DOI: 10.1111/infa.12530

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

Infants use contextual memory to attend and learn in naturalistic scenes

Kristen Tummeltshammer[1] | **Dima Amso**[2]

[1]Brown University, Providence, Rhode Island, USA

[2]Department of Psychology, Columbia University, New York, New York, USA

Correspondence
Dima Amso, Department of Psychology, Columbia University, (419E Schermerhorn Hall), 1190 Amsterdam Avenue, MC 5501 New York, NY 10027, USA.
Email: Da2959@columbia.edu

Funding information
National Institute of Mental Health, Grant/Award Numbers: F32-MH108278-01, R21-MH113870; National Science Foundation, Grant/Award Number: BCS-2051819

Abstract

Infants encounter new objects and learn about object features in relation to a rich and detailed visuospatial context. Using a contextual cueing task, recent work showed that 6- and 10-month-old infants search more efficiently for target objects in repeated rather than novel visuospatial contexts (i.e., arrays of shapes on a blank background). Here, we investigate whether infants' sensitivity to visuospatial context scales up to more complex and potentially more distracting, naturalistic scenes. In an eye-tracking task, 8-month-olds searched for a novel target object in colorful photographs of everyday environments (e.g., bedrooms and kitchens). Repeated ("Old") contexts co-varied with target locations, such that the target object appeared in exactly the same location on the same scene, while varying ("New") contexts contained target objects placed in different counterbalanced locations across a variety of scenes. Infants exhibited faster search times, more anticipation of target animation, and longer looking at targets that appeared in Old relative to New contexts. In a subsequent memory test, infants showed better recognition of label-object pairings for target objects that had appeared in Old, rather than New, contexts. These results indicate that infants can use visuospatial contextual information in complex naturalistic scenes to facilitate memory-guided attention and learning of object-paired labels.

1 | INTRODUCTION

Memory for the particular visuospatial contexts in which objects occur may be useful in facilitating young children's orienting of attention, object recognition, and learning of new object features. In human adults and some non-human animals, learning of associations between objects and their visuospatial contexts (called contextual cueing) has been shown to guide visual attention orienting to facilitate specific target object visual search (Brockmole & Henderson, 2006; Chun & Jiang, 1998, 1999; Goujon & Fagot, 2013; Olson & Chun, 2002; Wasserman, et al., 2014). Recent studies have shown that infants can also learn these associations and use them to increase search efficiency in simple displays, such as arrays of colorful shapes (Bertels, et al., 2016; Tummeltshammer & Amso, 2018). In other words, infants can use memory for visuospatial context to guide attention to an object. The goals of the present study were: (1) to determine whether these results would replicate in complex and potentially distracting complex naturalistic scenes, and (2) to determine whether engaging memory-guided attention supports learning of target-paired sounds or labels.

A replication and extension of infant contextual cueing, from artificial displays to naturalistic scenes, is warranted. Previous work has shown that scene dynamics may affect whether the viewer recruits memory for spatial contextual information during visual search; sometimes, target identification is faster if one searches de novo on each trial (Kunar et al., 2008; Wolfe, et al., 2011). Adult studies have found that search within complex and/or naturalistic visual scenes is more likely to engage memory-guided attention (Brockmole, et al., 2006; Brockmole & Henderson, 2006; Ehinger & Brockmole, 2008; Goujon, 2011; Goujon, et al., 2012; Hollingworth, 2009, 2012). In complex natural scenes, drawing on memories of contextual information may be more efficient than navigating many distracters in each scene in search of a target. In contrast, if the target is highly salient or the artificial display is very simple, visual search for a target may be most efficiently driven by relatively fast attention processes with little added benefit from the recruitment of memory for contextual information.

Importantly, infant dynamics between attention and memory engagement may differ from those of adults and require their own empirical consideration. The richness and complexity of natural scenes will often be correlated with the number of distracting elements, which may render infants less able to process the target-in-scene pairings and thus less likely to benefit from contextual repetition (Markant & Amso, 2022). In such a case, infants may default to a de novo search strategy on each trial, rather than one that uses visuospatial context to guide target visual search. Alternatively, the richness of natural scenes may provide robust contextual information, and like in adults, motivate the engagement of memory-guided attention. Having found that 6- and 10-month-old infants can use memory-guided attention to guide visual search in simple arrays, we next asked whether this result would extend to more complex and potentially challenging displays.

The second goal of the study was to test whether prioritized attention to the target during search in repeated visuospatial contexts would lead to better learning. The relationship between objects and their contexts has been shown to affect learning and memory processes: for example, faster recognition and enhanced identification of objects that appear in the same, rather than differing, contexts (Davenport & Potter, 2004; Oliva & Torralba, 2007; Palmer, 1975). In infants, several studies have shown that endogenously cued or prioritized attention leads to superior subsequent recognition memory (Amso & Johnson, 2006; Markant & Amso, 2013, 2016; Wu & Kirkham, 2010). Such enhancement may arise from the selective deployment of attention, leading to deeper processing of object features and stronger binding of multisensory elements (Hauer & MacLeod, 2006; Markant et al., 2015; Stokes, et al., 2012; Talsma, et al., 2010; Treisman, 1986). For example, prioritizing attention to the expected location of an event has been shown to support the binding of sensory elements from different modalities, such as sight and sound (Fiebelkorn, et al., 2010; Talsma, et al., 2010).

Word learning in particular challenges infants to make rapid multisensory associations between visual objects and their labels (Waxman & Gelman, 2009). Therefore, we asked whether infants' use of visuospatial contextual cues would facilitate learning of object-paired information, such as object labels, introduced in locations that have received attentional priority. For example, studies have demonstrated better word learning when objects were presented in predictable locations in 16- to 18-month-olds (Benitez & Smith, 2012) and alongside repeated distracter objects in 3-year-olds (Axelsson & Horst, 2014). Horst et al. (2011) found that toddlers learned more words when read the same story repeatedly than when exposed the same words across a variety of different stories. However, if an object is always observed (or its label always heard) in the same context, it may be difficult to isolate from its surroundings or to identify in a new context; thus, one could also imagine a benefit to learning new object information against a shifting background of more noisy or variable input (Twomey, et al., 2018).

