Acta Psychologica 243 (2024) 104124

Contents lists available at ScienceDirect

Acta Psychologica

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

Early development of saliency-driven attention through object manipulation

Lichao Sun [a,*], David J. Francis [b], Yukie Nagai [c], Hanako Yoshida [a]

[a] Department of Psychology, University of Houston, TX, United States
[b] Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, TX, United States
[c] International Research Center for Neurointelligence, University of Tokyo, Tokyo, Japan

ARTICLE INFO

Keywords:
Sustained attention
Visual saliency
Head-mounted eye-tracking
Object play

ABSTRACT

In the first years of life, infants progressively develop attention selection skills to gather information from visually clustered environments. As young as newborns, infants are sensitive to the distinguished differences in color, orientation, and luminance, which are the components of visual saliency. However, we know little about how saliency-driven attention emerges and develops socially through everyday free-viewing experiences. The present work assessed the saliency change in infants' egocentric scenes and investigated the impacts of manual engagements on infant object looking in the interactive context of object play. Thirty parent-infant dyads, including infants in two age groups (younger: 3- to 6-month-old; older: 9- to 12-month-old), completed a brief session of object play. Infants' looking behaviors were recorded by the head-mounted eye-tracking gear, and both parents' and infants' manual actions on objects were annotated separately for analyses. The present findings revealed distinct attention mechanisms that underlie the hand-eye coordination between parents and infants and within infants during object play: younger infants are predominantly biased toward the characteristics of the visual saliency accompanying the parent's handled actions on the objects; on the other hand, older infants gradually employed more attention to the object, regardless of the saliency in view, as they gained more self-generated manual actions. Taken together, the present work highlights the tight coordination between visual experiences and sensorimotor competence and proposes a novel dyadic pathway to sustained attention that social sensitivity to parents' hands emerges through saliency-driven attention, preparing infants to focus, follow, and steadily track moving targets in free-flow viewing activities.

1. Introduction

Visual attention develops rapidly during the first year of life. Effective attentional control is critical for directing gaze from distractors in the visual environment, helping maintain attention focus on target tasks, and promoting the development of object recognition (Amso & Johnson, 2006; Bruce & Tsotsos, 2009; Feldman & Friston, 2010). The developing attention system is initially predominantly guided by exogenous processing in an involuntary manner, and young infants are sensitive to salience characteristics, such as color, shape, and motion contrasts in their visual fields (Amso & Johnson, 2006; Dannemiller, 1998; Fantz & Miranda, 1975; Ruff, 1996; Tellinghuisen et al., 1999). As infants gain better control of eye movements and bodily engagement, infants gradually become pervasive in controlling visual selection and generating sustained attention, which is linked with word learning and advanced cognitive control (Colombo & Cheatham, 2006; Csibra et al., 2001; Ruff, 1996). While such developmental growth has been shown as a result of advanced motor control, increasing knowledge of symbolic representation, and enhanced social competence in interpreting a broad use of social cues and guidance (Kannass & Oakes, 2008; Yu et al., 2019), a growing number of developmental evidence suggests that parent-infant object play serves as a fundamental training ground for attuning effective visual search at the early stage of cognitive development. Recent studies using head-mounted eye trackers document that parents proactively engage in object manipulation and influence the *visual properties* of infants' egocentric scenes, such as the visual size of the viewed object (Burling & Yoshida, 2019; Mendez et al., 2023; Yu & Smith, 2012), centering of the scene (Axelsson et al., 2016; Bambach et al., 2016),

* Corresponding author at: Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204-5022, United States.
E-mail addresses: lsun21@cougarnet.uh.edu (L. Sun), dfrancis@uh.edu (D.J. Francis), nagai.yukie@mail.u-tokyo.ac.jp (Y. Nagai), hyoshida@central.uh.edu (H. Yoshida).

https://doi.org/10.1016/j.actpsy.2024.104124
Received 17 February 2023; Received in revised form 30 December 2023; Accepted 2 January 2024
Available online 16 January 2024

distribution of saliencies (Anderson et al., 2022; Sun, 2022). Despite the evident influence of parents in shaping infants' viewing experiences, there is still much to learn about how infants respond to changes in visual properties, such as saliency, in the context of parental scaffolding. Using head-mounted eye-tracking technology, we characterized the saliency dynamics in the infants' visual field during dyadic object play and examined the effects of object manipulation on the development of saliency-driven attention.

