gaze transitions were associated with visual inspection during difficult trials (Leckey et al., 2020). Future research should further clarify the conditions under which gaze transitions are more or less associated with decision accuracy.

These experiments began to examine how toddlers make memory decisions, however, there are several different questions that should be followed up in future research. For example, Experiment 2 was conducted utilizing a between-subject design, which was important to reduce the risk of potential carry-over effects, but future research with within-subject manipulations would help establish if the attentional manipulation is equally effective across participants regardless of their viewing behavior when no flicker is presented. Future research should also include additional manipulations (e.g., flicker presented at a different location of the screen) to establish whether orienting attention towards the target is necessary or whether orienting attention away from the novel item suffices. In addition, the verbal prompts for the retrieval phase of the Active condition for both experiments were provided before the image pairs were presented. We made this decision based on observations that toddlers tend to turn towards speakers and having the verbal prompts before the trial ensured the toddlers would be reoriented back towards the screen. However, presenting the instructions before the stimuli may impact participants' responses. For example, some toddlers could have forgotten the prompt and relied on the novelty signal to make their choice. Future research should examine the impact of the timing of instructions on toddlers' responses in visual paired comparison tasks to establish whether toddlers reverted to novelty because of when instructions were presented. Nevertheless, it is reassuring that the prompts were identical across all trials in our research; if toddlers forgot the question on the first few trials, they may have remembered better in later trials, but we did not find any order effect. Moreover, high accuracy levels have been found with perceptual decision tasks in which the experimenter uttered the prompt before toddlers saw the stimuli (Leckey et al., 2020); if toddlers forgot the questions, they should have performed poorly in that task as well.

Another potential limitation pertains to trial loss. In this study we had the toddlers complete 20 trials in one testing session, with an average of 3–4 trials being removed from analysis. This trial loss may have been due to toddlers losing focus during the longer task. We are reassured by the fact that we retained more than 70 percent of data in trial 20 across both conditions, but this may suggest that future studies should examine the effects of including additional breaks or establish the optimal number of trials in this age group, in order to eliminate potential fatigue while maximizing trial numbers, which is important for power (DeBolt et al., 2020). An additional caveat is that in our paradigm for Experiment 2 we chose to manipulate toddlers' gaze in order to assess how novelty preference impacted toddlers' memory decisions. However, prior research with the VPC paradigm has revealed that novelty preferences weaken with longer delays between familiarization and retrieval phases (e.g., Morgan & Hayne, 2011). With a longer delay, toddlers' memory decisions may have been less impacted, potentially resulting in higher decision accuracy. These future studies would be an important extension of the current experiments. Additionally, prior research using the head-turn procedure has shown that infants' behavior in the task changed as they gained more experience with the paradigm. Specifically, infants exhibited a stronger familiarity response when they had less experience with the task compared to more experience (Santolin et al., 2020). This research highlights the importance of taking infants' experience into account when interpreting results, and future research should investigate this with our memory task as well. It is plausible to predict that more experience may influence assessment and orientation to the novelty of individual items. Finally, although we did our best to recruit a diverse sample that was consistent with the racial and ethnic background of the region, our samples were more likely to include White, non-Hispanic families than expected based on the demographics of the region, which might influence the generalizability of our results.

In conclusion, toddlers may struggle to accurately act on their memories when novel stimuli are presented, and attentional manipulations may be used to help toddlers overcome this problem. Moreover, these results obtained with paradigms using multiple trials hold the promise to better connect evidence obtained with young children to findings from older children and adults. Future research should further explore the associations between attentional control and novelty preference as an explanation for why some toddlers can make accurate decisions while others perseverate on novel, or goal irrelevant, stimuli.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The data and analytic code necessary to reproduce the analyses presented here are publicly accessible. Data and analytic are available at the following URL: https://osf.io/sr9fq/. The materials necessary to attempt to replicate the findings presented here are widely and publicly accessible. The analyses presented here were not preregistered.