We presented 8-month-olds with a contextual cueing eye-tracking task, where they saw photographs of rich complex scenes (i.e., kitchens, bedrooms, living rooms, and backyards) that contained three additional novel objects: one target and two distracters that appeared in different quadrants of the screen. Some scenes repeated throughout the experiment (e.g., the same bedroom) with the same target object always appearing in the same location and flanked by the same arrangement of distracters. Other scenes varied (e.g., a variety of different kitchens) with the target object and distracters appearing in different locations, providing a baseline for infants' visual search across entirely new contexts. Notably, all objects were presented an equal number of times, in all four quadrants, and as either targets or distracters across both New and Old visuospatial context conditions. In this way, infants could not use statistical regularity of object or location alone as cues to facilitate search. Rather, the object must be contextualized within the visuospatial background. Based on previous studies, we predicted that infants would search more efficiently for objects in repeated (Old) visuospatial contexts compared to varying (New) contexts (e.g., Bertels, et al., 2016; Tummeltshammer & Amso, 2018). To address our second goal, we paired each target object with either a unique sound or pseudo-word label (e.g., “ding-ding”, “toma”) that played as the target became animated after 2 s of search. We included both sounds and pseudo-words because, while there is evidence that infants may associate visual objects with arbitrary sounds by 7 months of age (Bahrick, 1994; Gogate & Bahrick, 1998), word-object associations are typically not reliably formed until 12–14 months (Werker, et al., 1998; Woodward, et al., 1994). Thus, we included an association that we expected 8-month-olds to learn, and another that they might be challenged by and potentially show greater benefits from additional contextual cues. Following the contextual cueing task, we tested infants' memory of the sound-object and label-object pairings using the preferential looking method (i.e., playing the sound or label to elicit looking at the matching object). If they had learned to associate the sound or label with its corresponding object, we would expect longer looking to the target than to the equally visually familiar foil (Golinkoff, et al., 2013). We predicted that “Old” contexts would support better learning of new sound-object and label-object pairings due to increased attentional priority during visual search (e.g., Axelsson & Horst, 2014; Benitez & Smith, 2012).

2 | METHOD

2.1 | Participants

Thirty-one healthy full-term 8-month-old infants participated in the experiment (16 females, M = 8 months, 4.0 days, SD = 24.4 days). Two additional 8-month-olds were tested but not included due to inattention and/or equipment failure. According to parental report of race/ethnicity, 21 participants were Non-Hispanic White, 1 was Hispanic White, 3 were Hispanic Other Race, 4 were Black, and 2 were Asian. We note that the reported results are derived from a homogeneous sample of infants and may not be generalizable to all infants. Infants were recruited via local advertisements from the greater Providence, RI area, including northern and eastern Rhode Island and southeastern Massachusetts. Data collection took place between April 2017 and January 2018. All procedures followed were in accordance with ethical standards established in the Declaration of Helsinki and were approved by the Institutional Review Board at Brown University. Written informed consent was received from a parent or guardian for each child prior to any assessment or data collection. Families received compensation for their time and travel.

Sample size was determined based on the size of the effect of context (Old, New) on infants' search reaction time (RT) latency in Tummeltshammer and Amso (2018), which was Cohen's $d = 0.57$, a moderate effect. For a paired comparison test at alpha level 0.05 and power level 0.8, and allowing for an attrition rate of 20%, the minimum sample size was estimated to be $N = 27$ participants. Due to counterbalancing of contexts and target objects, we continued data collection until an equal number of infants had viewed each counterbalanced set, arriving at the final sample of 31 infants.

2.2 | Apparatus and stimuli

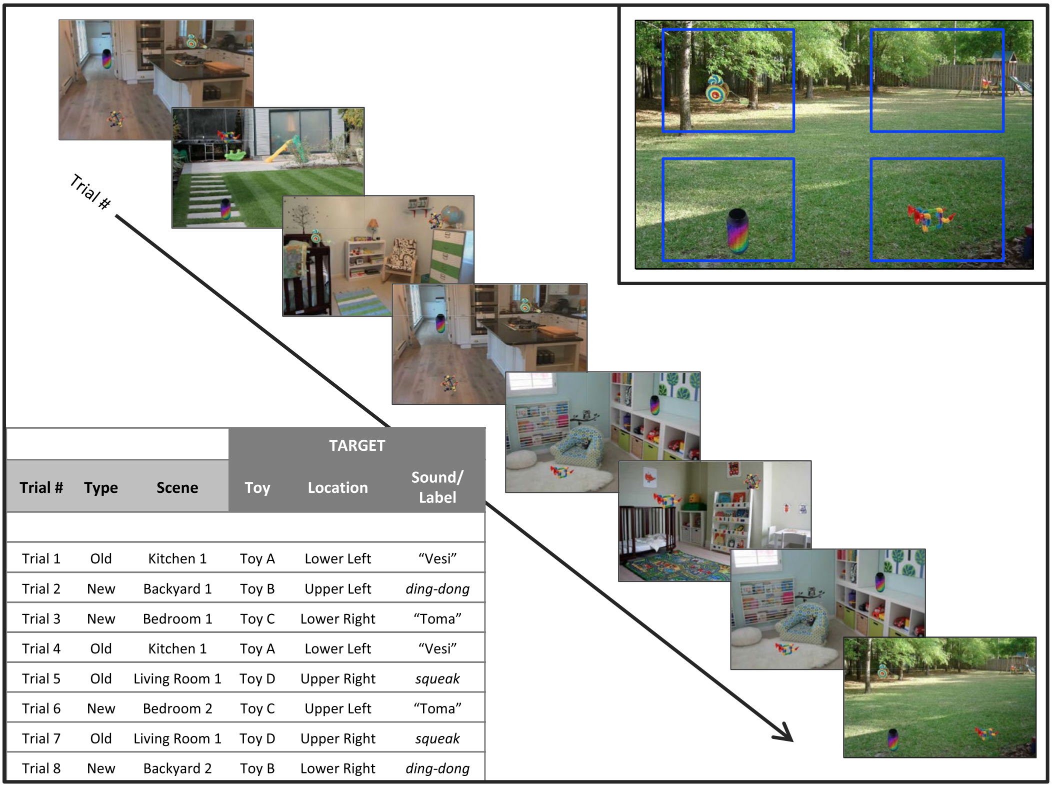
Eye movements were recorded using a remote eye tracker (SensoMotoric Instruments RED system) with a 22″ monitor. Stimuli were presented using the SMI Experiment Center software at a resolution of 1920 × 1080 pixels, and sounds were played through external stereo speakers. A digital video camera with infrared night vision (Canon ZR960) was placed above the monitor to observe and record infants' head movements.