1.1. Saliency features in early viewing experiences

Attention focus, often described as "spotlight", requires the detection of various visual features, which proceed simultaneously across the visual scenes (Itti & Koch, 2001; Treisman & Gelade, 1980). This attention processing involves two interrelated mechanisms: (1) an endogenous process derived through one's task goals, contextual knowledge, and prior experiences (e.g., Eriksen & Hoffman, 1972; Posner, 1980; Yantis & Egeth, 1999); and (2) an exogenous process that is involuntarily driven by physical saliency, which depends on the visual properties of stimuli and their corresponding contexts (Egeth & Yantis, 1997; Theeuwes, 1993, 1994; Yantis, 1993). Screen-based experimental paradigms systematically demonstrated that early visual attention is readily controlled by the physical saliency of stimulus over the screen (Borji & Itti, 2013; Itti et al., 1998; Itti & Koch, 2000). Specifically, developmental literature documents that infants' looking patterns are characterized by some saliency dimensions, and infants rapidly shift attention toward the stimuli with bright color, greater luminance, enhanced motion, and/or larger viewed size (Amso & Johnson, 2006; Dannemiller, 2000; Fantz & Miranda, 1975; Dannemiller, 2005; Oakes et al., 2002; Ruff, 1996; Tellinghuisen et al., 1999). In the first three months of life, newborns and young infants exhibited preferences for high contrast patterns and tended to look around the external counter of the stimuli on the screen (Frank et al., 2014; Wentworth et al., 2002). Furthermore, the sensitivity to the luminance and edge features in visual scenes persisted and remained useful in searching tasks for older children and adults (Açık et al., 2010; Karatekin, 2007; Madden et al., 2004).

Despite saliency-driven attention being extensively demonstrated in well-controlled screen-based experiments, we know relatively little about how saliency appears and becomes accessible in everyday viewing experiences. A growing number of observational studies employed head-mounted eye trackers and captured infants' momentary gaze fixation points in their corresponding scenery images (e.g., Franchak et al., 2011; Yu & Smith, 2016). These findings indicate the significant role of parent-child interactions as an early stage of social learning and highlight the visual contingencies between infants and parents during object play (Franchak et al., 2016; Smith et al., 2015; Suarez-Rivera et al., 2019). Examining infants' distinctive looking patterns associated with parental scaffolding makes it possible to assess visual dynamics, including saliency changes from the infant's point of view. Different from the experimental paradigm, which manipulates the target features of stimuli and controls the distractors on the screen, visual saliency in naturalistic viewing scenes constantly changes according to the physical appearances of stimuli (e.g., object size, color, texture, shape, etc.) and spatial distribution relative to the perceiver (e.g., visual angle). For example, the degree of spatial proximity between objects and infants is relevant to the duration of sustained attention and word-referent mapping (Samuelson et al., 2011; Axelsson et al., 2016). These findings highlight the dynamic perceptual experiences when one- to two-year-olds selectively attended to objects clustered in a visually salient environment and point to the importance of sensorimotor coordination in active visual exploration (e.g., Pereira et al., 2014; Yu & Smith, 2017).

1.2. The significance of sensorimotor development

The human visual experience is not a passive act of perceiving what surrounds us; instead, it occurs through the complex integration of neural, physiological, oculomotor, motor, social, cognitive, and multiple dynamic processes (e.g., Borjon et al., 2021; Chang et al., 2016; Colombo, 2001; Curtindale et al., 2019, p. 201; Franchak et al., 2018; McFarland et al., 2020; Mundy et al., 2000; Wass et al., 2018; Yu & Smith, 2016). From a social embodiment perspective, active visual selection is established by aligning one's eyes, head, and hands, often in collaboration with social partners.

Notably, infants' visual experiences and sensorimotor competence are tightly coordinated and influenced by each other. Visual experiences are spatially and temporally integrated with one's head (e.g., Borjon et al., 2021; Burling et al., 2013; Yu & Smith, 2012) and body movements (e.g., Iverson, 2022; Morse et al., 2015; Soska & Adolph, 2014). Infants' gaze and body movement are temporally bound together after birth (Robertson et al., 2001; Robertson & Johnson, 2009). As the sensorimotor system is still nascent, newborns and young infants cannot fully support their heads or bodies, and their visual exploration appears to be receptive and predominantly dependent upon parents' support (Deák et al., 2014; Ruff, 1996). For example, infants aged 3 to 11 months preferred to follow their parents' hand actions and attend to the handled objects rather than the static object alone during play (Deák et al., 2014, 2018). Object manipulation by parents not only provides visual accessibility but also affects the salient features of viewed objects (Brand et al., 2002; Koterba & Iverson, 2009). For example, the visual size of objects in the infants' egocentric scenes - a robust index of visual saliency - highlights the importance of the constant visual size of objects in maintaining attention (Mendez et al., 2023; Burling & Yoshida, 2019). One possible explanation is that the enhanced saliency features affected by object manipulation stabilize young infants' attention toward the handled object, which can be a potential mechanism underlying the role of parental scaffolding in navigating and maintaining infant attention during the first year of life.