This work was supported by a grant from the National Science Foundation (NSF; BCS1424058) to S.G. Any opinions, findings, and conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the NSF. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. S.L was supported by National Institute of Child Health and Human Development Grant F31HD102153.

References

Amso D, & Scerif G (2015). The attentive brain: insights from developmental cognitive neuroscience. Nature Reviews Neuroscience, 16, 606–619. 10.1038/nrn4025 [PubMed: 26383703]

Barr R, Dowden A, & Hayne H (1996). Developmental changes in deferred imitation by 6-to 24-month-old infants. Infant Behavior and Development, 19, 159–170. 10.1016/S0163-6383(96)90015-6

Bauer PJ, & Leventon JS (2013). Memory for one‐time experiences in the second year of life: Implications for the status of episodic memory. Infancy, 18, 755–781. 10.1111/infa.12005

Beckner AG, Cantrell LM, DeBolt MC, Martinez M, Luck SJ, & Oakes LM (2020). Visual short‐term memory for overtly attended objects during infancy. Infancy, 25, 347–370. 10.1111/infa.12332 [PubMed: 32749061]

Blanco NJ, & Sloutsky VM (2020). Attentional mechanisms drive systematic exploration in young children. Cognition, 202, 104327. 10.1016/j.cognition.2020.104327 [PubMed: 32464341]

Challis BH, & Sidhu R (1993). Dissociative effect of massed repetition on implicit and explicit measures of memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 115–127. 10.1037/0278-7393.19.1.115 [PubMed: 8423430]

Davidson MC, Amso D, Anderson LC, & Diamond A (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. Neuropsychologia, 44, 2037–2078. 10.1016/j.neuropsychologia.2006.02.006 [PubMed: 16580701]

DeBolt MC, Rhemtulla M, & Oakes LM (2020). Robust data and power in infant research: A case study of the effect of number of infants and number of trials in visual preference procedures. Infancy, 25, 393–419. 10.1111/infa.12337 [PubMed: 32744759]

DeCasper AJ, & Fifer WP (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208, 1174–1176. 10.1126/science.7375928 [PubMed: 7375928]

Dink JW, & Ferguson B (2015). eyetrackingR: An R library for eye-tracking data analysis. Retrieved from www.eyetracking-r.com.

Eichenbaum H, Otto T, & Cohen NJ (1992). The hippocampus—what does it do?. Behavioral and Neural Biology, 57, 2–36. 10.1016/0163-1047(92)90724-I [PubMed: 1567331]

Fantz RL (1958). Pattern vision in young infants. The Psychological Record, 8, 43–47.

Folke T, Jacobsen C, Fleming SM, & De Martino B (2016). Explicit representation of confidence informs future value-based decisions. Nature Human Behaviour, 1, 1–8.

Gottlieb J (2018). Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. Cortex, 102, 150–160. 10.1016/j.cortex.2017.08.019 [PubMed: 28919222]

Hagihara H, Ienaga N, Terayama K, Moriguchi Y, & Sakagami MA (2020). Looking represents choosing in toddlers: Exploring the equivalence between multimodal measures in forced-choice tasks. Infancy. 26, 148–167. 10.1111/infa.12377 [PubMed: 33341103]

Hayne H (2007). Infant memory development: New questions, new answers. In Oakes LM & Bauer PJ (Eds.), Short-and long-term memory in infancy and early childhood: Taking the first steps toward remembering (pp. 209–239), Oxford University Press.

Hayne H, Jaeger K, Sonne T, & Gross J (2016). Visual attention to meaningful stimuli by 1-to 3-year olds: implications for the measurement of memory. Developmental Psychobiology, 58, 808–816. 10.1002/dev.21455 [PubMed: 27753455]

Herbert J, & Hayne H (2000). The ontogeny of long-term retention during the second year of life. Developmental Science, 3, 50–56. 10.1111/1467-7687.00099

Hembacher E, & Ghetti S (2014). Don’t look at my answer: Subjective uncertainty underlies preschoolers’ exclusion of their least accurate memories. Psychological science, 25, 1768–1776. 10.1177/0956797614542273 [PubMed: 25015686]