Infants were presented with photographs of everyday environments: kitchens, children's bedrooms, living rooms, and backyards. The scenes were chosen to represent a typical, natural infant environment and were not edited for a precise balance of features; however, when necessary, the images were cropped or adjusted in Photoshop to control resolution (300 dpi), size (1680 by 1050 pixels), and average brightness (100 on a scale of 0–256). On each background scene, infants were presented with three of four possible unfamiliar toy-like objects (Figure 1), which were arranged in unique quadrants of the screen in order to avoid ambiguity in coding infants' looks. On each trial, the objects appeared static for 2 s, and then only the target object became animated for 4 s, looming within its quadrant as its unique sound or label played (non-speech sounds: *ding-ding*, *squeak*; pseudo-words: “toma”, “vesi”). The stimuli were edited and animated using Adobe Flash and Premiere Pro software packages. Sample trial videos, as well as the scene image files, are publicly available in a Databrary repository [nyu.databrary.org/volume/1367].

2.3 | Design and procedure

Infants were tested individually in a quiet room, seated 60 cm from the monitor on their caregiver's lap. A looming calibration stimulus was presented at five points (the four corners and center of the screen) to obtain the infant's point-of-gaze and validated at a minimum of four points to ensure accuracy. Average deviation was 1.84° (SD = 1.4°), suitable for assessing eye movements within the specified areas of interest.

Following successful calibration and before each trial, a colorful attention-grabbing stimulus drew infants' fixation to the center of the screen. The experimenter manually initiated each trial after



			TARGET		
Trial #	Type	Scene	Toy	Location	Sound/ Label
Trial 1	Old	Kitchen 1	Toy A	Lower Left	"Vesi"
Trial 2	New	Backyard 1	Toy B	Upper Left	*ding-dong*
Trial 3	New	Bedroom 1	Toy C	Lower Right	"Toma"
Trial 4	Old	Kitchen 1	Toy A	Lower Left	"Vesi"
Trial 5	Old	Living Room 1	Toy D	Upper Right	*squeak*
Trial 6	New	Bedroom 2	Toy C	Upper Left	"Toma"
Trial 7	Old	Living Room 1	Toy D	Upper Right	*squeak*
Trial 8	New	Backyard 2	Toy B	Lower Right	*ding-dong*

FIGURE 1 Example block of 8 trials. Here, the kitchen and living room scenes repeat on each "Old" trial with the targets always in the same locations, while a new bedroom and backyard are presented on each "New" trial with the targets in different counterbalanced locations. *Inset:* Example scene with AOIs in blue (used for analysis, not visible to participants).

ensuring the infant's fixation. The experiment consisted of a contextual cueing task, immediately followed by a preferential looking memory test, and lasted approximately 8 min.

2.3.1 | Contextual cueing task

All infants were exposed to two toy-like objects in Old contexts, which consisted of a specific scene with a fixed configuration of target and distracter objects and other scene background elements that repeated 12 times throughout the experiment. Thus, the target always appeared in the same location within each Old context. All infants were also exposed to two toy-like objects in New contexts, which consisted of 12 different scenes with different counterbalanced configurations of target and distracter objects and varying scene background elements. New scenes were never repeated and the target's location could not be predicted, providing a baseline of infants' visual search behavior in an entirely new visuospatial context.

To rule out any location or item probability effects, the four possible target objects appeared equally often in each quadrant throughout the experiment, and each object appeared as a distracter as often as a target. Hence, infants could not predict the identity or location of the target based on likelihood alone; any difference in performance could only be attributed to learning the visuospatial *context* in which the object appeared. Further, combinations of target object, location, background scene, and paired sound or label were counterbalanced across infants, such that targets repeated in Old contexts for some infants were presented in New contexts for other infants and vice versa. All infants were exposed to four trial types: Old Context/Sound, New Context/Sound, Old Context/Label, and New Context/Label. A total of 48 trials (4 contexts *x* 12 trials each) were presented in blocks of 8 trials (2 per context, see example in Figure 1) with a cartoon break inserted every 2 blocks (i.e., a 10-s clip of Sesame Street).

2.3.2 | IPLP recognition memory test

Following this exposure, infants were tested on the sound-object and label-object pairings using the Intermodal Preferential Looking Paradigm (IPLP). Infants viewed two objects side-by-side on a blank screen for 6 s while the sound or label corresponding to one of the objects played. The post-test consisted of 8 trials presented in 2 blocks with a brief cartoon interlude (2 Old/Old trials, 2 Old/New trials, 2 New/Old trials, and 2 New/New trials). An Old/Old IPLP trial, for example, consisted of two objects that had been presented as targets in the two Old visuospatial contexts during the contextual cueing task, but the sound or label played would match only one of the objects. Although all objects and sounds/labels were equally familiar, having appeared as targets an equal number of times during contextual cueing, the expectation is that infants who have learned the audiovisual pairing should allocate more attention to the object that matches the sound or label they hear (Golinkoff, et al., 2013).

2.4 | Data analysis

Eye movements were separated into discrete fixations using a temporal filter of 80 ms and a spatial filter of 150 pixels (equal to 3.63° visual angle). Areas of interest were uniformly delineated around the four quadrants of the screen (see Figure 1 inset), and fixations landing in the target AOIs were coded for their RT latency and duration.

The following contextual cueing task-dependent variables were computed to examine visual search in Old/New contexts: (1) Mean RT latency to fixate the target; (2) mean proportion of trials in which the target was fixated before it animated; (3) duration of looking at the target (as a proportion of total looking at the screen) *prior* to its animation; and (4) duration of looking at the target (as a proportion of total looking at the screen) *after* it animated and the sound or label played. Previous experiments have taken decreased latency RTs and higher rates of anticipatory looking as evidence of gains in spatiotemporal knowledge (e.g., Amso & Johnson, 2006; Kirkham, et al., 2007; Markant & Amso, 2013). Trials were excluded if the infant had no fixations within the target AOI. Infants supplied an average of 10.06 valid trials per condition (out of 12; range 4–12) and an average of 42.04 valid trials across all conditions (out of 48, $SD = 6.9$). Of the $N = 31$ tested infants, data from 3 infants were excluded due to an insufficient number of valid trials in a single condition (<4) or across all conditions (<24), resulting in a final sample of $N = 28$ infants included in the contextual cueing task data analyses.

