Cognition 229 (2022) 105256

Contents lists available at ScienceDirect

Cognition

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

Scene saliencies in egocentric vision and their creation by parents and infants

Erin M. Anderson *, Eric S. Seemiller, Linda B. Smith

Psychological and Brain Sciences, Indiana University, USA

ARTICLE INFO

Keywords:
Vision
Salience
Visual search
Egocentric vision
Parent-child interactions

ABSTRACT

Across the lifespan, humans are biased to look first at what is easy to see, with a handful of well-documented visual saliences shaping our attention (e.g., Itti & Koch, 2001). These attentional biases may emerge from the contexts in which moment-tomoment attention occurs, where perceivers and their social partners actively shape bottom-up saliences, moving their bodies and objects to make targets of interest more salient. The goal of the present study was to determine the bottom-up saliences present in infant egocentric images and to provide evidence on the role that infants and their mature social partners play in highlighting targets of interest via these saliences. We examined 968 unique scenes in which an object had purposefully been placed in the infant's egocentric view, drawn from videos created by one-year-old infants wearing a head camera during toy-play with a parent. To understand which saliences mattered in these scenes, we conducted a visual search task, asking participants ($n = 156$) to find objects in the egocentric images. To connect this to the behaviors of perceivers, we then characterized the saliences of objects placed by infants or parents compared to objects that were otherwise present in the scenes. Our results show that body-centric properties, such as increases in the centering and visual size of the object, as well as decreases in the number of competing objects immediately surrounding it, both predicted faster search time and distinguished placed and unplaced objects. The present results suggest that the bottom-up saliences that can be readily controlled by perceivers and their social partners may most strongly impact our attention. This finding has implications for the functional role of saliences in human vision, their origin, the social structure of perceptual environments, and how the relation between bottom-up and top-down control of attention in these environments may support infant learning.

When people *first* look at a scene, their attention is often drawn to areas with high contrast, greater luminance, visually large elements, and elements in the center of the scene (Clarke & Tatler, 2014; Itti, Koch, & Niebur, 1998; Najemnik & Geisler, 2005; Proulx & Egeth, 2008; Tatler, 2007; Wolfe & Horowitz, 2017). These saliences, which characterize human perception across the lifespan, capture attention by making some elements easier to see than others (Gerhardstein & Rovee-Collier, 2002; Itti & Koch, 2001). Saliences are sometimes discussed positively in terms of their evolutionary value in alerting us to danger (Itti & Koch, 2001; Posner, 1980) but are often framed negatively as distractors that need to be inhibited in goal-directed search (e.g., Corbetta & Shulman, 2002), or as Kaplan (1964) put it, only searching for lost keys under the streetlight. However, the strength and pervasiveness of bottom-up salience effects - even in tasks in which it would best to ignore them - suggests that they may be generally useful (Amso & Johnson, 2006; Bruce & Tsotsos, 2009; Feldman & Friston, 2010). Here, we provide new evidence for how saliences may play a functional role in everyday human attention. In brief, our hypothesis is that the keys *are* often under the streetlight because human behavior structures targets of attention so that they will be easier to see. We provide evidence for this hypothesis in analyses of infant field-of-view images captured during parent-infant play with objects.

We suspect that behaviorally-controlled saliences play a role in human attention throughout the life span. However, we focus on what attracts attention in infant field-of-view images for three reasons. First, human infants have weak top-down control of attention and have been shown to be highly susceptible to task-irrelevant saliences in laboratory studies (Frank, Amso, & Johnson, 2014; Tummeltshammer et al., 2014). Second, a large literature shows that parents often scaffold infant attention, helping them attend to an object of interest (Bornstein, 1985; Pêcheux et al., 1992; Suarez-Rivera, Smith, & Yu, 2019). While there is not yet evidence that parents exploit saliences by creating them around

* Corresponding author at: Department of Psychological & Brain Sciences, 1101 E. 10th Street, Bloomington, IN 47405, USA.
E-mail address: ema1@iu.edu (E.M. Anderson).

https://doi.org/10.1016/j.cognition.2022.105256
Received 16 February 2022; Received in revised form 9 August 2022; Accepted 11 August 2022
Available online 18 August 2022

targets of interest, parent-child interaction is one context in which this behavior seems likely to occur. Third, a growing literature based on infant field-of-view visual experiences shows that infants often move their bodies and objects in ways that increase the visual size of objects (e.g. Suanda et al., 2019; Yu & Smith, 2012), both in tandem with and independently of parent behavior. Visual size is a well-documented salience in the adult literature (Proulx & Green, 2011; Proulx & Egeth, 2008), raising the possibility that infant behavior enhances other attention-getting properties for targets of interest. For these reasons, parent-infant interaction during play seems likely to be a useful context for an initial study of how humans may intentionally or unintentionally create bottom-up saliences to shape their own and others' attention. Our method takes an egocentric vision approach (Mann, Kitani, Lee, Ryoo, & Fathi, 2014), studying visual saliences in the infant's field-of-view, as captured by a wearable camera in the context of free-flowing behavior by the parent and infant.

1. Salience in egocentric vision

The growing literature on egocentric vision in adults and infants has repeatedly raised the question of whether findings that emerge from highly controlled experimental studies operate in the same way – or play any role at all – in the context of free-flowing behavior (Foulsham, Walker, & Kingstone, 2011; Franchak, McGee, & Blanch, 2021; Hayhoe & Ballard, 2005; Yu & Smith, 2013). Accordingly, we first need to determine whether the suite of saliences that control human attention in experimenter-created scenes are detectable in infant egocentric images.

Salience is a psychological construct, defined in terms of experimentally measurable effects of different stimuli on human attention. In visual search tasks, perceivers are asked to find specified targets; these search times are used to measure and *define* what properties are salient (Verghese, 2001; Wolfe, Cave, & Franzel, 1989; Eckstein, 2011). We use this same measure of salience the visual properties in infant ego-centric images. The rationale for our approached is further explained with respect to Fig. 1 (A-C) which shows three kinds of scenes that might be used in a visual search experiment: simple search arrays (A), photographs of scenes (B), and infant egocentric images (C).

The majority of studies on human visual search used simple arrays with carefully manipulated and controlled visual properties (Fig. 1A). Findings from these paradigms have converged on a set of stimulus properties that matter under a variety of task conditions. Search time is faster when targets are visually larger than or otherwise contrast with the distractors around them (Proulx & Egeth, 2008; Wolfe & Horowitz, 2017 for a review). Increases in set size – the number of distractors or competing elements in an array – also increases the time needed for perceivers to visually locate the target (Wolfe, 2020). Additionally, there is a strong bias for perceivers to look first to items closer to the center of a scene or screen, generally considered a nuisance variable in need of control (Araujo, Kowler, & Pavel, 2001; Bindemann, 2010). Studies of infants and children indicate these same features in simple arrays are also salient early in development (e.g., Cavallina, Puccio, Capurso, Bremner, & Santangelo, 2018; Gerhardstein & Rovee-Collier, 2002).

Experiments using photographs of real-world scenes (Fig. 1B) show that many findings from simple visual search tasks extend to these more complex contexts (Tatler, Hayhoe, Land, & Ballard, 2011). For example, saliency "maps" locate areas of real-world scenes where hypothesized saliences are present and where multiple saliences converge (e.g., where elements contrast from their surround in terms of color, luminance and visual size). These studies have shown that saliences derived from simple arrays predict the location of attention and depth of visual processing within a real-world scene (Borji, Sihite, & Itti, 2013; Itti et al., 1998; Itti & Koch, 2001; Nuthmann, Clayden, & Fisher, 2021; Proulx & Egeth, 2008). As with simple arrays, studies of real-world scenes also show perceivers' strong bias to direct attention to the center of the scene (Hayes & Henderson, 2020; Tatler, 2007; van Renswoude, van den Berg, Raijmakers, & Visser, 2019). At the same time, some differences in the stimulus factors that influence spatial attention in simple search arrays and real-world scenes have been noted. For one, there is no principled way to define the number of distractors for photographs of real-world scenes and thus to determine set size; this difficulty limits researchers'

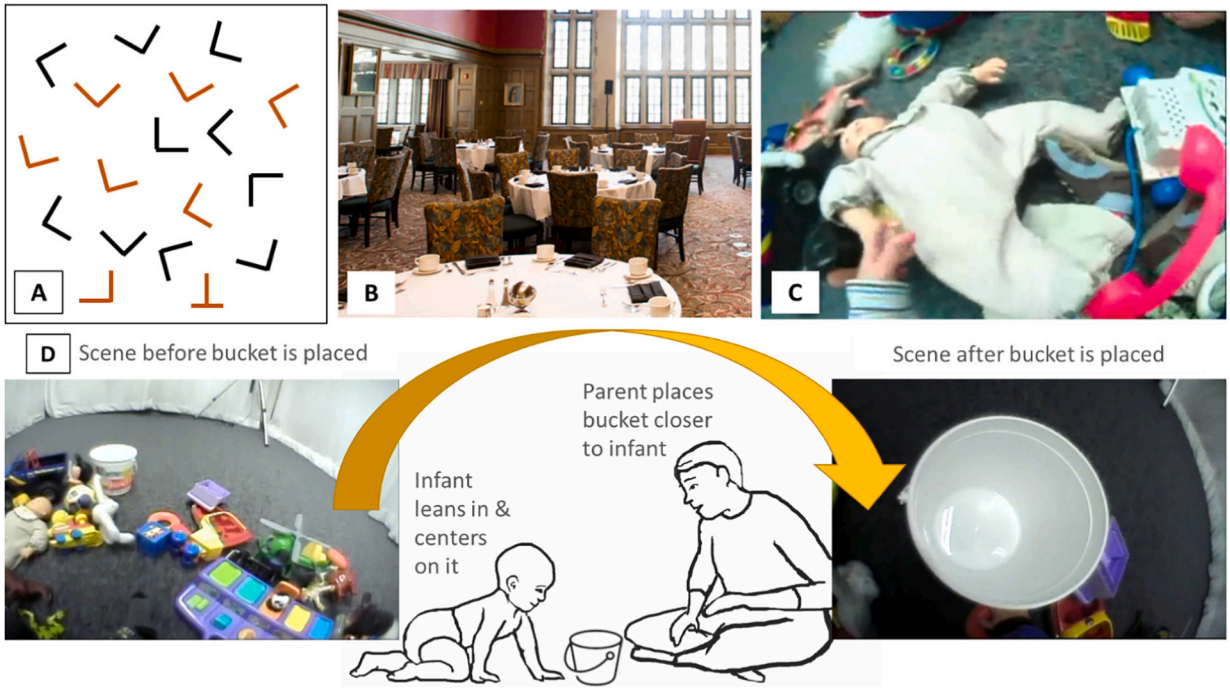

Fig. 1. *From top:* An example of types of images used in classic visual search tasks (A), real-world scene search tasks (see Sareen, Ehinger, & Wolfe, 2016) (B), and egocentric images captured via infant head cameras (C). The sort of everyday scene in C (whether an infant playroom or a cluttered drawer, cabinet) is different from typical search, because of the complexity of shapes, locations, sizes, and the convergence of multiple properties. (D) is an example of how visual properties could change through the actions of perceivers and social partners on the environment.

ability to make precise predictions about how long it should take to locate a target in such scenes (see Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). However, clutter (how many edges, or variability in properties; see Rosenholtz, Li, & Nakano, 2007) can markedly slow search times. For adults, semantic knowledge based on a lifetime of experiences also influences where viewers look for particular objects in scenes, and can even out-predict classical saliences like contrast (Henderson & Hayes, 2017; Võ, 2013).