As infants initiate more head and body actions and develop more flexible postures, their visual input becomes more dynamic and well-structured. Reaching emerges around 4–5 months of age, accompanied by sitting, as infants gain head and upper trunk control and become able to move their hands without support (Harbourne et al., 2013; Thelen et al., 1993). As infants eventually reach independent sitting between 5 and 8 months (Group & de Onis, 2006; Rochat & Goubet, 1995; Soska et al., 2010), they spend more time reaching, grasping, and handling objects (Corbetta et al., 2000; Rachwani et al., 2015; Soska & Adolph, 2014; also see a review by Iverson, 2022). Starting from the second year of life, children progressively exhibit more sophisticated manual actions on objects (Basilio & Rodríguez, 2017; Heiman et al., 2019; Ungerer & Sigman, 1984). The advances in motor competence incorporated with developing selective attention skills facilitate hand-eye coordination (e.g., Yu & Smith, 2017; Perone et al., 2008) and head-eye coordination (e.g., Schillingmann et al., 2015; Yoshida & Smith, 2008). The developmental change in the ways infants proactively interact with objects not only influences how objects appear in the infant's visual field but also helps stabilize infant attention during the exploration of various features of object representation (Adolph, 1997Burling & Yoshida, 2019;Kretch et al., 2014 Pereira et al., 2014).

In summary, the coordination between visual attention and motion systems establishes the stability of attention aligned with eyes, head, body, etc. (e.g., Hietanen, 2002; Robertson et al., 2001; Yoshida & Smith, 2008), and the developing sensorimotor skills greatly determine what visual input is accessible in the infant's egocentric scenes (Bambach et al., 2016; Kretch & Adolph, 2015). Considering the role of saliency-driven attention in early visual selection, the present study proposes a dyadic pathway to early sustained attention in the context of parent-infant object play: infant object looking is accounted for by the enhanced saliency aligned with object manipulation; and, the source of object manipulation dramatically changes from parents to infants during the first year of life.

1.3. Study goals & hypotheses

To understand the development of saliency-driven attention in naturalistic viewing, the present study investigated the effect of visual saliency on infant object looking in relation to object manipulation by the parent and/or infant during play. We applied the head-mounted eye-tracking method to capture the infant-centered visual scenes and estimated the momentary gaze allocation as well as saliency contours within infants' scenes. To examine the developmental change in saliency-driven attention, we compared the distribution of infant object looking between younger (3- to 6-month-old) and older (9- to 12-month-old) infants. These two age groups were established according to the developmental framework of early visual attention (e.g., Ruff, 1996) and the key milestones of motor development during these transition periods (e.g., reaching and sitting, object manipulation, Rochat, 1989; Rochat & Goubet, 1995; Iverson, 2010).

We have two main hypotheses on the impacts of saliency-driven attention: (H1) both age groups would exhibit attention toward objects aligned with enhanced saliency, and younger infants would be more sensitive to the saliency features in their visual scenes; (H2) saliency-driven attention varies as a joint function of object manipulation (by parent and/or infant) and age group. Specifically, we expected that younger infants would exhibit more saliency-driven attention on objects handled by parents in contrast to older infants who would look more at the self-manipulated object regardless of the saliency levels in the visual scenes.

2. Method

2.1. Participants

A sample of 30 parent-infant dyads, including 15 dyads classified into the younger group (infants with a mean age of 4.9 months, $SD = 1.1$; 7 males) and 15 dyads in the older group (infants with a mean age of 10.9 months; $SD = 1.2$; 10 males), completed the study. The sample of dyads was broadly representative of the ethnicity in the local community: non-Hispanic Caucasian (33 %), Hispanic (33 %), African American (13 %), Asian (3 %), bi-racial (13 %), and unknown (3 %). All the participating infants were typically developing, and they were all from comparable middle socioeconomic backgrounds (i.e., the median annual household income in Houston; U.S. Census Bureau, 2020). The present study was conducted according to guidelines laid down in the Declaration of Helsinki and its later amendments. The experimenter explained the study procedures to parents and obtained written informed consent forms before tasks. As a token of appreciation, all the participating dyads were provided a gift bundle. All procedures involving human subjects in the present study were approved by the Institutional Review Board at the university where the research took place.