For the IPLP memory test, areas of interest were delineated around the two objects (see Figure 4a) and all fixations landing in an AOI during the analysis window of 1000–6000 ms were summed as the measure of looking time to that object. Scores on the IPLP memory test were calculated as: (Looking Time to Matching Object—Looking Time to Non-matching Object)/(Total Looking Time). Thus, a positive score indicates more looking to the object that matched the sound or label, while a negative score indicates more looking to the non-matching object. Trials were excluded if the infant failed to fixate the display for a minimum of 500 ms during the analysis window. Infants viewed 2 trials of each condition and valid trials of the same type were averaged. Of the $N = 31$ tested infants, two infants were missing data from only one of the four IPLP conditions. Those missing values were replaced with group averages for those conditions.

3 | RESULTS

3.1 | Contextual cueing task: Visual search RTs

Mean RT latencies to targets were compared in a Context (Old, New) by Sound Type (Sound, Label) repeated measures ANOVA. Results show a significant main effect of Context, F (1,27) = 7.06, $p = 0.013$, $\eta_p^2 = 0.21$, and no effect of Sound Type or interaction with Context (all $p > 0.478$). Infants were faster to locate targets that appeared in Old contexts than in New contexts as shown in Figure 2a. Nineteen out of 28 infants had faster mean RT latencies to targets in Old compared to New contexts.

Next we examined whether faster search RT latencies had emerged through exposure to repeated presentations of the target objects in Old compared to New visuospatial contexts. Given that Old and New context conditions were interleaved across 48 trials of exposure, the variable "Exposure Number" (Figure 2b) refers to the condition-specific repetition of each trial type, rather than the overall trial number. Although we binned across 3 exposures for illustrative purposes in Figure 2b, data points were not binned or averaged in the analysis. Latencies were analyzed by fitting a linear mixed-effects model in R (R Core Team, 2020) using the lme4 package (Bates, et al., 2015). The model included fixed effects of Exposure Number and Context by Exposure Number interaction, as well as the random effect of Participant with Context as a random slope variable. Estimates of coefficients, standard errors, and corresponding t-statistics for the model are presented in Table 1, along with their estimated significance. The model showed a significant main effect of Exposure Number, Type III F (1,105.8) = 7.75, $p = 0.006$, as well as a significant interaction of Context by Exposure Number, Type III F (1,64.5) = 8.16, $p = 0.006$. As shown in Figure 2b, the interaction indicates a divergence in infants' response times across Old and New contexts and a relative *increase* in search efficiency with exposure to contextual regularities.

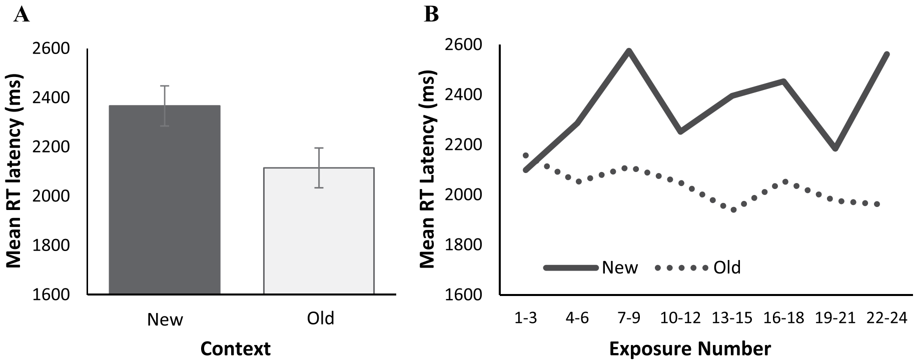


FIGURE 2 (a) Mean reaction time (RT) latency to fixate the target when presented in Old and New contexts. Error bars indicate standard error of the mean. (b) Mean change in RT latency across contexts, collapsed across Sound and Label trials.

TABLE 1 Summary of linear mixed-effects model for change in reaction time (RT) latencies with exposure

Fixed effects	Estimate	Std. Error	*t*-value	*p*-value
(Intercept)	2174.11	90.60	24.00	<0.001
Exposure number	34.19	12.10	2.83	0.010
Context * exposure number	−20.70	7.11	−2.91	0.008
Random effects	**Variance**		**Std. Dev. [CI]**	
Participant (Intercept)	393157		627.0 [305.0, 972.2]	
Context Participant	116600		341.5 [120.4, 552.4]	
Residual	1391417		1179.6 [1130.6, 1230.8]	
Number of observations: 1126, participants: 28				

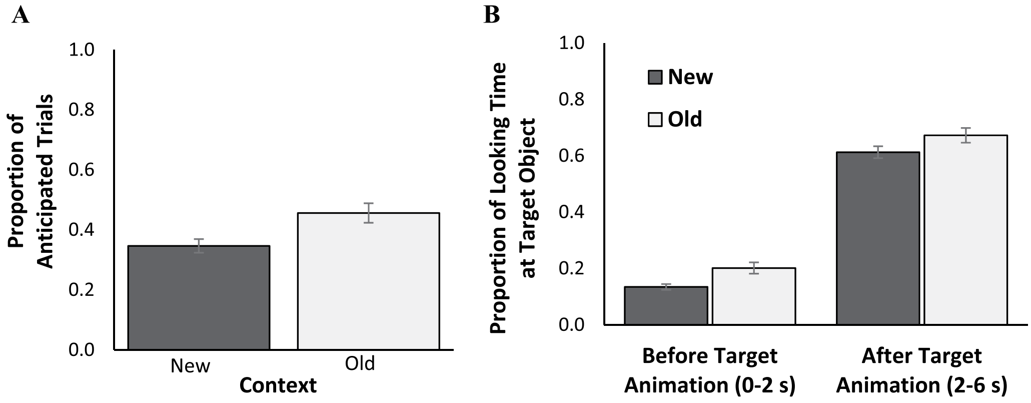


FIGURE 3 (a) Mean proportion of trials in which infants anticipated the target's animation. (b) Mean proportion of looking time to target object, relative to total looking time to the entire scene, before and after its animation at the 2-s mark. Error bars indicate standard error of the mean.