Pictures taken by photographers are real-world scenes, but they also have limitations. These photographs are created by perceivers with mature visual systems who hold the camera and frame the picture, thus may selectively bias the saliences in the images in different ways than those that characterize the field-of-view experiences of freely-behaving perceivers (Fathi, Ren, & Rehg, 2011; Foulsham et al., 2011; Hayhoe, 2000; Smith, Yu, Yoshida, & Fausey, 2015; Tatler et al., 2011). Fig. 1C shows a head-camera image captured by an infant playing with toys. Such an image has *uncontrolled* complexities with many saliences, some competing with each other and some correlated with each other. For example, one could speculate from extant research that the baby doll in the center of the image is salient due to its size and position, or that the red phone is salient due to its contrast with the lighter background, or perhaps the yellow object on the edge of the image would attract attention due to its unique shape and color relative to the scene; indeed, all of these may be simultaneously true. It is quite possible that if one used egocentric images in standard visual search tasks, one would find *no* systematic saliences as a result of this variability and mutual interference. If this were the case, it would be a meaningful finding in its own right, because it would suggest that the saliences studied in the laboratory have only minimal effects in attracting everyday attention. Thus, our analyses of the saliences in infant head-camera images (Fig. 1C) use the context of a standard visual search task with adult participants as the behavioral measure of salience. We specified the target of search for adults (the baby doll, the phone, the cow), measured their visual properties, quantified a suite of visual properties in the image suggested by traditional visual search studies, and related these to speed of search. The use adult performances in experimental tasks as the behavioral metric on infant experiences is one often used in research on the properties of developmental environments (see Gilette, Gleitman, Gleitman, & Lederer, 1999's Human Simulation Paradigm).

2. Behavior-created saliences

By hypothesis, if there are systematic saliences in infant egocentric images, their functional relevance would derive from their active creation by infants and their mature social partners. Fig. 1D illustrates one example of how a parent and an infant might jointly change the properties of the egocentric scene to make a target object more salient. In the left-most panel of Fig. 1D, there is a clutter of potential target objects; in the right most panel of Fig. 1D, there is one visually dominant object in the scene: white against the background blue rug, with minimal surrounding clutter. In our corpus of infant egocentric images of object play, this rightmost image was created by the *joint* behaviors of the parent placing the object in the child's field of view and the infant leaning in to take a close look. The literature on infant egocentric experiences suggests that these kinds of events are common and that they alter the structure of the scenes in ways helpful to infant attention (e.g., Bambach, Crandall, Smith, & Yu, 2018; Smith, Yu, & Pereira, 2011; Yu & Smith, 2012). Perceivers and their social partners can move objects toward or away from the perceiver and into uncluttered or cluttered regions, thereby structuring the scene saliences (Burling & Yoshida, 2019; Yoshida & Smith, 2008). These behaviors do not have to be intentional to be effective. We will return to this point in the General Discussion.

Our study provides an initial quantification of the visual properties and their saliences (as measured by adult visual search) as they are behaviorally created by the infant and their social partners in the context of active perception. We focus on scenes generated by typically developing 12-month-olds who are likely to benefit from parent scaffolding, and who, at the same time, have sufficient motor control to structure their own environment through purposeful movement, either in tandem with parent actions or on their own. To measure the effects of both parents' and infants' actions, we analyzed the properties of acted-upon objects within egocentric images and their effects on attention. If people systematically enhance visual targets for themselves and others, then the power of bottom-up saliences on human attention would not rest merely on their evolutionary history but would also reflect day-to-day experiences in infancy and perhaps across the lifetime.

3. Methods

3.1. Collection of the egocentric images

Infants ($n = 17$, 9 male), who were between 11.5 and 12.5 months of age, wore head cameras during 10 min of unstructured toy play with their parent. Play occurred on a carpeted floor in a 3-m by 4-m room with an assortment of 32 toys that were haphazardly placed on the floor. The toys sampled many common toy categories (e.g., animals, vehicles, characters like Mickey, a doll, ball, bucket, block, stacking ring, and baby-toy versions of objects like a rotary phone and keyboard; the complete list and photos are available at https://osf.io/4cvw5). Toys ranged in volume from 256 cm^3 to 12,472 cm^3 ($M = 2445$ cm^3). Although all parent-infant dyads had the same 32 categories of toys available, some versions of the toys were different between participants, i.e., different horses or different buckets. In total, 40 unique objects were used across dyads to instantiate the 32 categories.

The head camera (Watec 90^o diagonal field of view, recording at 30 Hz) was worn by the infant on a headband or hat with the camera lens set low over the eyes of the infant (Borjon et al., 2018). After the head camera was placed on the infant, the parent was told to play with their infants as they would normally and then parent and infant were left alone in the playroom for the 10-min play period.

3.2. Selection of images

The selection of images from the 170 min of head camera video (down-sampled at 5 Hz) was aimed at capturing the potential role of infants and their parents in structuring the scenes. The criteria for placement images were that the object had to have been moved by the parent or infant and released in the infant's head-camera field-of-view immediately before the selected image occurred. Because parents and infants act on objects manually (Yu & Smith, 2012), either parents' or infants' hands or both are present in nearly every image. An additional criterion for selecting an image was that the mover's hand was no longer in contact with the moved object.

The selected scenes were located by trained in-lab coders, who looked through images from each subject in the larger corpus in chronological order (see Supplemental Materials for video examples of parent and infant placements). This resulted in selection of 968 unique images (720 × 480 pixels) that were used to create two categories of search trials: ones in which the search target was the object that had placed by a parent or infant, and ones in which the target was visible in the scene but had not been placed. An individual image could be used for both categories with search directed to different objects, as described below. Because parents were very active in placing objects in the field of view of their infants, more than half of the selected images ($n = 529$) featured objects placed by parents. There were 291 images in which infants placed the objects and 147 images in which no in-view object had been immediately placed by a participant. The objects placed by participants were not a uniform sampling of the 40 available toys in the room. At the corpus level for parents, the 40 unique objects were placed in the scene from 0 to 59 times (*median* = 15); for infants, the 40 unique objects were placed between 0 and 26 times (*median* = 6). To create an equal number of *placed* and *not-placed* matched targets for the search

task (described below), we selected *not-placed* targets from images in which a different object had been placed.

3.3. Visual search task

We used performance in a visual search task to behaviorally measure the salience of targets in images where targets had or had not been placed. On each trial, adult participants were asked to find a single target within an infant egocentric scene. We used a version of the visual search task common in the literature (Treisman, 1977; Wolfe et al., 1989), presenting an image of the search target prior to every trial, and then presenting the scene, asking participants to respond as soon as they found the object. While the instruction image of the target was a single canonical-view photograph of the object, the egocentric scenes were uncontrolled real-world images (as shown in Fig. 2). As with searching for objects in the real world, the target in the scene was unlikely to present the same canonical view as in the instruction image, and may instead be partially occluded, embedded in a jumble of other objects, or have different shadows or variable lighting.

Due to COVID-19 protocols at the time of data collection, the search task was conducted online. This this online task could not control for the participant's screen size or the distance between participant and the screen. (We may reasonably estimate that screen size diagonals varied from 482.6 mm to 863.6 mm (Anwyl-Irvine, Dalmaijer, Hodges, & Evershed, 2021) and that distances ranged from 500 mm to 100 mm ("Computer Workstations eTool", 2022), with a greater distances typical for larger screens. Assuming a 600 mm distance and a 544 × 306 mm screen, we had a diagonal visual angle of 54.96° in our task). With this greater variability - typically controlled for in lab experiments - our online measure adds noise that may reduce the sensitivity the behavioral measure of saliences. Therefore, saliences captured by this method are ones that remain potent across viewing conditions, and so are also likely relevant to the highly variable viewing conditions of real life.

3.3.1. Participants

Participants (n = 156) were recruited through Mechanical Turk (MTurk). Criteria for participant inclusion was that they were over 18 years of age, resided in the United States, and had at least a 95% approval rating in MTurk, indicating successful completion of 95% of previous tasks they had enrolled in on the site. Participants were compensated $3.75 for a 15-min task.

3.3.2. Search trials

Each trial consisted of a specified search target and one of the 968 selected images. These target-image pairings were created as follows. For each image with a participant placed object, that object served as a search target for that image. For each of these **placed target trials,** we created a target-matched **not-placed target trial** by selecting images in which that matching target was present but not placed. These matched target trials of not-placed images could be formed from images in which a different object had been placed or from images with no placed objects. From these options, the target matched trials were selected randomly with the only criterion being that the target was in the image and not in contact with any hand.

We used this procedure to create two between-subject search conditions, one using only parent-placed objects as targets and randomly selected matching not-placed targets, and the other using only infant-placed objects and randomly selected matching not-placed targets. Sets of 160 trials (80 placed target and 80 not-placed target) trials were created for presentation to the participants. Within these sets, each unique trial, defined by the combination of target and image, was presented to 12 participants. The trials for each participant were created by the random sampling (without replacement) of possible target-image pairs for their respective condition with the constraint of a maximum 12 samplings across participants for each unique target-image trial. Given these procedures, individual participants saw each unique scene (while searching for *different* targets) an average of 1.2 times (*median* = 1) and searched for the same target category (but in different unique scenes) an average of 4.5 times during the task (*median* = 3). There were 22,080 total number data points across subjects and trials.

The search task was created in javascript using the jsPsych library (de Leeuw, 2015). All search trials were constructed from the selected 968 images (480 × 720 pixels; 12.93° x 19.3°). Instruction images (560 × 560 pixels, 15.06° x 15.06°) for each of the 40 unique toys presented the search target for the trial in a canonical view on a white background, as shown in Fig. 2.