2.2. Task setting and materials

Participating dyads were observed in a laboratory setting that contained a child-friendly set of a table and chairs, a container of eight toy objects, a list of words, eye-tracking equipment, recording cameras, and computer devices. Experimenters observed the play session behind the black curtains, which were also used to cover the walls and door to minimize potential distractions, such as unfamiliar equipment and devices in the room. In the play session, the parent and the infant sat across a 60 cm × 60 cm × 40 cm table, which was used as a surface for jointly interacting with the toy objects (Fig. 1A). The toys were matched with the word list (i.e., bunny, eat, cookie, car, put, drink, open, bear), and they were all selected according to the early comprehended words from the McArthur Child Development Inventory (Fenson et al., 1994, 2000).

2.3. Procedures

Parents were told that the goal of the object play was to observe how parents interact with children as naturally as possible, and they were encouraged to freely use toy objects and incorporate a specified word according to the audio instruction. As such, we expected parents to have sufficient opportunities to present all the toys and show variation in

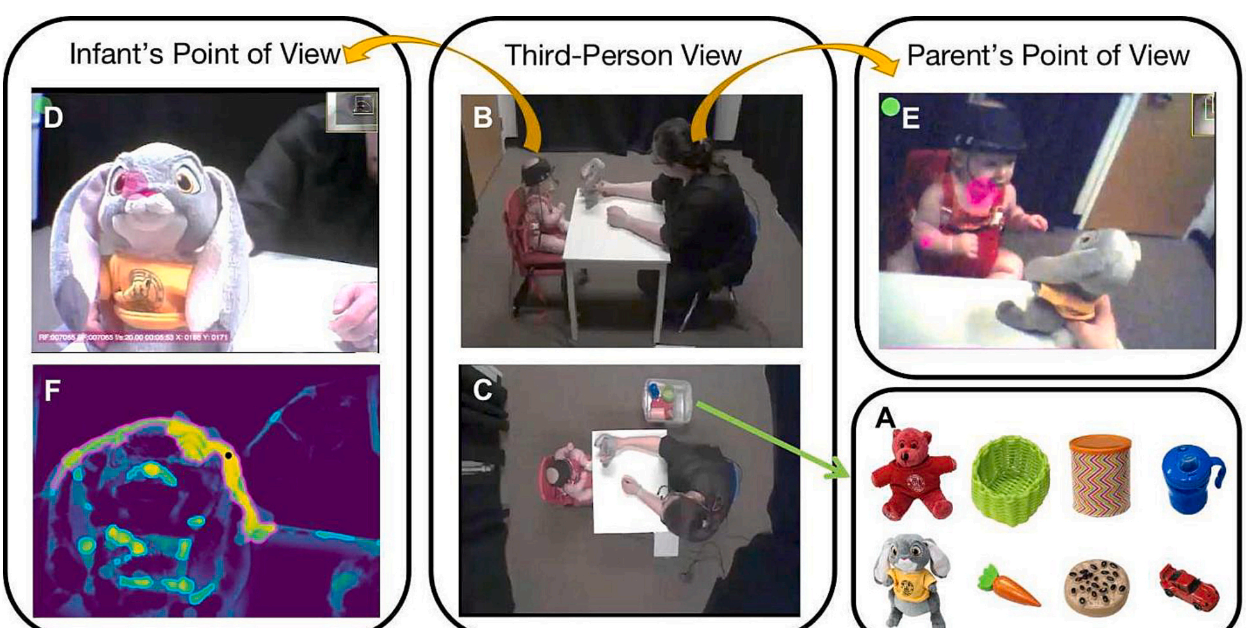

Fig. 1. The recording structure of a parent–infant play session. (A) Eight toy objects used throughout the play session. (B) Third-person view from the wall-mounted camera. (C) Third-person view from the ceiling camera. (D) The infant's view from the head-mounted camera (with a pink dot indicating the child's fixation point). (E) The parent's view from the head-mounted camera (with a larger pink dot indicating the parent's fixation point and a smaller pink dot representing the estimated gaze movement with respect to the corneal reflection). (F) The saliency map of the infant's corresponding scenery image. The top salient contour is labeled with the pink color with the center as a black dot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