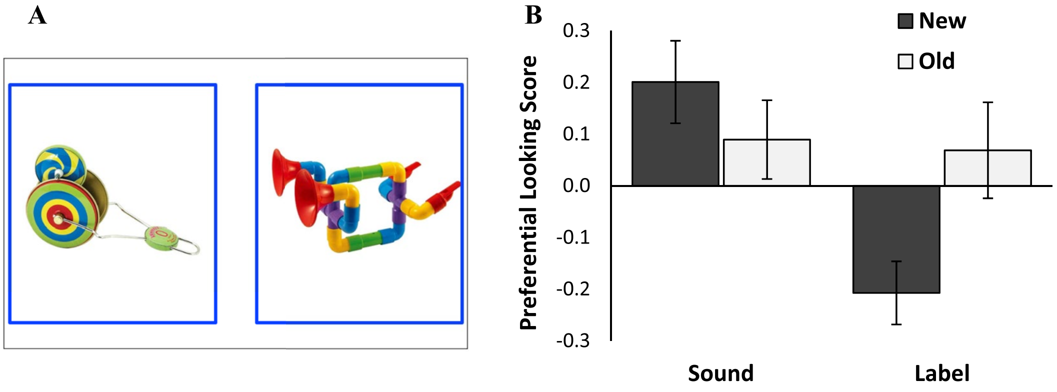


FIGURE 4 (a) Example Intermodal Preferential Looking Paradigm (IPLP) display with AOIs in blue (used for analysis, not visible to participants). (b) IPLP scores for objects associated with sounds or labels that had appeared in Old versus New contexts. Error bars indicate standard error of the mean.

3.2 | Contextual cueing task: Anticipation of target animation

For convergent evidence of contextual cueing, we examined the proportion of trials in which infants anticipated the target's animation, fixating it prior to becoming animated at the 2-s mark. A Context (Old, New) by Sound Type (Sound, Label) repeated measures ANOVA showed that infants anticipated more in Old contexts than in New contexts, $F\ (1{,}27) = 9.72$, $p = 0.004$, $\eta_p^2 = 0.27$ (Figure 3a). There was no effect of Sound Type or interaction with Context (all $p > 0.159$). Twenty out of 28 infants had a higher proportion of anticipated target animations in Old compared to New contexts.

3.3 | Contextual cueing task: Duration of looking at targets

We also considered whether context had an effect on infants' duration of looking at the target objects, comparing mean looking time to the target as a proportion of total looking time to the entire scene. We expected that a consistent context would be more likely to influence infants' attention during search, that is, *before* the target animated, while the presence of a sound or label would be more likely to influence infants' attention during the sound/labeling event, that is, *after* the target animated. Results of a Context (Old, New) by Sound Type (Sound, Label) by Time Interval (Before Animation, After Animation) repeated measures ANOVA showed a significant main effect of Context, $F\ (1{,}27) = 9.77$, $p = 0.004$, $\eta_p^2 = 0.27$, as infants looked longer at targets in Old compared to New contexts. There was also a main effect of Time Interval, $F\ (1{,}27) = 569.27$, $p < 0.001$, $\eta_p^2 = 0.96$, as infants looked longer at targets *after* they became animated. However, there was no significant effect of Sound Type or significant interactions of Sound Type, Context, and Time Interval (all $p > 0.256$), indicating that the effect of Context was present across both time intervals and Sound Type conditions (Figure 3b). Indeed, separate planned comparisons for the intervals before and after target animation showed significant differences in looking time between Old and New contexts both during the Before Animation interval (Old $M = 0.20$, $SE = 0.01$; New $M = 0.14$, $SE = 0.02$; $t\ (27) = 3.06$, $p = 0.005$) and during the After Animation interval (Old $M = 0.67$, $SE = 0.02$; New $M = 0.61$, $SE = 0.02$; $t\ (27) = 2.23$, $p = 0.034$). Twenty-two out of 28 infants looked longer at targets in Old compared to New contexts before they animated, and 19 out of 28 infants looked longer at targets in Old contexts after they animated.

As an added manipulation check, we also computed mean proportions of looking time to the target relative to the other two distracter objects when all objects were still (as opposed to relative to total looking at the entire scene) for comparison against 33% chance. In New contexts, proportions of looking to the target before it animated did not significantly differ from chance ($M = 0.35$, $SE = 0.02$, $t\ (27) = 1.04$, $p = 0.308$); infants directed attention to all three objects similarly. Instead, in Old contexts, proportions of looking to the target before it animated were significantly greater than chance ($M = 0.41$, $SE = 0.03$, $t\ (27) = 2.65$, $p = 0.013$) as infants looked longer at the target object than at the distracter objects.

3.4 | IPLP sound/object pairing recognition memory test

To examine the effect of repeated context on learning of target-paired sound/label information, we compared scores on the IPLP recognition memory test (where a positive score indicates longer looking to the object correctly paired with the sound or label). A Context (Old, New) by Sound Type (Sound, Label) repeated measures ANOVA showed a significant main effect of Sound Type, $F\ (1{,}30) = 6.35$, $p = 0.017$, $\eta_p^2 = 0.18$, as sound-object pairings were recognized better than label-object pairings (Figure 4a). Moreover, infants recognized the object with which a sound was paired above what would

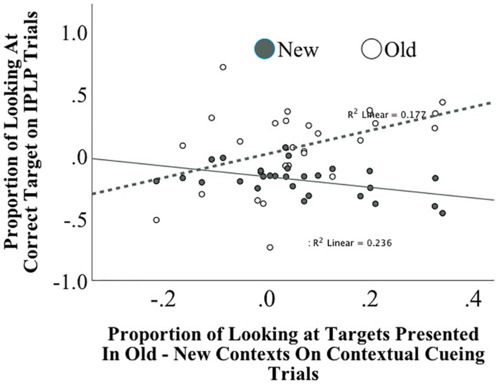


FIGURE 5 Illustrates the relationship between attention, here proportion of looking time, to the target after it animated when presented in a repeated Old context or in a variety of New contexts (*x*-axis), and proportion of looking at the matching labeled object on Intermodal Preferential Looking Paradigm (IPLP) memory trials.

be expected by chance ($M = 0.15$, $SE = 0.05$, t (30) = 2.76, $p = 0.010$), but did not recognize the object with which a label was paired more or less than would be expected by chance ($M = -0.07$, $SE = 0.06$, t (30) = 1.19, $p = 0.245$). This main effect of Sound Type was qualified by a significant Context by Sound Type interaction, F (1,30) = 7.59, $p = 0.010$, $\eta_p^2 = 0.20$. Figure 4b shows that infants recognized the sound-object pairings to a similar extent regardless of whether they had been presented in Old or New contexts (Old $M = 0.09$, $SE = 0.08$; New $M = 0.20$, $SE = 0.08$; t (30) = 0.97, $p = 0.341$). In contrast, there was a significant difference in recognition of label-object pairings depending on whether the pairing was situated in an Old or New context (Old $M = 0.07$, $SE = 0.09$; New $M = -0.21$, $SE = 0.06$; t (30) = 2.62, $p = 0.014$).