3.3.3. Procedure

Each participant began the task with 8 practice trials, followed by 160 search trials. Randomly interspersed with the search trials were 5 trials in which the search target was not present. Given the purposely uncontrolled scene properties and the wide variability in object viewpoints across scenes, these target-absent trials were not "catch" trials in the usual sense but were included to help with instructions and to let participants know that "timing out" and not finding some targets was expected. Participants were told the following: "We want to know how quickly people can find toys in cluttered playtime scenes. At the beginning of each trial, you will see a picture of a toy. The picture shows the toy you will be looking for in the trial. Then a scene will appear, like this (example scene). The toy in the scene might be from a different angle or have different lighting than in the first picture you were shown. As soon as you see the toy you are looking for in the scene, press the "F" key. Please respond as quickly as possible." Participants were told that if they could not find the target, to just let the trial timeout and that there

Fig. 2. Examples of targets and infant head-camera images used in the visual search task when the targets were placed by parents, placed by infants, were not placed, or absent.

were some trials in which the target was not present.

Participants were shown an example target and example scene as part of the instructions. They were told that although targets might repeat across trials, they should only search for the current target. After the instructions, participants were given eight practice trials with feedback. On one of these practice trials, the target was absent. In the feedback, participants were reminded to press the "F′ key when a target was present and to refrain from pressing any key and to let the scene time out when it was absent. After the practice trials, each participant was given the test trials without feedback. Then the target image appeared, along with another prompt to press the spacebar when they were ready to search. After pressing the spacebar, the scene appeared. The scene presentation was timed and disappeared after 5000 ms. Other elements of the study were self-paced so that online participants would be required to provide input between trials. Once the searches were completed, the participant was given a unique code to submit their HIT on Mechanical Turk.

3.3.4. Response time

Response time was the latency between the appearance of the scene and participant pressing the F key to indicate finding their target. If participants did not respond for the 5 s duration of the search trial, their response time was set to 5000 ms (the maximum presentation time). Participants did not find the present targets within 5000 ms on 9.4% of the experimental trials and responded within the 5000 ms limit an average on 2.5 of the 5 absent trials.

3.4. Visual measures

For each target-image trial used in the search task, the target was bounded with a 4-point polygon using MATLAB's image labeler. In-lab coders were instructed to draw polygons that captured as much of the visible object as possible, while staying close to the boundaries of the object. A third of the targets (32.6%) were bounded by two independent coders, and their average agreement was 0.84 using an intersection-over-union approach. For targets with <0.5 agreement ($n = 3$), an independent third coder resolved the disagreements. **Visual size** was calculated as the angular area of the polygon (°). **Distance from center** was determined by subtracting the center point of the polygon (x,y) from the center point of the whole scene (360, 240) on the x- and y-axis and finding the hypotenuse in Cartesian coordinates.

The remaining visual measures were determined with respect to a 400 × 400 pixel region of interest (ROI) whose center coincided with the center of the target (or as close to the target's center as possible, if the target was very near the image edge). **Set size** was defined as the number of toys in the ROI. In-lab coders were given the cropped (400 by 400 pixel) ROI images and instructed to place a point on each toy that was visible in MATLAB's image labeler. Coders could see the points they had previously placed. The number of points placed was then exported for each ROI image. Because the basic task was placing dots on a maximum of 40 possible objects that the coders were experts in recognizing and because the dots were then counted by an algorithm, accuracy was high. **Quantity of edges** in the ROI was determined using a canny edge detection algorithm in MATLAB. This is a clutter measure commonly used in computer vision; we set the threshold for edges to 0.45 and discarded those below 0.25. This threshold eliminated edges that were based on the textured carpet. The outputs were averaged across the entire ROI and then we divided these values by the maximum value for the full set of images. These resulting values indicate whether the **amount of edges** was high or low relative to our stimulus set. Target-surround differences in edges were by calculating the **proportion of edges** belonging to the target object compared to the rest of the ROI. This last measure indicates how much of the visual information belongs to the target object and not the surround. Because there is no consensus as to how to measure clutter or amount of stuff in real-world scenes (Rosenholtz et al., 2007; Wolfe et al., 2011) we also computed parallel measures to the reported quantity and target-surround edge measures based on gabor filters and observed the same pattern of results. We measured the luminance **contrast** of the object within the immediate surround of the ROI (Sebastian, Seemiller, & Geisler, 2020). The ROI was converted to LAB color space to mirror human vision more closely. For the target object (the pixels in the polygon) and its surround (full ROI), we found the root mean squared (RMS) contrast of the intensity values in the L channel, and then determined what proportion of this contrast was accounted for by the object, compared to the surround. We also analyzed the a & b color channels for red-green RMS contrast and blue-yellow RMS contrast, i.e., not whether an object is red or blue but how much it differed from its surround in terms of color. Color contrast results were similar to the luminance contrast analyses and thus we report only the latter in this paper. The code for all visual property measures and the corresponding data is available at https://osf.io/4cvw5.

4. Results & discussion

4.1. Parent and infant placement of objects

We first examined the properties surrounding target objects, depending on whether they had been placed or not. An initial question was whether there were selection effects based on physical sizes of the objects, which varied considerably. As noted in the Methods section, parents and infants did not uniformly place the available objects, and a selection effect could bias the visual properties. For example, if parents or infants only placed physically larger objects, this tendency might lead to larger (and thus attention-grabbing) visual sizes, which could also influence measures of the proportion of visible edges belonging to the object, or how many other objects were in view (set size). We found that infants were in fact biased to place smaller, easier to hold objects. The mean physical size of objects placed by infants ($M = 2019.17$ cm^3, $SD = 2213.17$) was smaller than of the mean of the available objects in the room ($M = 2464.38$ cm^3, $SD = 2805.01$). Related to this, the infant-selected distribution of object sizes significantly differed from a random sampling of available objects, $t(575) = 2.68$, $p < .001$ (see Fig. 3A). Parent-placed objects were physically larger than infant-placed objects ($M = 2626.85$ cm^3, $SD = 2753.06$) and were an unbiased sample that did not differ from a random distribution, $t(1102) = -1.14$, $p = .25$. Additionally, the physical size of an object did not strongly predict its visual properties within the infant egocentric images: there was a weak correlation in general between physical size and visual size ($r = 0.30$), as well as between physical size and the proportion of edges belonging to the object in an image ($r = 0.29$). There was no correlation between an object's physical size and its centeredness in egocentric images ($r = 0.03$), between physical size and surrounding set size ($r = 0.03$), between physical size and the amount of edges in the ROI ($r = -0.03$), or between physical size and contrast ($r = 0.03$).

Turning to the visual properties surrounding target objects, we asked how these properties changed when they were placed by parents or infants. To better control for the physical properties of individual objects, we randomly paired each placed target with an instance of the same target when it was present but not placed (e.g., if the phone was the placed object in a scene, it was paired with a companion scene where the phone was in view but was not the placed object). This allowed us to compare Target Type - the properties of placed versus not placed targets - and as well as differences between who did the placing. We begin by collapsing across who did the placing, because the infant wearing the head camera is always playing a role in in the scene properties. Parent-placed does not mean the parent behavior on its own determined the visual properties of the target object: because the images are from the infant point of view, the saliences in the parent-placed objects are influenced by both parent placement and any body movements by the infant that affects their view. We used Linear Mixed Effect models for our analyses. Models were tested in the R environment (R Core Team,

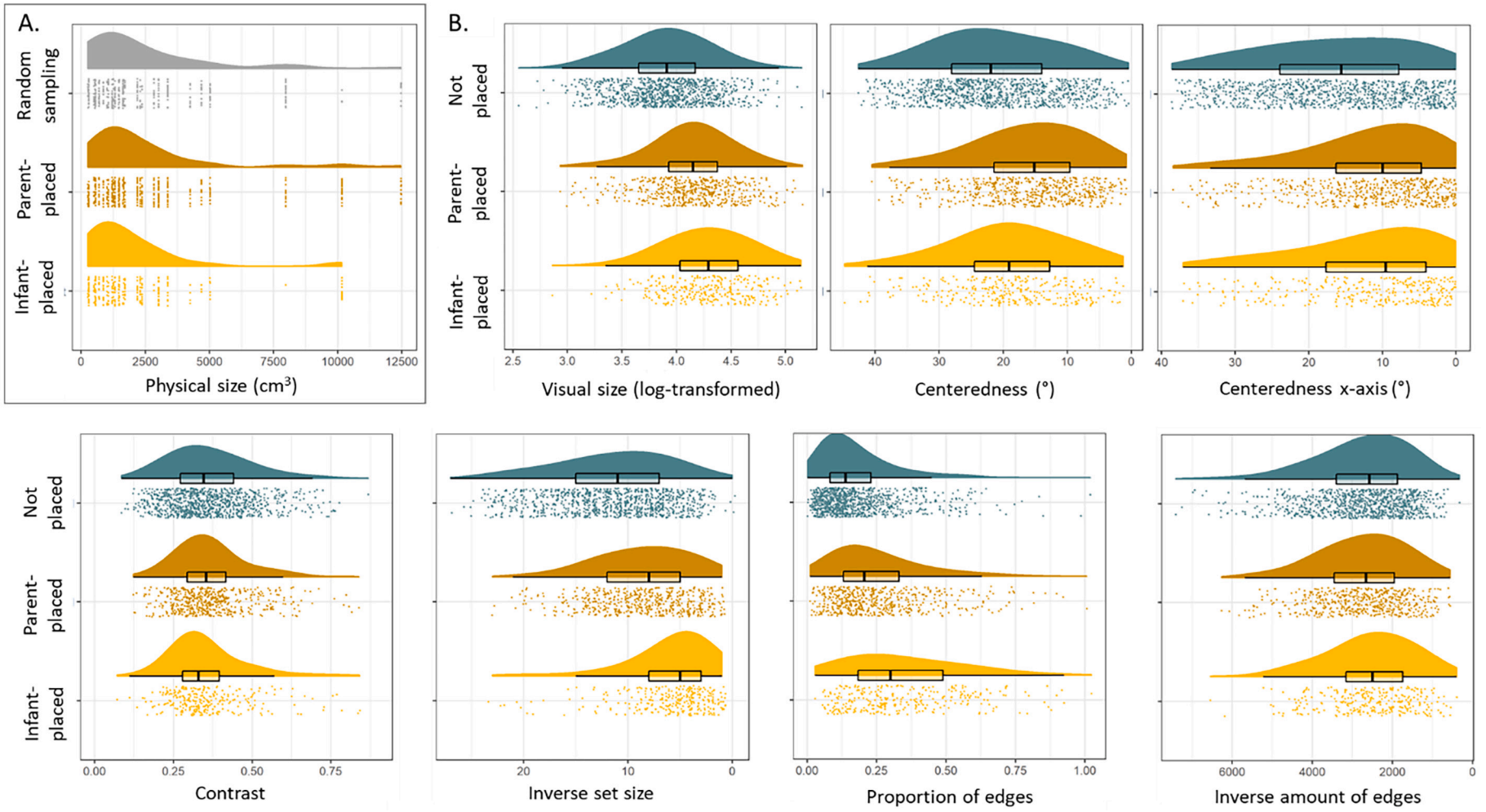

Fig. 3. Raincloud plots showing A) the physical size of objects placed by parents or infants, compared to a random sampling of available objects and B) relationships between visual properties of the object in the image and whether the target was placed by parents, by infants or was not placed. In each plot, the solid-colored area shows the density distributions, the individual dots are the raw data points, and the box plot indicates the median of the data (the thick bar in the middle) and the 25th and 75th percentiles (the lower and upper hinges).