Horst JS, & Samuelson LK (2008). Fast mapping but poor retention by 24-month-old infants. Infancy, 13, 128–157. 10.1080/15250000701795598 [PubMed: 33412722]

Imuta K, Scarf D, & Hayne H (2013). The effect of verbal reminders on memory reactivation in 2-, 3-, and 4-year-old children. Developmental Psychology, 49, 1058–1065. 10.1037/a0029432 [PubMed: 22822936]

Jankowski JJ, Rose SA, & Feldman JF (2001). Modifying the distribution of attention in infants. Child Development, 72, 339–351. 10.1111/1467-8624.00282 [PubMed: 11333070]

Johnson EG, Leckey S, Davinson K, & Ghetti S (2020). Associative binding in early childhood: Evidence from a preferential looking paradigm. Developmental Psychobiology, 62, 266–278. 10.1002/dev.21904 [PubMed: 31404482]

Johnson MH, & Tucker LA (1996). The development and temporal dynamics of spatial orienting in infants. Journal of Experimental Child Psychology, 63, 171–188. 10.1006/jecp.1996.0046 [PubMed: 8812042]

Krajbich I (2019). Accounting for attention in sequential sampling models of decision making. Current Opinion in Psychology, 29, 6–11. 10.1016/j.copsyc.2018.10.008 [PubMed: 30368108]

Kucker SC, McMurray B, & Samuelson LK (2018). Too much of a good thing: How novelty biases and vocabulary influence known and novel referent selection in 18-month old children and associative learning models. Cognitive Science, 42, 463–493. 10.1111/cogs.12610 [PubMed: 29630722]

Leckey S, Selmeczy D, Kazemi A, Johnson EG, Hembacher E, & Ghetti S (2020). Response latencies and eye gaze provide insight on how toddlers gather evidence under uncertainty. Nature Human Behaviour, 4, 928–936. 10.1038/s41562-020-0913-y

Lerche V, Voss A, & Nagler M (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. Behavior Research Methods, 49, 513–537. 10.3758/s13428-016-0740-2 [PubMed: 27287445]

Maris E, & Oostenveld R (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of Neuroscience Methods, 164, 177–190. 10.1016/j.jneumeth.2007.03.024 [PubMed: 17517438]

Mather E, & Plunkett K (2012). The role of novelty in early word learning. Cognitive Science, 36, 1157–1177. 10.1111/j.1551-6709.2012.01239.x [PubMed: 22436081]

McKee RD, & Squire LR (1993). On the development of declarative memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 397–404. 10.1037/0278-7393.19.2.397 [PubMed: 8454964]

Morgan K, & Hayne H (2011). Age-related changes in visual recognition memory during infancy and early childhood. Developmental Psychobiology, 53, 157–165. 10.1002/dev.20503 [PubMed: 20945410]

Morrison CM, Chappell TD, & Ellis AW (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. The Quarterly Journal of Experimental Psychology Section A, 50, 528–559. 10.1080/027249897392017

Mumby DG, Gaskin S, Glenn MJ, Schramek TE, & Lehmann H (2002). Hippocampal damage and exploratory preferences in rats: memory for objects, places, and contexts. Learning & Memory, 9, 49–57. [PubMed: 11992015]

Munoz M, Chadwick M, Perez-Hernandez E, Vargha-Khadem F, & Mishkin M (2011). Novelty preference in patients with developmental amnesia. Hippocampus, 21, 1268–1276. 10.1002/hipo.20836 [PubMed: 20882542]

Musen G, & Treisman A (1990). Implicit and explicit memory for visual patterns. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 127–137. [PubMed: 2136749]

Nelson K, & Fivush R (2004). The emergence of autobiographical memory: a social cultural developmental theory. Psychological Review, 111, 486–511. 10.1037/0033-295X.111.2.486 [PubMed: 15065919]