Learning a novel label-object pairing in a repeated (Old) context seemed to boost infants' later recognition of that label-object pairing compared to when it had been learned in a variety of different (New) contexts. In the latter case, infants had a significant preference to look at the foil object that did not match the label. We asked whether the differences in recognition memory by Context in the Label condition were linked to visual attention during the contextual cueing task. We conducted a general linear model on IPLP scores in the Label condition and included the following 6 continuous variables: Target Looking Duration (proportion of time spent looking at the target after it animated and its label was played for 4 s) on (1) Label/Old trials and on (2) Label/New trials, and (3) the interaction of Target Looking Duration on Label/Old by Target Looking Duration on Label/New context trials; RT Latency to the target on (4) Label/Old trials and on (5) Label/New trials, and (6) the interaction of RT Latency on Label/Old by RT Latency on Label/New context trials. The analysis resulted in a main effect of Context, F (1,21) = 5.77, $p = 0.026$, $\eta_p^2 = 0.22$. Label/Old context performance on the IPLP post-test did not differ from chance ($M = 0.07$, $SE = 0.09$, t (30) = 0.74, $p = 0.465$), whereas Label/New context performance was significantly below chance ($M = -0.21$, $SE = 0.06$, t (30) = −3.39, $p < 0.002$), indicating significant looking at the non-matching object when the label paired with a target situated in a variable (New) contexts was played.

The analysis also yielded significant interactions of Context by Target Looking Duration on Label/Old trials, F (1,21) = 7.95, $p = 0.010$, $\eta_p^2 = 0.28$, Context by Target Looking Duration on Label/New trials, F (1,21) = 6.47, $p = 0.019$, $\eta_p^2 = 0.24$, and a three-way interaction of Context by Target Looking Duration on Label/New trials by Target Looking Duration on Label/Old trials, F (1,21) = 6.86, $p = 0.016$, $\eta_p^2 = 0.25$. Figure 5 illustrates this result using unstandardized predicted values from the model. Infants who looked longer at targets in Old contexts during the contextual cueing task (ostensibly offering greater opportunity for learning when the animation and labeling event occurred) tended to have higher recognition scores for those targets' label-object pairings *and at the same time* preferred the non-matching foil when presented with labels that had been paired with target objects in New contexts.

4 | DISCUSSION

Visual search is affected by the viewer's familiarity with a scene or space. Memories of the visual environment can be drawn upon to deploy attention efficiently and prioritize locations that were important in the past (Oliva & Torralba, 2007). The present study has demonstrated that 8-month-old infants are sensitive to contextual regularity and engage memory-guided attention when tasked to search for target objects in complex naturalistic scenes. Further, the engagement of memory-guided attention has been found to enhance processing of information presented at the locus of attention, leading to faster recognition, deeper encoding, and stronger learning (e.g., Benitez & Smith, 2012). Consistently, we observed that infants' sensitivity to visuospatial context affected their learning of novel object labels with better recognition of label-object pairings that had been presented repeatedly in the same context.

Our first prediction, which infants would search more efficiently for objects appearing in repeated scenes, was supported by multiple metrics. Infants oriented faster to targets and tended to anticipate their animation more often in Old contexts than in New contexts (Figures 2 and 3a). Perhaps a consequence of increased visual search efficiency in familiar contexts, infants also looked longer at targets in repeated contexts than in varying contexts, and this effect was apparent both before and after the targets became animated (Figure 3b). This result is consistent with the findings of Bertels et al. (2016) who observed longer looking to repeated compared to novel search arrays. Further, these results extend the findings of Tummeltshammer and Amso (2018) from simple artificial displays to more complex photographs of naturalistic scenes. As noted in the introduction, scene dynamics can alter whether it is most efficient to use memory-guided attention or to search de novo on each trial. Adult studies show that memory-guided attention is more likely to be engaged in natural than artificial scenes and in testing contexts that necessitate head and eye movements (Võ & Wolfe, 2012, 2015). In infants, complex scene structures might pose a different challenge: The richness and complexity of the natural scene structure could have been too *distracting* for infants to efficiently learn coherent item-in-context associations and use memory-guided attention (Markant & Amso, 2022). However, 8-month-old infants indeed became faster to detect targets presented in repeated contexts even in our more complex scenes.

We note the artificiality of using computer programs to superimpose objects onto our scenes as a limitation of this work. Superimposing objects onto the scene photographs is not ideal; however, it was necessary to provide the experimental control we required (e.g., ensuring objects were equidistant from the center, counterbalanced so that they appeared equally often in the 4 quadrants, appearing as targets in some scenes and as distracters in others). Scenes were chosen to allow the integration of objects into the background elements as naturally as possible (e.g., resting on a shelf). We also note a precedent established by other studies, which have examined visual search in naturalistic scenes using either target stimuli artificially embedded in natural background (Brockmole & Henderson, 2006; Ehinger & Brockmole, 2008; Goujon, 2011; Henderson, et al., 2009) or computer-rendered illustrations of real-world scenes (Brockmole, et al., 2006; Hollingworth, 2009, 2012). We believe that the complexity of our displays was sufficient to test the primary question of whether infants would exhibit contextual cueing effects in richer, more naturalistic scenes.

Moreover, evidence from adults indicates that contextual cueing may be accomplished on both local and global scales (i.e., in reference to specific items or to an entire display; Bar, 2004; Mack & Eckstein, 2011; Torralba et al., 2006), although there is some indication that scene-based cues may overshadow item or array-based cues when both are presented (Brooks, et al., 2010; Rosenbaum & Jiang, 2013). In our study, a number of features defined visuospatial context and thus covaried with the location of the target on Old context trials, including the scene background elements and the configuration of target and distracter objects contextualized within the global scene. This study cannot pinpoint whether or to what extent each of these visuospatial redundancies was used by infants for increasing visual search efficiency. It is also not clear that infants would parse the scene into a local spatial array of target and distracter objects separated from the repeated arrangement of other elements in the background. Having established the general value of contextual cues here, future work within this paradigm may focus on isolating whether and how infants use these local and global contexts to guide attention.