2021) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), with significance *p*-values obtained using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) which uses the Satterthwaithe approximation (Luke, 2017).

The model specification was *Property ~ Type*Who + (Type | object) + (Who | dyad)*, where the fixed effect term of type was whether the target was placed or not-placed, the fixed effect term of who was whether the target was moved by the parent or the infant, and there was an interaction term between these. Random effects of placement were entered for individual toys that served as targets and random effect of who were entered for each parent-infant dyad. We found that visual size and centeredness, the two body-centric measures, corresponded with placement. Placed targets were visually larger than when they were otherwise in the scene (M_p = 2485°, SD = 2340°; M_u = 1327°, SD = 1432° area), and were closer to the center of the scene (M_p = 17.2°, SD = 8.79; M_u = 21.3, SD = 9.28° from center). See Table 1 for details.

Clutter measures were also associated with placement of the object in the scene. Placed targets were in ROIs with fewer distractor objects (M_p = 7.69, SD = 4.54; M_u = 11.4, SD = 5.37 toys), and a higher proportion of edges belonged to the target object compared to its surround (M_p = 0.29, SD = 0.18; M_u = 0.19, SD = 0.14). Contrast of the target within its surround did not differentiate target types: it was similar whether the target was placed (M = 0.36, SD = 0.12) or not-placed (M = 0.37, SD = 0.14). The amount of edges in the ROI was also similar between the two types of target objects (M_p = 2678, SD = 1087; M_u = 2756, SD = 1164 edges).

For most of the visual properties, there was also an interaction between Target Type (placed or not) and who it was placed by, indicating some differences between parent-placed targets and infant-placed targets. Objects were larger in view (M = 3025.47°, SD = 2702.10°) when they were infant-placed than when they were parent-placed (M = 2193.88°, SD = 2063.56°), but parent-placed targets were closer to scene center (M = 16.1° away, SD = 8.53°) than infant-placed targets (M = 19.1°, SD = 8.93°). The measures of centeredness on the x- and y-axis revealed that infant- and parent-placed objects were similarly centered on the x-axis (M_i = 12.10°, SD = 9.42° from center; M_p = 11.70°, SD = 8.71°), with no interaction between target type and who placed it. However, infant-placed objects were below center on the y-axis and thus closer to the infant themselves (M_i = 12.80°, SD = 6.82; M_p = 9.02° from center, SD = 6.31°), leading to a Type by Who interaction for centeredness on the y-axis (see Table 3).

For the clutter measures, infant-placed targets had ROIs with smaller set sizes (M = 5.96, SD = 4.07) compared to parent-placed targets (M = 8.62, SD = 4.51), and there were fewer edges for infant-placed targets (M = 2545.06, SD = 1078.89) than for parent-placed targets (M = 2750.08, SD = 1086.05). Similarly, a greater proportion of edges in view belonged to targets that were infant-placed (M = 0.35, SD = 0.20) than those that were parent-placed (M = 0.25, SD = 0.17). Finally, contrast in the ROI was similar across placements by infants (M = 0.35, SD = 0.12) and parents (M = 0.37, SD = 0.12).

Having a sense of how the visual properties differed between parent-placed, infant-placed and not-placed objects, we next asked how well these categories predicted the speed with which outside observers could locate these targets within the egocentric scene. To do this, we analyzed the visual search times for the placed and not-placed targets using the following model specification: *RT ~ Type * Condition + (Type | object) + (Type | respondent)*, where Type (placed or not-placed) and Condition (whether the respondent was in the parent- or infant-placed condition) were entered as fixed effects, and the target and the visual search respondent (MTurk subject) were entered as random effects. Collapsing across parents and infants, placed objects were found on average at 881 ms (SD = 815) while not-placed targets were found on average at 1048 ms (SD = 960). There was a main effect of whether the target was placed or not, with an estimate that the shift from a placed target to a not-placed target would increase search time by β = 107.79 ms (SE = 40.82 ms), $t(232)$ = 6.64, p = .009. There was no main effect of condition, β = −27.80 ms (SE = 94.19 ms), $t(154)$ = −0.30, p < .77. and collapsing across placed and not-placed, targets were found at similar speeds (M_{parent} = 955.92 ms, SD = 900.34; M_{infant} = 996.27 ms, SD = 848.21). There was however, an interaction between whether the target was placed and whether it was in the infant-placed or parent-placed condition, β = 90.39 ms (SE = 42.67 ms), $t(208)$ = 2.12, p < .035. Targets placed by parents were found faster (M = 868 ms, SD = 820) than targets placed by infants (M = 954 ms, SD = 787), while not-placed targets were found within similar timeframes, whether they appeared in the parent-placed condition (M = 1049.37, SD = 969.73) or in the infant-placed condition (M = 1040.76 ms, SD = 905.90); see Fig. 4B.

To summarize, the visual properties of placed objects differed from those of not-placed objects on most of the traditional visual measures that predict salience in search tasks. Placed objects were larger in view, more centered and appeared in regions of interest with reduced clutter. Moreover, while the overall centering of the target in the scenes (primarily due to centering on the y-axis) was more strongly associated with parent-placed than infant-placed targets, most of the visual properties were similar for parent- and infant-placed objects, and distinguished from not-placed objects. Not all visual properties distinguished objects that had been acted upon from those otherwise in the scene: neither the amount of edges in the ROI, nor the amount of contrast were significantly altered by parent and infant behavior. Finally, parent-and infant-placed objects were found more rapidly than not-placed objects in the adult search task.

4.2. Properties that support rapid search in egocentric images

These analyses of visual properties that commonly predict salience in search tasks do not mean that they are salient in the context of egocentric image. Visual salience is a psychological property, determined by the effect of some visual property on visual attention. The next set of

Table 1
Coefficient estimates from a mixed-effects models of visual properties by parent and infant placement. The model specification was Property ~ Type*Who + (Type | object) + *(Who | dyad)*.

	Type (placed to not-placed)		*Who (infant to parent)*		*Type x Who interaction*	
Variable	B (*SE*)	t-value	B (*SE*)	t-value	B (*SE*)	t-value
Visual size (°)	−1451.92 (*228.67*)	−6.35 **	−1105.93 (*116.67*)	−9.48**	706.55 (*159.26*)	4.44**
Centeredness (° overall) †	1.74 (*0.74*)	2.36*	−3.18 (*0.65*)	−4.86**	3.74 (*0.90*)	4.15**
Centeredness (° x-axis) †	4.97 (*0.78*)	6.36**	−0.78 (*0.69*)	−1.12	−0.35 (*0.95*)	−0.37
Centeredness (° y-axis) †	−3.20 (*0.56*)	5.68**	−3.71 (*0.49*)	−7.49**	5.76 (*0.68*)	8.42**
Set size	3.80 (*0.50*)	7.53**	2.99 (*0.31*)	9.55**	−1.04 (*0.44*)	−2.39*
Edge proportion	−0.14 (*0.02*)	−7.48 **	−0.12 (*0.01*)	−11.59**	0.07 (*0.01*)	4.96**
Amount of edges	135.09 (101.29)	1.33	346.56 (*76.08*)	4.55**	−189.37 (*104.72*)	−1.81
Contrast †	0.02 (*0.02*)	1.20	0.02 (*0.02*)	1.13	−0.04 (*0.02*)	−1.97*

Notes: * indicates p-values <0.05. ** indicates p-values <0.001. † For this variable, the full model did not converge, so the model specification is Property ~ Type*Who + (1|object) + (1|dyad).