Newcombe NS, Balcomb F, Ferrara K, Hansen M, & Koski J (2014). Two rooms, two representations? Episodic-like memory in toddlers and preschoolers. Developmental Science, 17, 743–756. 10.1111/desc.12162 [PubMed: 24628962]

Oakes LM, Baumgartner HA, Kanjlia S, & Luck SJ (2017). An eye tracking investigation of color–location binding in infants' visual short-term memory. Infancy, 22, 584–607. 10.1111/infa.12184 [PubMed: 28966559]

Oudeyer PY, & Smith LB (2016). How evolution may work through curiosity-driven developmental process. Topics in Cognitive Science, 8, 492–502. 10.1111/tops.12196 [PubMed: 26969919]

Parker S, Heathcote A, & Finkbeiner M (2020). Using evidence accumulation modeling to quantify the relative contributions of spatial attention and saccade preparation in perceptual tasks. Journal of Experimental Psychology: Human Perception and Performance, 46, 416–433. 10.1037/xhp0000723 [PubMed: 32223293]

Pascalis O, & Bachevalier J (1999). Neonatal aspiration lesions of the hippocampal formation impair visual recognition memory when assessed by paired-comparison task but not by delayed nonmatching-to-sample task. Hippocampus, 9, 609–616. 10.1002/(SICI)1098-1063(1999)9:6<609::AID-HIPO1>3.0.CO;2-A [PubMed: 10641753]

Pascalis O, de Schonen S, Morton J, Deruelle C, & Fabre-Grenet M (1995). Mother's face recognition by neonates: A replication and an extension. Infant Behavior and Development, 18, 79–85. 10.1016/0163-6383(95)90009-8

Pathman T, & Ghetti S (2014). The eyes know time: A novel paradigm to reveal the development of temporal memory. Child development, 85, 792–807. 10.1111/cdev.12152 [PubMed: 23962160]

Pinheiro J, Bates D, DebRoy S, & Sarkar D R Core Team (2018). nlme: linear and nonlinear mixed effects models. R package version 3.1–137.

Ratcliff R, Smith PL, Brown SD, & McKoon G (2016). Diffusion decision model: Current issues and history. Trends in Cognitive Sciences, 20, 260–281. 10.1016/j.tics.2016.01.007 [PubMed: 26952739]

Ratcliff R, Perea M, Colangelo A, & Buchanan L (2004). A diffusion model account of normal and impaired readers. Brain and Cognition, 55, 374–382. 10.1016/j.bandc.2004.02.051 [PubMed: 15177817]

Richmond J, Colombo M, & Hayne H (2007). Interpreting visual preferences in the visual paired-comparison task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33, 823–831. 10.1037/0278-7393.33.5.823 [PubMed: 17723062]

Richmond J, & Nelson CA (2007). Accounting for change in declarative memory: A cognitive neuroscience perspective. Developmental Review, 27, 349–373. 10.1016/j.dr.2007.04.002 [PubMed: 18769510]

Richmond J, Sowerby P, Colombo M, & Hayne H (2004). The effect of familiarization time, retention interval, and context change on adults' performance in the visual paired comparison task. Developmental Psychobiology, 44, 146–155. 10.1002/dev.10161 [PubMed: 14994266]

Robinson AJ, & Pascalis O (2004). Development of flexible visual recognition memory in human infants. Developmental Science, 7, 527–533. [PubMed: 15603285]

Rose SA (1983). Differential rates of visual information processing in full-term and preterm infants. Child Development, 54, 1189–1198. 10.2307/1129674 [PubMed: 6354626]

Rose SA, Gottfried AW, Melloy-Carminar P, & Bridger WH (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. Developmental Psychology, 18, 704–713. 10.1037/0012-1649.18.5.704

Ross-Sheehy S, Oakes LM, & Luck SJ (2011). Exogenous attention influences visual short-term memory in infants. Developmental Science, 14, 490–501. 10.1111/j.1467-7687.2010.00992.x [PubMed: 21477189]

Rossion B, & Pourtois G (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. Perception, 33, 217–236. 10.1068/p5117 [PubMed: 15109163]