Our second prediction, that a consistent context would boost learning of new sound-object and label-object pairings, was partially supported by infants' performance on the recognition memory post-test. Namely, we observed an interaction, such that the effect of context depended on whether the target-paired information was a label or a non-speech sound. On the IPLP post-test, infants recognized novel sound-object pairings at above-chance levels regardless of whether they had been learned in Old or New contexts. However, their recognition of label-object pairings differed significantly between labels that had been learned in Old and New contexts. Specifically, when the label-object pairing had been learned in a repeated (Old) context, infants showed a significant preference for the matching labeled object, whereas they looked longer at the non-matching foil object when the label-object pairing had been learned across a variety of (New) contexts.

One explanation for the difference in the effect of context on learning of sound-object and label-object associations is their complexity and the relative challenge they pose for our 8-month-old participants. While there is evidence that infants may associate visual objects with arbitrary sounds by 7 months of age (Bahrick, 1994; Gogate & Bahrick, 1998), word-object associations are typically not reliably formed until 12–14 months (Werker, et al., 1998; Woodward, et al., 1994). At 13 months, infants seem to assign sounds and labels to objects equally well (Campbell & Namy, 2003; Woodward & Hoyne, 1999). Thus, the relative ease at which 8-month-olds were able to associate non-speech sounds with target objects may have precluded any influence of contextual regularity on their learning, whereas the difficulty of word-object associations at this age offered a better opportunity to measure its effect. Examining these conjectures with younger and older infants is warranted.

Figure 5 shows that longer looking to the label-object pairing, when presented in repeated Old relative to varying New contexts, was correlated with better recognition memory for that pairing on the IPLP post-test. This result is fairly straightforward. However, longer looking to the label-object pairing in Old relative to New contexts was *also* correlated with more looking at the *non-matching* object on IPLP trials for which the label-object pairing had been learned in New contexts. Recall that the label-object pairings were all equally familiar with the only difference being the context in which they occurred. While a preference for the non-matching object was unexpected, it is consistent with a number of studies that have found shifting preferences as a result of encoding strength (e.g., Bahrick, et al., 1997; Richmond & Nelson, 2009). For example, Bahrick and colleagues observed that over a longer retention interval, null preferences for intermediately encoded visual memories shifted to novelty preferences. According to this interpretation, if infants had only weakly or partially encoded label-object pairings presented in the New context condition, they may have shifted attention away from the matching target object in the IPLP post-test. That is, infants who took advantage of the repeated context to engage memory-guided attention may have found the New context condition comparably challenging or distracting, thereby resulting in weakly or partially encoded label-object pairings and hindering word learning. Future work is needed to clarify this finding and its interpretation.

Our results support a pathway through which presenting a new label-object pairing in a stable, repeated context prioritizes infants' attention to the labeled object, leading to longer looking and ostensibly stronger encoding, which in turn enables more robust context-independent recognition in the future. Such a mechanism requires the coordination of both memory processes (e.g., recognition of the scene elements and retrieval of learned knowledge of the environment's structure) and attentional guidance processes (e.g., sensitivity to salient sensory input and top-down control in line with task demands or goals), which are still developing in infancy and early childhood. There is evidence of both implicit memory and long-term retention of learned contextual information in human infants (Cuevas & Sheya, 2019; Rovee-Collier & Cuevas, 2009); however, whether they engage episodic memory and to what extent memory skills are hippocampally or cortically mediated remain points of debate (Gomez & Edgin, 2016). To our knowledge, the earliest evidence of hippocampal activation in such a contextual memory task is in 3-year-olds (Prabhakar, et al., 2018). In this study, Prabhakar and colleagues asked toddlers to play with two separate toys in two separate rooms, while a novel song played in each room. Having probed the children about which room each toy was in, a subsequent fMRI session revealed that better memory for item-in-context was associated with greater hippocampal activation for the song that had played in the corresponding context. The present study contributes to this distinguished literature on infant memory by offering behavioral evidence, in 8-month-olds, of memory for past events as well as elements of the spatial context in which they happened.

To conclude, our data suggest that infants can use contextual memory to prioritize attention to locations that were important in the past, leading to faster visual search and longer looking times. Further, this prioritization of attention in stable, repeated contexts may promote better learning and recognition of target-paired information, such as novel label-object pairings. This work has important implications for understanding developing attention and memory systems, and for possible interventions in a variety of neurodevelopmental disorders where spatial memory is affected.

ACKNOWLEDGMENTS

We thank Diego Placido for help with recruitment and Bella Ehrlich for help with stimulus development. This work was funded in part by an R21 MH113870 (to DA), NSF BCS-2051819 (to DA) and an NRSA fellowship (1-F32-MH108278-01) to KT from the National Institutes of Health. The authors declare that they have no conflicts of interest with regard to the funding sources for this study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the authors upon request. Stimuli used in this study are publicly available in a Databrary repository [nyu.databrary.org/volume/1367].

ORCID

Kristen Tummeltshammer https://orcid.org/0000-0003-1745-8854
Dima Amso https://orcid.org/0000-0001-6798-4698

REFERENCES

Amso, D., & Johnson, S. P. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology*, *42*(6), 123–1245. https://doi.org/10.1037/0012-1649.42.6.1236

Axelsson, E. L., & Horst, J. S. (2014). Contextual repetition facilitates word learning via fast mapping. *Acta Psychologica*, *152*, 95–99. https://doi.org/10.1016/j.actpsy.2014.08.002

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology*, *6*(2), 111–123. https://doi.org/10.1207/s15326969eco0602_2

Bahrick, L. E., Hernandez-Reif, M., & Pickens, J. N. (1997). The effect of retrieval cues on visual preferences and memory in infancy: Evidence for a four-phase attention function. *Journal of Experimental Child Psychology*, *67*(1), 1–20. https://doi.org/10.1006/jecp.1997.2399

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benitez, V. L., & Smith, L. B. (2012). Predictable locations aid early object name learning. *Cognition*, *125*(3), 339–352. https://doi.org/10.1016/j.cognition.2012.08.006

Bertels, J., San Anton, E., Gebuis, T., & Destrebecqz, A. (2016). Learning the association between a context and a target location in infancy. *Developmental Science*, *20*(4), 1–10. https://doi.org/10.1111/desc.12397

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 699–706. https://doi.org/10.1037/0278-7393.32.4.699

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. https://doi.org/10.1080/13506280500165188

Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: Evidence from contextual cueing. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1406–1418. https://doi.org/10.1037/a0019257

Campbell, A. L., & Namy, L. L. (2003). The role of social-referential context in verbal and nonverbal symbol learning. *Child Development*, *74*(2), 549–563. https://doi.org/10.1111/1467-8624.7402015