their manual engagements during play. Experimenters also provided an explicit example before the study started, "Feel free to use whatever toys you choose. The first theme word is 'bunny'. You can show the (toy) bunny alone, or you can make (toy) bunny play with (toy) bear together." Parents would hear a word prompt every 40 s, and thus, the play session consisted of 40 s × 8 words for a total playtime of 5 min and 20 s, which would generate approximately 9600 frames per dyad for further behavioral annotations and analyses.

2.4. Equipment

To capture egocentric views and corresponding gaze movements, both the parent and the infant were equipped with a head-mounted Watec (WAT-230A) miniature color camera gear with a supplementary eye tracker. The head-mounted camera placed on the forehead recorded the visual field from the individual's egocentric views (54.4° horizontal by 42.2° vertical; e.g., Pereira et al., 2010; Smith et al., 2015; Yoshida & Smith, 2008). The eye tracker was adjusted and faced the individual's right eye in order to specify the focus of attention by its corresponding corneal reflection (e.g., Burling & Yoshida, 2019; Franchak et al., 2011; Sun & Yoshida, 2022). Correspondence between the images from the head-mounted camera and eye tracker was achieved using a manual calibration procedure. Before and after the play session, research assistants would randomly point to the stickers on a 60 cm × 40 cm board using a rattling toy to direct an infant's attention to target areas. Parents also completed the same calibration procedure by following pointing on the board.

To fully capture the individual's hand actions during the play session, two digital video cameras and an audio recorder were mounted on the wall and the ceiling to capture an overall view of the play scene (Fig. 1B & C). The videos of the third-person scenes were synchronized with the videos captured by the head-mounted cameras and then merged into a single video according to the shared timestamp by Adobe Premiere (Fig. 1D taken from the infant's egocentric views and Fig. 1E taken from the parent's egocentric views). All the exported videos had a resolution of 640 × 480 pixels and were recorded at a sampling rate of 30 Hz. The following analyses focused on the moments within the play session only, and we excluded frames with eye blinks or interruptions due to camera adjustment. On average, each parent-child dyad had 5634 frames ($SD =$ 206) recording both parents' and infants' visual scenes as well as its moment-to-moment attention allocation.

2.5. Data processing

To examine the change in saliency-driven attention, we determined infants' moment-to-moment gaze behaviors and calculated saliency features in their corresponding egocentric scene. Specifically, there were three major steps in data processing: (1) the estimation of the fixation point in the egocentric views, (2) the calculation of the saliency distribution, and (3) the behavioral annotation of the referential input.

2.5.1. The estimation of gaze location

To estimate momentary changes in the direction of eye gaze in infants' scenes, both the scenery views from the head-mounted camera and the eye views obtained by the eye-tracker were synchronized by Yarbus software (Positive Science). A minimum correlation of 0.85 between images was obtained during video processing. The location of an infant's eye gaze in the frame was represented in the coordinates of x and y in the image of 640 × 480 pixels. Any estimated gaze location landing outside of the frame was excluded from further analyses (e.g., Kretch & Adolph, 2015).

2.5.2. Saliency distribution computation

To capture the momentary saliency changes, the videos of the infants' head-camera views were fed into a saliency computation program (Nagai & Rohlfing, 2009), which employed Itti's saliency algorithm (e.g., Itti et al., 1998; Itti & Koch, 2000). First, the computation model extracted five primitive features, including color, intensity, orientation, flicker, and motion, by linearly filtering an input video. Specifically, color, intensity, and orientation are static features performed on each frame separately (e.g., Itti et al., 1998; Nagai & Rohlfing, 2009). In contrast, both the flicker and motion channels detect moving locations by examining temporal variabilities in the intensity of pixels across frames (e.g., Mahdi et al., 2018; Schlesinger et al., 2007). The integration of static and dynamic features allows Itti's algorithm to generate more comprehensive saliency maps that highlight both spatially distinct and temporally changing regions. Here, we equally weighed all the conspicuity maps from five channels and combined them into a single saliency map with salient contours valued from 0 to 1.