Rovee-Collier C, & Cuevas K (2009). Multiple memory systems are unnecessary to account for infant memory development: an ecological model. Developmental Psychology, 45, 160. 10.1037/a0014538 [PubMed: 19209999]

Samuelson LK, Kucker SC, & Spencer JP (2017). Moving word learning to a novel space: A dynamic systems view of referent selection and retention. Cognitive Science, 41, 52–72. 10.1111/cogs.12369 [PubMed: 27127009]

Santolin C, Garcia-Castro G, Zettersten M, Sebastian-Galles N, & Saffran JR (2021). Experience with research paradigms relates to infants' direction of preference. Infancy, 26, 39–46. 10.1111/infa.12372 [PubMed: 33111438]

Sokolov EN (1963). Perception and the Conditioned Reflex. New York: Macmillan.

Tummeltshammer KS, Mareschal D, & Kirkham NZ (2014). Infants' selective attention to reliable visual cues in the presence of salient distractors. Child Development, 85, 1981–1994. 10.1111/cdev.12239 [PubMed: 24646174]

Vandekerckhove J, Tuerlinckx F, & Lee MD (2011). Hierarchical diffusion models for two choice response times. Psychological Methods, 16, 44–62. 10.1037/a0021765 [PubMed: 21299302]

Voss A, Nagler M, & Lerche V (2013). Diffusion models in experimental psychology: A practical introduction. Experimental Psychology, 60, 385–402. 10.1027/1618-3169/a000218 [PubMed: 23895923]

Wiecki TV, Sofer I, & Frank MJ (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. Frontiers in Neuroinformatics, 14. 10.3389/fninf.2013.00014

Wu R, & Kirkham NZ (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. Journal of Experimental Child Psychology, 107, 118–136. 10.1016/j.jecp.2010.04.014 [PubMed: 20627258]

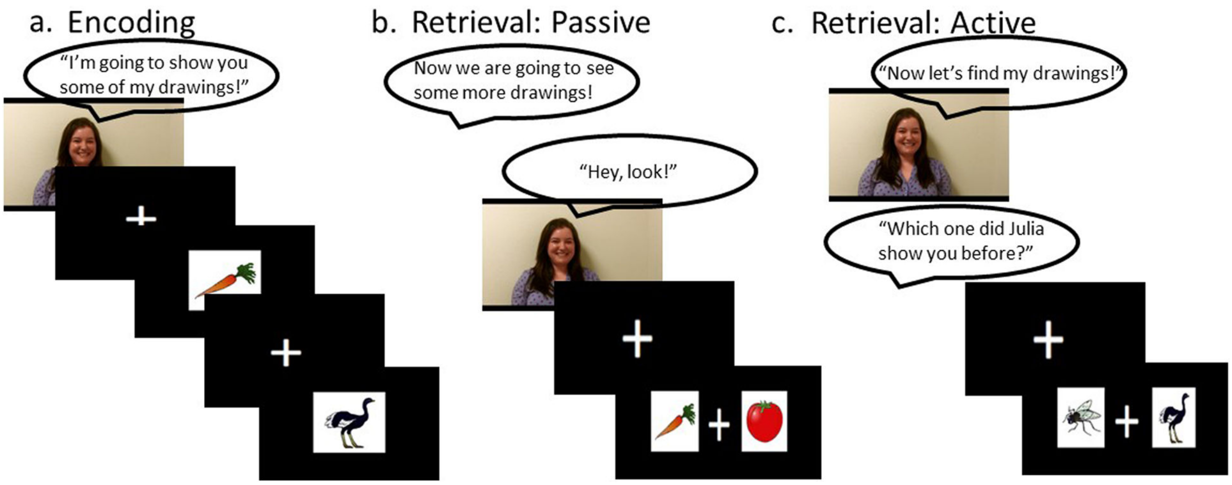

Fig 1. Eye Tracker Task Design in Experiment 1.