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365. https://doi.org/10.1111/1467-9280.00168

Cuevas, K., & Sheya, A. (2019). Ontogenesis of learning and memory: Biopsychosocial and dynamical systems perspectives. *Developmental Psychobiology*, *61*(3), 402–415. https://doi.org/10.1002/dev.21817

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

Ehinger, K. A., & Brockmole, J. R. (2008). The role of color in visual search in real-world scenes: Evidence from contextual cueing. *Perception and Psychophysics*, *70*(7), 1366–1378. https://doi.org/10.3758/PP.70.7.1366

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex*, *20*(1), 109–120. https://doi.org/10.1093/cercor/bhp083

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–149. https://doi.org/10.1006/jecp.1998.2438

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science*, *8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Gomez, R. L., & Edgin, J. O. (2016). The extended trajectory of hippocampal development: Implications for early memory development and disorder. *Developmental Cognitive Neuroscience*, *18*, 57–69. https://doi.org/10.1016/j.dcn.2015.08.009

Goujon, A. (2011). Categorical implicit learning in real-world scenes: Evidence from contextual cueing. *The Quarterly Journal of Experimental Psychology*, *64*(5), 920–941. https://doi.org/10.1080/17470218.2010.526231

Goujon, A., Brockmole, J. R., & Ehinger, K. A. (2012). How visual and semantic information influence learning in familiar contexts. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1315–1327. https://doi.org/10.1037/a0028126

Goujon, A., & Fagot, J. (2013). Learning of spatial statistics in nonhuman primates: Contextual cueing in baboons (Papio papio). *Behavioural Brain Research*, *247*, 101–109. https://doi.org/10.1016/j.bbr.2013.03.004

Hauer, B. J., & MacLeod, C. M. (2006). Endogenous versus exogenous attentional cueing effects on memory. *Acta Psychologica*, *122*(3), 305–320. https://doi.org/10.1016/j.actpsy.2005.12.008

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, *9*(1), 32. https://doi.org/10.1167/9.1.32

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, *17*(1–2), 273–291. https://doi.org/10.1080/13506280802193367

Hollingworth, A. (2012). Task specificity and the influence of memory on visual search: Comment on Võ and Wolfe (2012). *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1596–1603. https://doi.org/10.1037/a0030237

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 17. https://doi.org/10.3389/fpsyg.2011.00017

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559–1571. https://doi.org/10.1111/j.1467-8624.2007.01083.x

Kunar, M. A., Flusberg, S., & Wolfe, J. M. (2008). The role of memory and restricted context in repeated visual search. *Perception & Psychophysics*, *70*(2), 314–328. https://doi.org/10.3758/pp.70.2.314

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, *11*(9), 1–16. https://doi.org/10.1167/11.9.9

Markant, J., & Amso, D. (2013). Selective memories: Infants' encoding is enhanced in selection via suppression. *Developmental Science*, *16*(6), 926–940. https://doi.org/10.1111/desc.12084

Markant, J., & Amso, D. (2016). The development of selective attention orienting is an agent of change in learning and memory efficacy. *Infancy*, *21*(2), 154–176. https://doi.org/10.1111/infa.12100

Markant, J., & Amso, D. (2022). Context and attention control determine whether attending to competing information helps or hinders learning in school-aged children. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(1), e1577. https://doi.org/10.1002/wcs.1577

Markant, J., Worden, M. S., & Amso, D. (2015). Not all attention orienting is created equal: Recognition memory is enhanced when attention orienting involves distractor suppression. *Neurobiology of Learning and Memory*, *120*, 28–40. https://doi.org/10.1016/j.nlm.2015.02.006

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. https://doi.org/10.1016/j.tics.2007.09.009

Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*(3), 273–302. https://doi.org/10.1080/13506280042000162

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition*, *3*(5), 519–526. https://doi.org/10.3758/BF03197524

Prabhakar, J., Johnson, E. G., Nordahl, C. W., & Ghetti, S. (2018). Memory-related hippocampal activation in the sleeping toddler. *Proceedings of the National Academy of Sciences*, *115*(25), 6500–6505. https://doi.org/10.1073/pnas.1805572115

Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: Evidence from eye tracking. *Developmental Science*, *12*(4), 549–556. https://doi.org/10.1111/j.1467-7687.2009.00795.x

Rosenbaum, G. M., & Jiang, Y. V. (2013). Interaction between scene-based and array-based contextual cueing. *Attention, Perception, and Psychophysics*, *75*(5), 888–899. https://doi.org/10.3758/s13414-013-0446-9

Rovee-Collier, C., & Cuevas, K. (2009). Multiple memory systems are unnecessary to account for infant memory development: An ecological model. *Developmental Psychology*, *45*(1), 160–174. https://doi.org/10.1037/a0014538

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences*, *109*(6), E360–E367. https://doi.org/10.1073/pnas.1108555108

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, *255*(5), 114B–125B. https://doi.org/10.1038/scientificamerican1186-114b

Tummeltshammer, K., & Amso, D. (2018). Top-down contextual knowledge guides visual attention in infancy. *Developmental Science*, *21*(4), e12599. https://doi.org/10.1111/desc.12599

Twomey, K. E., Ma, L., & Westermann, G. (2018). All the right noises: Background variability helps early word learning. *Cognitive Science*, *42*, 413–438. https://doi.org/10.1111/cogs.12539

Võ, M. L., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 23–41. https://doi.org/10.1037/a0024147

Võ, M. L., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences*, *1339*(1), 72–81. https://doi.org/10.1111/nyas.12667

Wasserman, E. A., Teng, Y., & Brooks, D. I. (2014). Scene-based contextual cueing in pigeons. *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*(4), 401–418. https://doi.org/10.1037/xan0000028

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-mont-old infants. *Developmental Psychology*, *34*(6), 1289–1309. https://doi.org/10.1037/0012-1649.34.6.1289

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, and Psychophysics*, *73*(6), 1650–1671. https://doi.org/10.3758/s13414-011-0153-3

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, *70*(1), 65–77. https://doi.org/10.1111/1467-8624.00006

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, *30*(4), 553–566. https://doi.org/10.1037/0012-1649.30.4.553

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118–136. https://doi.org/10.1016/j.jecp.2010.04.014

How to cite this article: Tummeltshammer, K., & Amso, D. (2023). Infants use contextual memory to attend and learn in naturalistic scenes. *Infancy*, *28*(3), 634–649. https://doi.org/10.1111/infa.12530