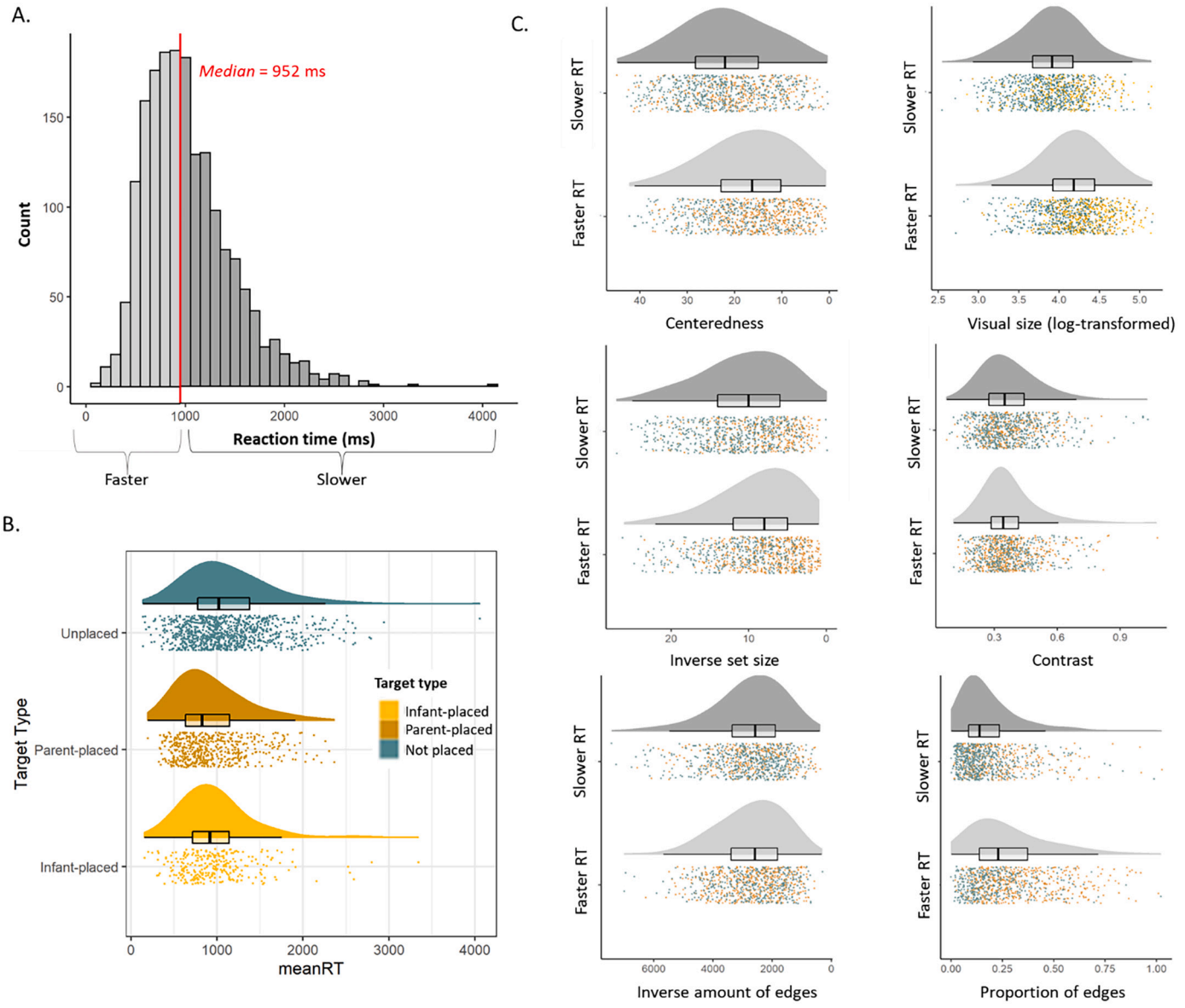

Fig. 4. *A.* The distribution of reaction times, averaged so that each data point reflects a unique scene & target. Red line indicates the median of the distribution. *B.* The relationship between reaction time and whether target was placed by the parent, infant or not placed. *C.* Raincloud plots showing relationships between reaction time bin (faster vs. slower) and each visual property. See Fig. 3 caption for explanations of plot elements. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

analyses asking how the measured visual properties of the placed and not-placed objects impacted adult visual search times. Fig. 4A shows the distribution of response times on all visual search trials. These were skewed with most responses falling under a second (*median* = 940 ms) but with a long tail of slower responses. As a first step to understanding the visual properties of these egocentric scenes, we partitioned the search trials into faster and slower categories using a median split; that is, "fast" searches had response times of less than a second, and "slow" searches were all longer. Then, for each of the measured properties, we examined how the trials with faster response times differed from those with slower response times. We used Linear Mixed Effect models to determine, for each measured property, whether that property differed between the so-defined faster and slower response categories. We used this approach rather than a linear regression because linearity and normal distributions are key assumptions for regression, and it was unclear that search time or the visual properties of the scene met either assumption (Osborne & Waters, 2002). Model comparisons confirmed this non-linearity: the mixed effects approach showed significantly better fit than the linear regression model across all indicators (e.g., AIC) and variables, except for amount of edges (see Supplemental Materials for model comparison and regression analysis results). The mixed effects model specification for each measure was *Property* ~ *RT bin* + *(RT bin | object)*, where the fixed effect term was the reaction time bin (from fast to slow) and a random effect was entered for individual toys that served as the target. For all analyses, the details of the statistical analyses are provided in Table 2.

Both visual size and centeredness of the target contributed to rapid search as shown in Fig. 4A. Targets in the faster-found category were typically visually larger (M_f = 2382.47°, SD = 2231.99°) and closer to center (M_f = 17°, SD = 8.77° from center) than slower-found targets (M_s = 13,229°, SD = 1569.67° angular area; M_s = 21.7°, SD = 9.2° from center).

Search was faster when there were fewer competitors -that is, when there were fewer edges in the region around the target, and a greater

Table 2
Coefficient estimates from a mixed-effects models of visual properties for fast and slow reaction time bins. The model specification was Property ~ RT bin + (RT bin | object).

Variable	B	SE	t-value
Visual size (°) †	−842.20	83.26	−10.11 **
Distance from center (overall) †	42.25	3.84	11.00 **
Distance from center (x-axis) †	33.21	4.11	8.09 **
Distance from center (y-axis) †	19.27	2.93	6.57 **
Set Size †	1.76	0.25	7.11 **
Edge proportion	−0.07	0.009	−7.51 **
Amount of edges	62.69	59.38	1.06
Contrast	<0.01	<0.01	0.11

Notes: * indicates $p = .047$, ** indicate <0.001. † indicates model of Property ~ RT bin + (1|object) because the model with random slope did not converge or the model fit was singular.

proportion of these edges belonged to the target. Specifically, ROIs surrounding faster-found targets were populated by fewer toys (M_f = 8.80, SD = 5.13 toys) than ROIs associated with slower-found targets (M_s = 10.55, SD = 5.39 toys) and fewer edges (M_f = 2688.85, SD = 1091.96; M_s = 2750.73, SD = 1164.80). Finally, for faster-found targets, a greater proportion of the edges in the ROI belonged to the target than to the surround, which was less likely to be the case for slower-found targets (M_f = 0.27 edges part of the target, SD = 0.18; M_s = 0.24, SD = 0.15). The two categories of search times, faster and slower, did not differ in the amount of contrast (M_f = 0.36, SD = 0.37; M_s = 0.12, SD = 0.14).

Like vision in the real world, the search task was one in which the view of the object and its scene context varied markedly, making visual recognition of an individual object potentially quite difficult. In this visual context, targets were found more rapidly if they were visually large and centered, two properties that can be directly influenced by behaviors that affect the spatial relation of the head camera (and thus the infant's head) to the designated target. These two properties are known to be associated with infant visual interest in an object as infants lean in or hold objects close and center objects at the midline of the body and in the head-centered field of view when attending to them (Bambach, Smith, Crandall, & Yu, 2016; Borjon et al., 2018). Targets were also easier to find in this context if there were fewer competitors for visual attention near the target, as measured by the number of toy objects (set size), amount of edges, and proportion of edges belonging to the target as opposed to the surround. These properties are potentially influenced by selective placement of objects relative to the surrounding area. Although there was considerable between-image variability in the contrast of the targets, high contrast did not support more rapid search within our images. The finding that contrast did not predict search time could be due to a specific aspect of the context which was one set of toys in a playroom. However, there was considerable variability in the contrast in the images and this variability was not found to be associated with search time, a fact that suggests that contrast may not be a potent salience in active vision in a complex visual world. We will return to these issues in the general discussion.

To summarize: compared to the simple arrays used in laboratory tasks, our search task using infant egocentric images is much more difficult, with diverse objects, lighting, and perspectives. Nonetheless, many of the same salience properties indicated in classic experiments–visual size, centering, and surrounding clutter – determined speed of search.

4.3. Converging properties that support rapid search

In complex scenes, distinct visual properties may correlate and co-occur, lead to joint effects on attention. First, we looked at how strongly measures were correlated, using Spearman's rank-order correlation (see Table 3). Between all measures, only four showed correlations above 0.50. There was a strong correlation between the overall measure of centeredness of the target and the x-axis centeredness sub-measure, $r(1806) = 0.83$, $p < .001$, and a medium relationship between the overall centeredness measure and the y-axis centeredness, $r(1806) = 0.52$, $p < .001$. This stronger relationship between x-axis centering and overall centering is consistent with previous findings that infants attend to and handle of objects of interest at the vertical midline of the body (Bambach et al., 2016). (The two sub-measures of centering in the image – y-axis centering and x-axis centering, showed a weak but reliable correlation, $r(1806) = 0.07$, $p < .01$, see Fig. 5A). Additionally, the proportion of edges clutter measure as strongly related to the visual size of the target = $r(1806) = 0.87$, $p < .001$, and had a negative medium relationship with set size, $r(1806) = -0.53$, $p < .001$.

Next, we examined the relationship between the convergence of multiple saliences and the effect of this on search time. Targets that are characterized by more salience properties should, in general, be found faster (e.g. Buetti, Xu, & Lleras, 2019; Itti & Koch, 2001). Segmenting targets by their reaction time bin, we see that, of the 968 targets in the faster-found group, 64% were above the median on (log-transformed) visual size, 61% were above the median on centeredness, 62% were above median on proportion of edges, and 54% were above the median on the inverse of set size (e.g., having fewer distractors). Only 50% of targets in the faster-found group were above the median on amount of edges, and only 49% on contrast. Focusing on the top four properties of visual size, centeredness and proportion of edges, we looked at their convergence across all targets in our corpus, regardless of search time: 18% of targets were above the median on all four saliences, 26% of the targets were above the median for three saliences, while 16% had two above-median saliences, 24% had a single above-median salience, and 17% were below the median on all saliences. Between faster- and slower-found targets, 26.7% of faster-found targets had four saliences, 29% had three saliences, 14.8% had two saliences, 18.8% had one salience, and 10.6% had none. In the slower-found group, only 7.2% had three saliences, 19.6% had 3 saliences, 16.4% had 2 saliences, 32.2% had only one salience and 34.6% had none. See Fig. 5B for the linear relationship between search time and number of above-median saliences.

We then determined when each target type (parent-placed, infant-placed or not-placed) had multiple saliences (Table 4). More than half of parent-placed objects and 70% of infant-placed objects were characterized by having 3 or 4 above-median saliences. Although these correspondences create a confound in the precise determination of the

Table 3
Spearman correlations between saliences (df = 1806).