(a). The encoding phase for both the Passive and Active conditions. Toddlers saw a video of a female experimenter introducing the pictures as her drawings and then toddles saw one image on the screen one after another, (b). Retrieval phase of the Passive condition. Experimenter conducting the session told them that they were going to see more drawings and then toddlers saw two images (1 old and 1 new image) on the screen preceded by a video of the female experimenter saying "Hey, look" and were given no instructions other than to look at the screen, (c). Retrieval phase of the Active condition. A video of the female experimenter introduced the condition by inviting toddlers to find her drawings. Toddlers then saw pairs of images (1 old and 1 new image) and were asked by the experimenter to point to the image that they had seen previously.

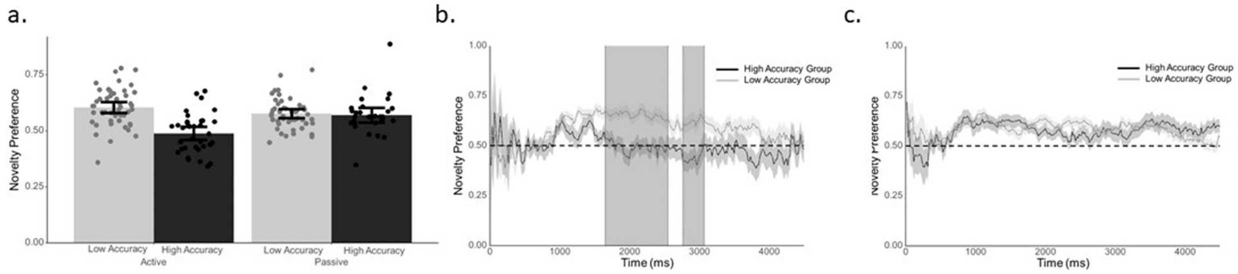

Fig 2. Novelty Preference in the Active and Passive Conditions for Experiment 1.

(a). Average novelty preference in the Active and Passive conditions as a function of accuracy group (low accuracy, in gray vs high accuracy, in black). Points represent individual data points and data are jittered on the horizontal axis to avoid stacking, (b). Novelty preference across the trial as a function of time in the Active condition as a function of accuracy group (low accuracy, in gray vs high accuracy, in black). Gray boxes indicate time in the trial where novelty preference for low and high accuracy groups were significantly different from one another. Black dotted line indicates chance looking preference, (c). Novelty preference across the trial as a function of time in the Passive condition as a function of accuracy group (low accuracy, in gray vs high accuracy, in black). The groups did not differ at any point in time. Black dotted line indicates chance looking preference. Error bars are 95% confidence intervals.