Consequently, the contours of salient regions and their estimated saliency values were computed in each frame of an infant's egocentric views. The present study primarily focused on the location of the top salient contour and its corresponding value for each frame of the infant's scenes. To further identify the correspondence between the saliency map and the estimated location of the infant's gaze, the frames of attention attributed by the visual saliency were identified when the area of the top salient contour overlapped with the spotlight circle of the gaze point with a radius of 15 degrees (Bambach et al., 2013a, 2013b; Kretch & Adolph, 2015). For example, Fig. 1D&F illustrates a moment of saliency-driven attention during play: the estimated gaze was allocated at the object handled by the parent, which also corresponded with the top salient contour in the saliency map.

2.5.3. Behavioral annotation

To further identify parents' and infants' manual engagement during the play session, frame-by-frame behavioral annotation was completed in the Datavyu coding software (Datavyu Team, 2014). Two well-trained coders who were blind to the purpose of the study annotated the individual's gaze patterns and hand actions for both the parent and the infant in each dyad. All the individual's target behaviors, including (1) gaze behaviors and (2) object handled actions, were annotated individually and were further time-stamped together according to the timeline of the play session (see examples in Fig. 2).

Reliability was measured by randomly selecting 25 % of the frames for each dyad and checking inter-rater coding agreement for the behaviors of interest. For example, the inter-rater reliability of the infant's gaze behaviors was 88.7 % ($SD = 4.0$ %, ranging from 82.9 % to 94.9 %) as assessed by Cohen's kappa of 0.80 ($SD = 0.07$), indicating strong agreement among raters (Cohen, 1968; McHugh, 2012). Additionally, the inter-rater reliability also falls into the reliability range obtained in other eye-tracking studies (e.g., 89 % for Sun & Yoshida, 2022; 84 % for Yoshida et al., 2020; 82 % to 95 % in Yu & Smith, 2017; 83 % in Chang et al., 2016).

2.5.3.1. Gaze behaviors (i.e., parent and infant object looking). Infants' gaze patterns were annotated according to the four regions of interest (ROIs) as follows: (1) toy objects, (2) parent's face, (3) parent's hands, and (4) their own hands (see Fig. 2). These four ROIs have been demonstrated to be most frequently shown in the egocentric view in the contexts of object dyadic play (e.g., Deák et al., 2018; Sun et al., 2022; Sun & Yoshida, 2022; Yu & Smith, 2013; Yu et al., 2019). Specifically, we extracted the episodes of infant attention on target objects when sustained attention has been demonstrated to link with early social learning (Kannass & Oakes, 2008; Tomasello & Farrar, 1986; Yu et al., 2019). To further verify the scaffolding role of the parent's handled actions on the attended object, we identified and counted the number of attention shifting from other ROIs prior to attention on the target object.

2.5.3.2. Object handled actions. Both parents' and infants' hand actions in relation to the toy objects were annotated frame-by-frame from both the wall and ceiling cameras. The moments of handled action were