	Dist from center x	Dist from center y	Dist from center	Visual size	Edge prop-ortion	Set size	Edge amount
Dist from center y	0.07**						
Dist from center	0.83***	0.52***					
Visual size	−0.19***	−0.18***	−0.28***				
Edge proportion	−0.07**	−0.04	−0.11***	0.93***			
Set size	−0.01	−0.13***	−0.05	−0.42***	−0.53***		
Edge amount	−0.26***	0.31***	0.39***	−0.05*	−0.11***	0.43***	
Contrast	0.00	−0.01	0.00	0.06**	0.14***	−0.03	0.09***

*Note. Asterisks indicates significance values based on two-tailed probability * $p < .05$, ** $p < .01$, *** $p < .001$.*

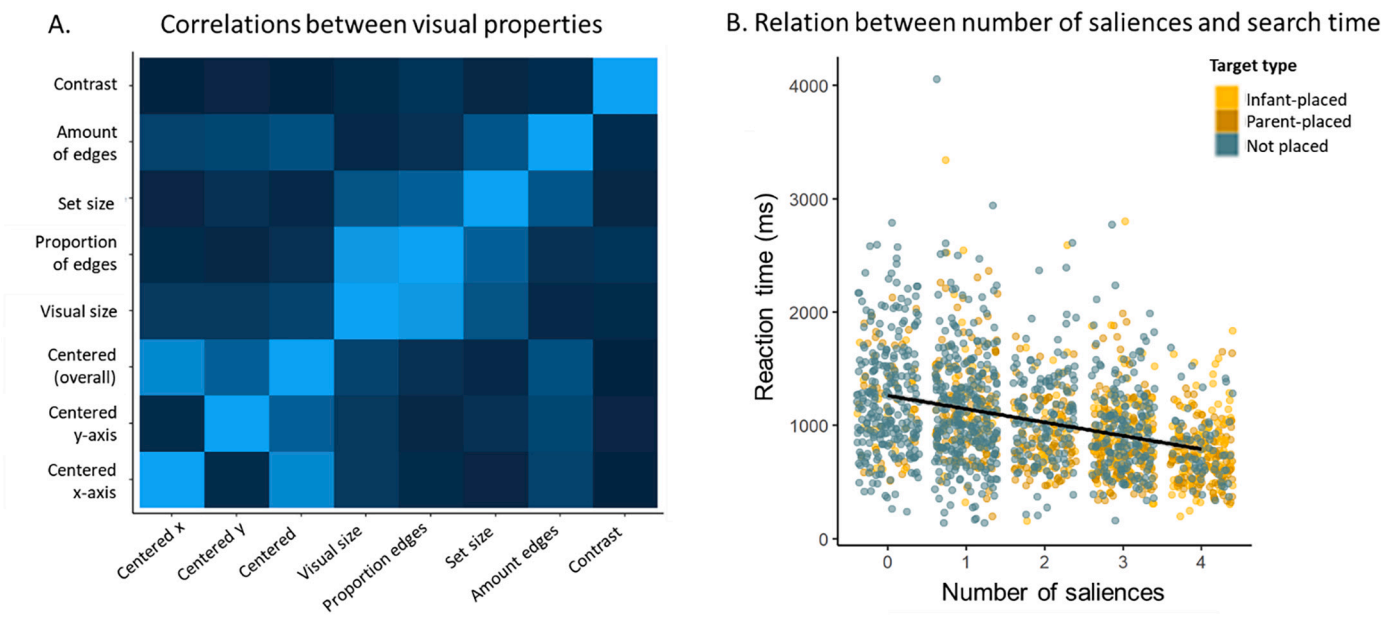

Fig. 5. *A.* Matrix of the strength of Spearman rank-order correlations between visual property measures, with lighter blues indicating stronger correlations (positive or negative) and darker blues showing weaker ones. *B.* Plot showing the relationship between the number of significant visual properties that each unique target is above the median on (visual size, overall centeredness, proportion of edges and set size) and mean search time for that target in the visual search task. Dot color indicates whether the target was parent-placed, infant-placed or not-placed.

Table 4

Number of saliences	Infant-placed	%	Parent-placed	%	Not-placed	%
0	18	6%	43	8%	257	27%
1	37	13%	114	21%	310	32%
2	32	11%	104	19%	146	15%
3	104	35%	153	28%	182	19%
4	105	35%	135	24%	66	7%
Total	*296*	*100%*	*549*	*100%*	*961*	*100%*

Note. Percentages are based on the total count of targets in each category (infant-placed, parent-placed, and not-placed).

saliences that led to rapid search by the adult subjects, they strongly indicate the role of infant and parent behavior in biasing the input to make targets of interest easier to see. By analogy, parents and infants are biased to place the keys under the streetlight. Targets characterized by multiple saliencies are, in general, found more rapidly than those without; this is true for infant-placed, parent-placed, and for randomly selected not-placed targets. However, the likelihood of multiple saliences is increased for people-placed targets relative to the not-placed targets.

5. General Discussion

The main contributions of the study are these: First, although searching for targets in egocentric images is a difficult and different task from the search tasks used in more traditional studies of salience, most of the saliences documented in those paradigms were also strongly related to search time for targets in infant-egocentric images. This result provides a foundation for understanding visual attention in egocentric vision. While low-level contrast did not predict search time in the current study, many of the visual properties that attract attention in controlled laboratory studies also clearly matter for attention in uncontrolled egocentric images. Second, the target saliences most associated with rapid search were also the properties that differentiated targets that were placed in the scenes by the infant or parent from targets that were not placed. This suggests that the key saliences are ones that are easily influenced by the in-the-moment behaviors of people in the everyday world. This finding has implications for the functional role of saliences in human vision, their origin, the social structure of perceptual environments, and how the relation between bottom-up and top-down control of attention in these environments may support infant learning.

5.1. *The functions and origins of salience*

Bottom-up saliences that automatically attract visual attention are generally understood as reflecting evolutionary adaptations in support of species-important tasks (Itti & Koch, 2001), including alerting the orienting system to danger, food, or new information. The present findings indicate that human behavior influences the presence of these saliences. By moving their bodies and by moving objects, behavior directly affects the salience of objects in the egocentric view and in the social context for other perceivers. Many biological organisms move their sensory surfaces toward a target of interest in the service of the extraction of sensory information (Hofmann et al., 2013; Kleinfeld, Ahissar, & Diamond, 2006; Taub & Yovel, 2020; see also, Lungarella & Sporns, 2006). Primates systematically shift their gaze to targets of interest in the moment, a behavior that increases the target's visibility and precision as the fovea captures an image that is higher acuity and cortically overrepresented compared to the periphery (Azzopardi & Cowey, 1996; Stewart, Valsecchi, & Schütz, 2020). In brief, controlling the input through bodily actions is a common biological solution to efficient perception, albeit one that is currently understudied in human vision.

The four saliences that characterized placed objects in the egocentric scenes and were associated with rapid visual search all concern the spatial layout of the target in the scene: visual size, centering, and measures of clutter (set size and proportion of edges belonging to the target). These are scene properties directly affected by the spatial relation of the perceiver to visual objects in the three-dimensional world. This correspondence points to one potential origin for why these saliences play such a strong role even in adult attention. If people and their social partners regularly move their bodies and objects in ways that make momentary targets spatially prominent, then these saliences could systematically characterize relevant task information in the egocentric view, from infancy onward. There are two nuances worth noting in this

proposal. First, learning predictive relations through active vision need not exclude other explanations of how saliences guide attention, such as a high signal-to-noise ratio between the object and the scene (e.g., Itti & Koch, 2001; Wolfe & Horowitz, 2017). Instead, our findings suggest how the signal might have an increasing effect on attention over time if predictive relations exist between specific visual properties and what is relevant in the current moment. Second, changing the scene properties through changing the body does not need to be intentional in the sense of having the top-down goal to attend to a target. Certainly, placements by parents may be deliberate attempts to scaffold infant attention. However, the infant placements are less likely to be intentional bids for parent's attention and are also unlikely to initiate infant attention, occurring as they do at the end of the infant's handling of the object. These are clear open questions for future research. Nonetheless, parents' and infants' placements – occurring within an unspecified time window of infant attention to the placed object – shared many of the same saliences. The implication is that in the everyday life of infants, the social relevance of an object and attention to that object is likely to be correlated with a suite of visual properties.

In a seminal paper on human spatial knowledge, Clark (1973) argued that that the origins of spatial concepts would be found in the morphology of the human body, its posture, and its behavior in the 3D world, and pointed to critical roles for both evolutionary and developmental history in these origins. The statistics of human visual experience over both evolutionary and developmental time are intertwined with and determined by the body and its behavior. The bias to look to the center of scenes is likely related to the bias (in active vision) to look at the world with head and eyes aligned and the bias to instrumentally act on objects at the body's midline (Bambach et al., 2016; Franchak et al., 2021; Luo & Franchak, 2020). Similarly, the egocentric visual statistics of instrumental actions on objects may well be why it is so difficult to counter the attentional influence of visually large objects (e.g., Proulx & Egeth, 2008). The everyday statistics of egocentric vision could also be a factor in why low-level contrast was not related to either placement or to rapid search: this is a salience that may not be easily in-the-moment modified by perceivers across everyday tasks. In sum, the potency and continuing functionality of biases to look to visually large, centered, and foregrounded images may lie in how people act in the physical world to optimize the visual input for themselves and for others. The origin and functionality of these and other saliences such as contrast may lie in our more ancient history, in the sensitivities biased by the statistics of scenes of nature (Geisler, 2008) or evolutionarily experienced threats (Itti & Koch, 2001). These are open issues for future research including the examination of saliences in egocentric images in contexts beyond the one selected for this study.

5.2. Relations between bottom-up and top-down control of attention

Bottom-up attention is traditionally characterized in terms of automatic mechanisms that operate on low-level sensory input, whereas top-down attention is knowledge-based using internal goals and predictive relations in the input to guide attention (Connor, Egeth, & Yantis, 2004). In adults, the two processes interact dynamically to intelligently modulate attention, and a long line of experiments have manipulated the relative strength of these factors to determine how that competition is resolved (e.g., Corbetta & Shulman, 2002; Treisman, 2006). In infants, visual attention is often most strongly controlled by external saliences (Kwon, Setoodehnia, Baek, Luck, & Oakes, 2016; Pomaranski, Hayes, Kwon, Henderson, & Oakes, 2021) with top-down modulation developing incrementally from infancy well into childhood (Best & Miller, 2010; Colombo, 2001; Diamond, 2013). In late infancy, individual differences in top-down control of visual attention have been repeatedly shown to predict much later developments in the self-regulation of attention (Johansson, Marciszko, Gredebäck, Nyström, & Bohlin, 2015; Papageorgiou et al., 2014), which in turn relates to achievements in many other cognitive domains (Best, Miller, & Naglieri, 2011; Espy et al., 2004; Richland & Burchinal, 2013).