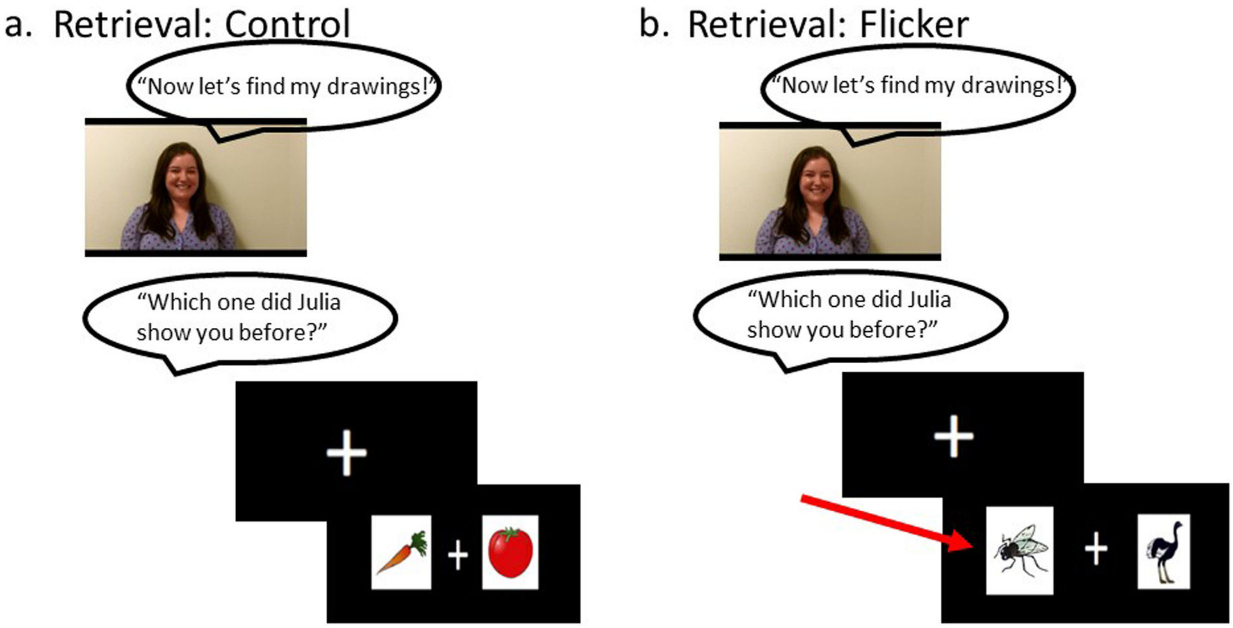

Fig 3. Attentional Manipulation Conditions for Active Condition in Experiment 2.

(a). Retrieval phase of the Control condition in the Active condition. This condition was identical to the retrieval phases of Experiment 1. Toddlers were invited by a video of the female experimenter to either look at the images (Passive condition; Not shown here) or to find her drawings and then asked to point to the picture they had seen before by the experimenter (Active condition; Shown above), (b). Retrieval phase of the Flicker condition in the Active Condition. Toddlers were invited by a video of the female experimenter to either look at the images (Passive condition; Not shown here) or to find her drawings and then asked to point to the picture they had seen before by the experimenter (Active condition; Shown above). For all trials, the old image flickered, once, 1 second into the trial. Red arrow indicates the image that was enlarged. Arrow is for demonstration purposes only. It was not included in the task.

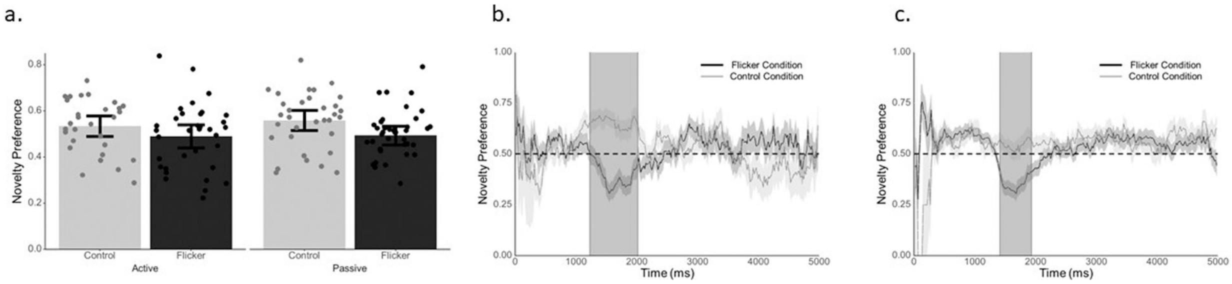

Fig 4. Novelty Preference in the Active and Passive Conditions for Experiment 2.

(a). Average novelty preference in the Active and Passive conditions as a function of attention condition (Control, in gray vs Flicker, in black). Points represent individual data points and data are jittered on the horizontal axis to avoid stacking, (b). Novelty preference across the trial as a function of time in the Active condition as a function of attention condition (Control, in gray vs Flicker, in black). Gray box indicates time in the trial where novelty preference for Control and Flicker conditions were significantly different from one another. Black dotted line indicates chance looking preference, (c). Novelty preference across the trial as a function of time in the Passive condition as a function of attention condition (Control, in gray vs Flicker, in black). Gray box indicates time in the trial where the novelty preference for Control and Flicker conditions were significantly differentfrom one another. Black dotted line indicates chance looking preference. Error bars are 95% confidence intervals.