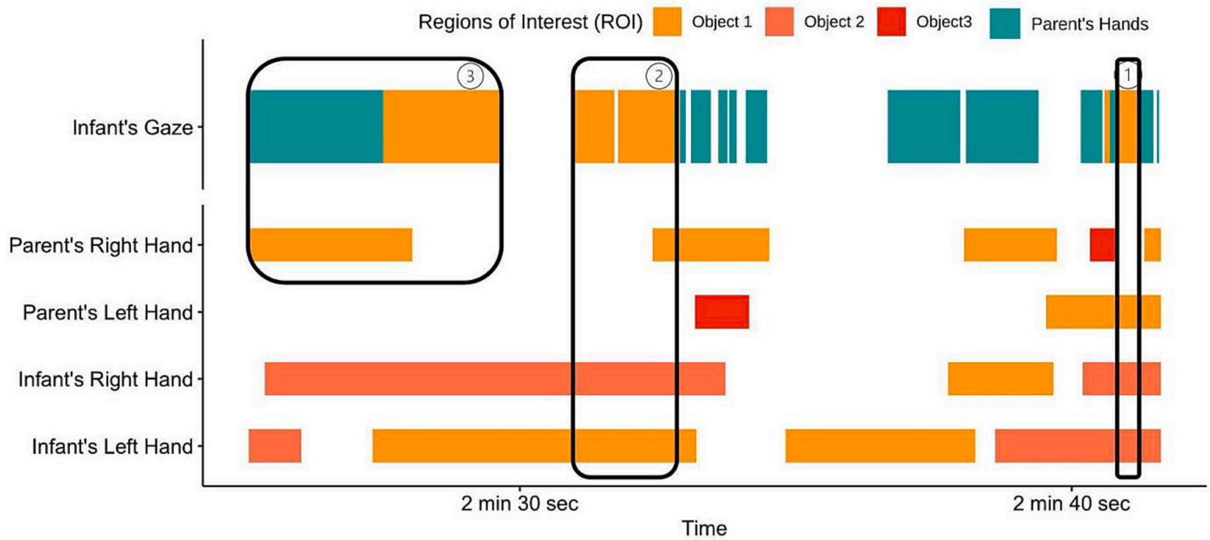

Fig. 2. The multimodal behavioral annotation of a video clip of a parent–infant play episode. ① An example of the infant maintaining attention to the object handled by the parent, whereas the infant is holding other toy objects. ② A moment of the infant attending to the object held by the infant. ③ An example of infant gaze shifting from the parent's hands to the object.

identified by the onset/offset of an individual starting to reach, handle, and/or manipulate any toy objects. Individual's two hands (right and left) were annotated separately and labeled with respect to the type of object (see Fig. 2). Because the present research was primarily concerned with the episodes of infant object looking, we compared and matched all the hand actions based on the corresponding object to which the infant attended and further annotated the presenter by defining which person displayed the attended object in the infant's scenes. Example ① in Fig. 2 presents the case when the parent and infant handled different objects, and we explicitly defined the parent as the presenter, corresponding to the identical object attended to by the infant at this moment. In addition to the presenter of the attended object, we also specified the collaborative moment that the attended object was handled by parents and infants together. This annotation is important to validate the emergence of hand-eye coordination with parental assistance and to further examine changes in the extent of the infant's manual engagement over age (e.g., Burling et al., 2013; Elmlinger & Yu, 2019; also see Example ② in Fig. 2). In sum, all the episodes of object handling were identified by the following conditions: (1) none of object handling, referring to the moments of the object standing on the table alone; (2) object handled by the parent alone; (3) object handled by the infant alone; and (4) object handled by the infant and parent together.

2.6. Analytic approach

The subsequent data organization and analyses were conducted in the R environment (R Core Team, 2022). Considering the varied play styles among and within parent-infant dyads (e.g., Bakeman & Adamson, 1984; Sosa, 2015), a series of hierarchical generalized linear models were applied, and the dependent variables were clustered by dyads. Thus, lmer and glmer functions of the R package lme4 (Ver 1.1–27.1, Bates et al., 2015) were used for estimating the present mixed-effect models fitting with Poisson distribution for count data with a log function.

By hypotheses, infant object looking might be primarily supported by the visual saliency features and changed over time. The left-skewed distribution of the top salient score in the present work and adult literature (e.g., Marat et al., 2009) highlights the importance of securing a relative measure of saliency variability, especially given our focus on age group differences in saliency-driven attention. Accordingly, we centered the scores to the grand mean of the top salient contours across all the scenes (e.g., Kreft et al., 1995) and classified the scenes by the relative saliency into three levels: "lower than 1 SD [-1 SD]", "on average [Mean]", and "higher than 1 SD [+1 SD]". As additional motion and contrasts were involved when object manipulation occurred, we expected that infants in both age groups would attend to the objects in a relatively higher salient scene, and younger infants would generate more object looking even with relatively lower saliency levels (H1). Corresponding with advanced motor skills and social competence, older infants would initiate more self-generated actions and attend more to the objects across various saliency levels (H2). In comparison, younger infants tended to have more saliency-driven attention to the objects handled by parents (H2).

3. Results

3.1. Preliminary analyses

Before testing our hypotheses on saliency-driven attention, we first characterized the distributions of infant object looking with respect to object manipulation. Developmental literature indicates that infants exhibit a predominant transition in the role of dyadic play (e.g., Bakeman & Adamson, 1984; Soska & Adolph, 2014). Thus, our first analyses aimed to validate the developmental changes that younger infants' object looking would be primarily guided by parental object manipulation. In contrast, older infants were expected to become more proactive in reaching and handling actions, resulting in more gaze shifting and longer attention to self-manipulated objects.