The present findings add new insights about this developmental trend from more externally-controlled to internally-controlled attention. By their first birthday, both infants and their parents play a role in scaffolding infants' attention by controlling the saliences in their input. Young infants' gaze is often pulled and held by external saliences (Atkinson, 1992; Johnson, Posner, & Rothbart, 1991); after their first birthday, however, infants make considerable strides in controlling their visual attention (Atkinson & Braddick, 2012; Ruff & Capozzoli, 2003; see Rosen, Amso, & McLaughlin, 2019 for a review). During this period of rapid growth, individual differences in the self-regulation of attention emerge that predict later executive functioning (Holmboe, Pasco Fearon, Csibra, Tucker, & Johnson, 2008; Morales, Mundy, Crowson, Neal, & Delgado, 2005; Rosen et al., 2019; Ross-Sheehy, Reynolds, & Eschman, 2020). A critical focus for future research should be how the control of external saliences by infants and their social partners impacts the development of attention. Momentary movements of the body and of objects of interest can reduce conflicting demands on attentional control by minimizing competition. One open hypothesis is that, with their increasing motor autonomy, and perhaps scaffolded by their parent's behavior in creating optimal contexts for selective attention, infants begin to gain control over their visual attention by first controlling the input. If so, then the development of the self-regulation of attention may not lie solely in resolving internal competitions but in learning to behaviorally reduce external competitions.

5.3. Social context of visual attention and its development

Both data and theory link individual differences in infants' top-down control of attention to parent behavior in guiding and scaffolding attention to objects (Bornstein, 1985). Parents have been shown to use behaviors such as looking at the object, gestures, handling of the object and talk to encourage infant sustained attention to an object (e.g., Baldwin, 1993; Suarez-Rivera et al., 2019; Yu & Smith, 2016). The present findings indicate that parents also scaffold their infants' attention by organizing the visual input – placing objects in ways that make them more salient than their surrounds. Increasing the external salience provides a bottom-up aid, which by some accounts helps young perceivers strengthen internal controls (Méndez, Yu, & Smith, 2021; but see Wass et al., 2018 for an alternate account). Rosen et al. (2019) have proposed that the moment-to-moment decisions of sustaining or shifting visual attention in late infancy train and strengthen neural connections from visual areas to the pre-frontal cortex, and have proposed that parent scaffolding of these moment-to-moment decisions play a key role in the emergence of individual differences with life-long consequences (see also Werchan & Amso, 2017). This hypothesis is a clear target for future work. The scaffolding of infants' visual attention by their mature partners in social interactions is also known to play a role in early word learning: words are more likely to be learned when parents name objects at moments in which the object is more visually predominant and contextually the most likely referent for heard name (e.g., Medina, Snedeker, Trueswell, & Gleitman, 2011; Pereira, Smith, & Yu, 2014). As far as we know, parents' role in creating perceptual scenes that visually direct attention to the intended referent has not been directly studied (but see Burling & Yoshida, 2019; Yoshida & Smith, 2008) but may also be an important factor in the development of self-regulated attention.

Are the present results unique to the developmentally important context of parent-infant social interactions? We suspect not; social partners are likely to play a strong role in determining the visual statistics of human experience throughout the lifespan. It well known that adults structure their own task environments to support cognitive processes such as memory, attention and ease of action (Ballard, Hayhoe, & Pelz, 1995; Clark & Chalmers, 1998; Evans, 1980; Hutchins, 1995; Kirsh, 1995), fixating on goal-relevant objects important to the task at hand, just before the moment when they will need them (e.g., Hayhoe & Ballard, 2005; Land & Hayhoe, 2001; Tatler et al., 2011). In social

contexts, such behaviors would support common ground and joint coordinated action. While there is little work studying the scene saliences of peer-to-peer object manipulation, the impact of hand gestures in adults has been well-studied. This behavior is communicative and influences the listener's understanding (Broaders & Goldin-Meadow, 2010; Hostetter, 2011). At the same time, considerable research suggests that gestures are produced primarily as an aid to the gesturer's own cognition (e.g., Goldin-Meadow, 2011; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Iverson & Goldin-Meadow, 1998 for a review); the gestures similarly affect the cognition of the listener because both conversational partners share similar psychologies. Analogous to this, moving and placing objects in an adult social context could be done to intentionally organize the scene statistics to affect a social partner's attention, but the results of such actions may often be unintentional, done primarily for the actor's own benefit.

5.4. Limitations and next steps

The current work lies at the start of a program of research on the functionality of bottom-up saliences in egocentric vision, their creation through human behavior, and their potentially critical role in infant visual attention and learning. As a first study, there are many open questions concerning the generality of findings across contexts, tasks, and the ages of the perceivers and their social partners. First, in the present study, we analyzed the saliences in egocentric images in which an object was placed by a participant, picked up, put down, and left in the infant field of view. We focused here in order to measure search time based on saliences without the additional cue of hand contact. Hands have been well-documented as strong top-down cues that guide where adults look in a scene (Niimi, 2020), but they also can create bottom-up saliences of motion, foregrounding, and visual size (see Burling & Yoshida, 2019; Schillingmann, Burling, Yoshida, & Nagai, 2015). Second, we examined these saliences in one important but highly specific developmental context: toy play between parents and infants. The generalizability of these observed saliences, as measured by adult search performances, needs to be extended to other egocentric scenes. Further analyses of the low-level visual statistics of these scenes is also in order, as there are likely other factors that guide attention and which may differentiate parent- and infant-created scenes beyond those we observed. Third, we used adult search performance as a functional measure of salience, and did so in an online search task, which may have added noise. In addition to comparing these results to adult performance in more controlled settings, a demonstration that these saliences are also used by infants is necessary, although challenging (see Amso & Johnson, 2006; Frank et al., 2014; Kwon et al., 2016). Fourth, we have proposed that the correlations between salience and attention to task relevant information may support the intelligent use of salience to guide attention in everyday tasks. The empirical un-packing of this idea will likely require going beyond visual attention to include the role of multi-modal saliences (holding and seeing), reward, and the social behavior of parents.

5.5. Conclusion: saliences and search in everyday life

A further and we believe theoretically important issue raised by the present study concerns the nature of the search task: the adult participants were shown a picture of an object presented in a canonical view and were asked to find that object in a scene in which the object could present an infinite number of views given orientation, occlusion, and in some cases bendable parts. The saliences that speeded the participants' searches are likely ones that matter in everyday life, in which the visual similarity to the memory for the searched-for object can be quite minimal. Prior analyses of similarity among the views of objects in egocentric images indicate that the views that parents and especially infants generate are different compared to the views typically presented in experiments and far more variable (Bambach et al., 2018; Pereira, James, Jones, & Smith, 2010). A critical question for understanding both object constancy and object recognition are the visual properties that support humans' outstanding visual recognition of objects. We conjecture that the saliences observed here – visual size, centering, and limited surrounding clutter – likely play a critical role in the extraction of shape properties that enable successful search for the same object despite the substantial variability in the image projected on the retina. This speculation highlights the importance of studying egocentric vision more generally. The egocentric view constitutes the experiences on which visual development depends and defines the day-to-day tasks that perceivers must solve. Understanding how our considerable knowledge about vision and attention from highly controlled laboratory studies plays out in the context of active vision is an exciting emerging area of research (Bonnen et al., 2021; Franchak et al., 2021). Ultimately, the results suggest that there are bottom-up saliences that can be readily controlled by perceivers and their social partners, and that these saliences may be potent in a variety of contexts because people create them in these contexts, tending to place the keys where they are easy to see.

CRediT authorship contribution statement

Erin M. Anderson: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Eric S. Seemiller:** Methodology, Software, Writing – review & editing. **Linda B. Smith:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition.

Data availability

All data reported in this paper is archived in https://osf.io/4cvw5/

Acknowledgments

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number 5 T32 HD007475 supporting Erin M. Anderson, as well as by the National Science Foundation, award number BCS-1842817, and NIH Award Numbers R01HD10462 and 1R01EY032897 awarded to Linda B. Smith. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Science Foundation. We are grateful to parents and infants who participated in the research. We thank members of the IU Cognitive Development Lab for their help. All data from the experiments reported in this paper is archived in OSF repository https://osf.io/4cvw5/

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105256.

References

Amso, D., & Johnson, S. P. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology, 42*(6), 1236–1245. https://doi.org/10.1037/0012-1649.42.6.1236

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods, 53*, 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. *Vision Research, 41*(25–26), 3613–3625. https://doi.org/10.1016/s0042-6989(01)00196-1

Atkinson, J. (1992). Early visual development: Differential functioning of parvocellular and magnocellular pathways. *Eye, 6*(2), 129–135.

Atkinson, J., & Braddick, O. (2012). Visual attention in the first years: Typical development and developmental disorders. *Developmental Medicine and Child Neurology, 54*(7), 589–595.

Azzopardi, P., & Cowey, A. (1996). The overrepresentation of the fovea and adjacent retina in the striate cortex and dorsal lateral geniculate nucleus of the macaque monkey. *Neuroscience, 72*(3), 627–639. https://doi.org/10.1016/0306-4522(95)00589-7

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*(5), 832–843. https://doi.org/10.1037/0012-1649.29.5.832

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Bambach, S., Crandall, D. J., Smith, L. B., & Yu, C. (2018). Toddler-inspired visual object learning. *Advances in Neural Information Processing Systems, 31*.

Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. (2016). Objects in the center: How the infant's body constrains infant scenes. In *, 2016. Joint IEEE International Conference on Development and Learning and Epigenetic Robotics* (pp. 132–137).

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

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641–1660.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327–336.

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research, 50*(23), 2577–2587. https://doi.org/10.1016/j.visres.2010.08.016

Bonnen, K., Matthis, J. S., Gibaldi, A., Banks, M. S., Levi, D. M., & Hayhoe, M. (2021). Binocular vision and the control of foot placement during walking in natural terrain. *Scientific Reports, 11*(1), 20881. https://doi.org/10.1038/s41598-021-99846-0

Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision Research, 91*, 62–77. https://doi.org/10.1016/j.visres.2013.07.016

Borjon, J. I., Schroer, S. E., Bambach, S., Slone, L. K., Abney, D. H., Crandall, D. J., & Smith, L. B. (2018). A view of their own: Capturing the egocentric view of infants and toddlers with head-mounted cameras. *Journal of visualized experiments: JoVE., 140*, Article e58445.

Bornstein, M. H. (1985). How infant and mother jointly contribute to developing cognitive competence in the child. *Proceedings of the National Academy of Sciences, 82*(21), 7470–7473.

Broaders, S. C., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science, 21*(5), 623–628. https://doi.org/10.1177/0956797610366082

Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision, 9*(3), 5.

Buetti, S., Xu, J., & Lleras, A. (2019). Predicting how color and shape combine in the human visual system to direct attention. *Scientific Reports, 9*, 20258.

Burling, J. M., & Yoshida, H. (2019). Visual constancies amidst changes in handled objects for 5-to 24-month-old infants. *Child Development, 90*(2), 452–461.

Cavallina, C., Puccio, G., Capurso, M., Bremner, A. J., & Santangelo, V. (2018). Cognitive development attenuates audiovisual distraction and promotes the selection of task-relevant perceptual saliency during visual search on complex scenes. *Cognition, 180*, 91–98. https://doi.org/10.1016/j.cognition.2018.07.003

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. http://www.jstor.org/stable/3328150.

Clark, H. H. (1973). Space, time, semantics, and the child. In *Cognitive development and acquisition of language* (pp. 27–63). Academic Press.

Clarke, A. D. F., & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research, 102*, 41–51. https://doi.org/10.1016/j.visres.2014.06.016

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology, 52*(1), 337–367.

Computer Workstations eTool. (2022). Retrieved from *https://www.osha.gov/etools/computer-workstations*.

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology, 14*(19), R850–R852.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience, 3*(3), 201–215. https://doi.org/10.1038/nrn755

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11*(5), 14. https://doi.org/10.1167/11.5.14

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*(1), 465–486. https://doi.org/10.1207/s15326942dn2601_6

Evans, G. W. (1980). Environmental cognition. *Psychological Bulletin, 88*(2), 259.

Fathi, A., Ren, X., & Rehg, J. M. (2011). Learning to recognize objects in egocentric activities. *CVPR 2011*, 3281–3288. https://doi.org/10.1109/CVPR.2011.5995444

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*, 215.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Franchak, J. M., McGee, B., & Blanch, G. (2021). Adapting the coordination of eyes and head to differences in task and environment during fully-mobile visual exploration. *PLoS One, 16*(8), Article e0256463. https://doi.org/10.1371/journal.pone.0256463

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology, 118*, 13–26.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59*, 167–192. https://doi.org/10.1146/annurev.psych.58.110405.085632

Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology, 81*(2), 194–215. https://doi.org/10.1006/jecp.2001.2649

Gilette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*(2), 135–176. https://doi.org/10.1016/s0010-0277(99)00036-0

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 595–607.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522.

Hayes, T. R., & Henderson, J. M. (2020). Center bias outperforms image salience but not semantics in accounting for attention during scene viewing. *Attention, Perception & Psychophysics, 82*(3), 985–994. https://doi.org/10.3758/s13414-019-01849-7

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7*(1–3), 43–64. https://doi.org/10.1080/135062800394676

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour, 1*, 743–747. https://doi.org/10.1038/s41562-017-0208-0

Hofmann, V., Sanguinetti-Scheck, J. I., Künzel, S., Geurten, B., Gómez-Sena, L., & Engelmann, J. (2013). Sensory flow shaped by active sensing: Sensorimotor strategies in electric fish. *Journal of Experimental Biology, 216*(13), 2487–2500.

Holmboe, K., Pasco Fearon, R. M., Csibra, G., Tucker, L. A., & Johnson, M. H. (2008). Freeze-frame: A new infant inhibition task and its relation to frontal cortex tasks during infancy and early childhood. *Journal of Experimental Child Psychology, 100*(2), 89–114. https://doi.org/10.1016/j.jecp.2007.09.004

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*(2), 297–315. https://doi.org/10.1037/a0022128

Hutchins, E. (1995). *Cognition in the wild* (no. 1995). In *MIT press. - Big Idea is Cognition is a System Rather than Individual Thing, Embedded in Culture*.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience, 2*, 194–203. https://doi.org/10.1038/35058500

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*(6708), 228.

Johansson, M., Marciszko, C., Gredebäck, G., Nyström, P., & Bohlin, G. (2015). Sustained attention in infancy as a longitudinal predictor of self-regulatory functions. *Infant Behavior and Development, 41*, 1–11.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience, 3*, 335–344. https://doi.org/10.1162/jocn.1991.3.4.335

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science* (p. 11). San Francisco: Chandler Pub. Co. ISBN 9781412836296.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 73*(1–2), 31–68.

Kleinfeld, D., Ahissar, E., & Diamond, M. E. (2006). Active sensation: Insights from the rodent vibrissa sensorimotor system. *Current Opinion in Neurobiology, 16*(4), 435–444.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Kwon, M.-K., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology, 52*(4), 537–555. https://doi.org/10.1037/dev0000080

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25–26), 3559–3565. https://doi.org/10.1016/s0042-6989(01)00102-x

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Lungarella, M., & Sporns, O. (2006). Mapping information flow in sensorimotor networks. *PLoS Computational Biology, 2*(10), Article e144.

Luo, C., & Franchak, J. M. (2020). Head and body structure infants' visual experiences during mobile, naturalistic play. *PLoS One, 15*(11), Article e0242009. https://doi.org/10.1371/journal.pone.0242009

Mann, S., Kitani, K. M., Lee, Y. J., Ryoo, M. S., & Fathi, A. (2014). An introduction to the 3rd workshop on egocentric (First-Person) vision. *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 827–832. https://doi.org/10.1109/CVPRW.2014.133

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *PNAS, 108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

Méndez, A., Yu, C., & Smith, L. B. (2021). One-year old infants control bottom-up saliencies to purposely sustain attention. *PsyArXiv*. https://doi.org/10.31234/osf.io/8x6ur

Morales, M., Mundy, P., Crowson, M., Neal, A. R., & Delgado, C. (2005). Individual differences in infant attention skills, joint attention, and emotion regulation behaviour. *International Journal of Behavioral Development, 29*(3), 259–263.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*(7031), 387–391. https://doi.org/10.1038/nature03390

Niimi, R. (2020). Interacting hands draw attention during scene observation. *Attention, Perception, & Psychophysics, 82*, 1088–1098. https://doi.org/10.3758/s13414-019-01881-7

Nuthmann, A., Clayden, A. C., & Fisher, R. B. (2021). The effect of target salience and size in visual search within naturalistic scenes under degraded vision. *Journal of Vision, 21*(4), 2. https://doi.org/10.1167/jov.21.4.2

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation, 8*(2).

Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychological Science, 25*(7), 1371–1379. https://doi.org/10.1177/0956797614531295

Pereira, A., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review, 21*(1), 178–185. PMCID:PMC3883952.

Pereira, A. F., James, K. H., Jones, S. S., & Smith, L. B. (2010). Early biases and developmental changes in self-generated object views. *Journal of Vision, 10*(11), 22. https://doi.org/10.1167/10.11.22

Pomaranski, K. I., Hayes, T. R., Kwon, M. K., Henderson, J. M., & Oakes, L. M. (2021). Developmental changes in natural scene viewing in infancy. *Developmental Psychology, 57*(7), 1025–1041. https://doi.org/10.1037/dev0001020

Proulx, M. J., & Egeth, H. E. (2008). Biased competition and visual search: The role of luminance and size contrast. *Psychological Research, 72*(1), 106–113. https://doi.org/10.1007/s00426-006-0077-z

Proulx, M. J., & Green, M. (2011). Does apparent size capture attention in visual search? Evidence from the Muller-Lyer illusion. *Journal of Vision, 11*(13), 21. https://doi.org/10.1167/11.13.21

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

van Renswoude, D. R., van den Berg, L., Raijmakers, M., & Visser, I. (2019). Infants' center bias in free viewing of real-world scenes. *Vision Research, 154*, 44–53. https://doi.org/10.1016/j.visres.2018.10.003

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science, 24*(1), 87–92. https://doi.org/10.1177/0956797612450883

Rosen, M. L., Amso, D., & McLaughlin, K. A. (2019). The role of the visual association cortex in scaffolding prefrontal cortex development: A novel mechanism linking socioeconomic status and executive function. *Developmental Cognitive Neuroscience, 39*, Article 100699. https://doi.org/10.1016/j.dcn.2019.100699

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 17. https://doi.org/10.1167/7.2.17

Ross-Sheehy, S., Reynolds, E., & Eschman, B. (2020). Evidence for attentional phenotypes in infancy and their role in visual cognitive performance. *Brain Sciences, 10*(9), 605. https://doi.org/10.3390/brainsci10090605

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology, 39*(5), 877–890. https://doi.org/10.1037/0012-1649.39.5.877

Sareen, P., Ehinger, K. A., & Wolfe, J. M. (2016). CB database: A change blindness database for objects in natural indoor scenes. *Behavior Research Methods, 48*(4), 1343–1348. https://doi.org/10.3758/s13428-015-0640-x

Schillingmann, L., Burling, J. M., Yoshida, H., & Nagai, Y. (2015). Gaze is not enough: Computational analysis of infant's head movement measures the developing response to social interaction. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 2104–2109.

Sebastian, S., Seemiller, E. S., & Geisler, W. S. (2020). Local reliability weighting explains identification of partially masked objects in natural images. *Proceedings of the National Academy of Sciences of the United States of America, 117*(47), 29363–29370. https://doi.org/10.1073/pnas.1912331117

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science, 14*(1), 9–17. PMCID:PMC3050020.

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development, 16*(3), 407–419.

Stewart, E. E. M., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision, 20*(12). https://doi.org/10.1167/jov.20.12.2. Article 2.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology, 55*(1), 96–109. PMID: 30489136 PMCID: PMC6296904.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 1–17. https://doi.org/10.1167/7.14.4

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision, 11*(5), 5. https://doi.org/10.1167/11.5.5

Taub, M., & Yovel, Y. (2020). Segregating signal from noise through movement in echolocating bats. *Scientific Reports, 10*(1), 1–10.

Treisman, A. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception & Psychophysics, 22*(1), 1–11. https://doi.org/10.3758/BF03206074

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron, 31*(4), 523–535. https://doi.org/10.1016/s0896-6273(01)00392-0

Võ, M. L.-H. (2013). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition, 126*(2), 198–212.

Wass, S. V., Clackson, K., Georgieva, S. D., Brightman, L., Nutbrown, R., & Leong, V. (2018). Infants' visual sustained attention is higher during joint play than solo play: Is this due to increased endogenous attention control or exogenous stimulus capture? *Developmental Science, 21*(6), Article e12667.

Werchan, D. M., & Amso, D. (2017). A novel ecological account of prefrontal cortex functional development. *Psychological Review, 124*(6), 720–739. https://doi.org/10.1037/rev0000078

Wolfe, J. M. (2020). Visual search: How do we find what we are looking for? *Annual Review of Vision Science, 6*, 539–562. https://doi.org/10.1146/annurev-vision-091718-015048

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, & Psychophysics, 73*(6), 1650–1671.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 419–433. https://doi.org/10.1037/0096-1523.15.3.419

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*, 0058.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy : the official journal of the International Society on Infant Studies, 13*(3), 229–248. https://doi.org/10.1080/15250000802004437

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*(2), 244–262.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS One, 8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology, 26*(9), 1235–1240 (PMID: 27133869).