3.1.1. The distribution of object manipulation over development

To specify the role of the presenter - parents, infants, or together - in displaying objects, we found a significant interaction between the presenter and age on object manipulation, $\beta = 1.64$, 95 % CI[*1.60, 1.70*], $p < .001$. The follow-up pairwise contrasts confirmed a robust effect of the presenter: parents in both age groups predominantly displayed target objects during play, followed by the parent and the child handled together, the parent and the child handled different objects, by the infant his/herself, and object stood alone by the least. Moreover, parents of younger infants displayed significantly more objects than those of older infants, *z-ratio* $= 4.55$, $p < .001$. Compared with younger infants, older infants not only handled more objects independently (*z-ratio* $= 10.49$, $p < .001$) but also interacted with objects differently from their parents at the same moment (*z-ratio* = 2.68, $p = .007$; also see Fig. 3).

3.1.2. The development of infant gaze shifting toward objects

To demonstrate the significant role of handled actions in guiding infant object attention, we also examined the distribution of gaze shifting across the age groups: How and where did the infant shift attention prior to the object the most? This question is important to address the social origin of attention following and evidence the importance of manual engagements in supporting visual selection under naturalistic circumstances (Deák et al., 2014; Yu & Smith, 2015). Here, we counted infant attention shifting according to its prior attention locations, i.e., parent's face, parent's hands, or child's hands (see Example ③ in Fig. 2). Infants in both age groups had a similar distribution of gaze shifting in the following order, from the greatest to least: parent's hands ($M = 20.05$, $SD = 3.83$), parent's face ($M = 5.21$, $SD = 1.10$), and the child's own hands ($M = 1.31$, $SD = 0.35$). In addition, older infants shifted more from their own hands to the target objects than younger infants did, *z-ratio* = 4.01, $p < .001$.

3.1.3. Infant object attention accompanied by handled actions

There was a significant interaction between the presenter and the age group on the number of frames in which infant object looking was accompanied by object manipulation, $\beta = -0.68$, 95 % CI[$-0.73, -0.61$], $p < .001$. The post-hoc contrasts revealed that infants predominantly spent more attention on the target object in coordination with the parent's handled actions ($M = 1130.69$, $SD = 167.10$), followed by the handled actions by the parent and the child together ($M = 282.0$, $SD = 41.82$), child alone ($M = 195.90$, $SD = 29.11$), and object stood on the table alone ($M = 238.40$, $SD = 35.38$) by the least. Specifically, older infants had significantly more frames that their attention was allocated to the object handled by themselves in contrast with younger infants, *z-ratio* = 8.38, $p < .001$.

3.2. H1: infant object attention accompanied by saliency change

Table 1 summarizes the effects of saliency and age group on infant object attention and highlights the significant interactions, $\beta = 0.62$, 95 % CI[$0.58, 0.67$], $p < .001$. Specifically, infants in both age groups showed a similar saliency order as follows: infants had significantly more attention toward objects with a saliency score of 1 SD above, followed by a mean saliency score, and a score of 1 SD below by the least. In other words, infants spent more attention on the target objects with increasing saliency in their visual scenes. Moreover, the contrasts between age groups by saliency showed that younger infants had more attention to objects in the scenes with a mean salient score or below than older infants, *z-ratio* = 2.12, $p = .034$ (see also Fig. 4).

Table 1
The model summary of infant object attention predicted by saliency changes and age group.

Variable	Estimate	Std. error	Statistic	*P*-value	Conf. Low	Conf. High
(Intercept)	5.86	0.15	39.58	<.001	5.57	6.16
Saliency levels	0.77	0.02	50.87	<.001	0.74	0.80
Age group: older	−0.44	0.21	−2.12	.034	−0.86	−0.03
Saliency × age group	0.48	0.02	20.34	<.001	0.44	0.53

3.3. H2: The effect of visual saliency in predicting infant object attention in relation to object manipulation by parents and/or infants

Furthermore, we investigated the effects of saliency and object manipulation in infant object looking by counting the number of frames with infant object looking during the play session. There was a significant three-way interaction among saliency class, presenter, and age group (see the model in Table 2). Specifically, younger infants spent significantly more time on the handled object by parents than the older ones, especially when the relative saliency score was below the average (−1 SD: *z-ratio* = 4.47, $p < .001$). When the object was handled by infants alone, older infants had significantly more object looking than younger ones did, regardless of the saliency change in the visual scenes (−1 SD: *z-ratio* = 9.96, $p < .001$; Mean: *z-ratio* = 7.57, $p < .001$; +1 SD: *z-ratio* = 7.92, $p < .001$). When parents and infants handled the attended object together, infants in both age groups had comparable amounts of saliency-driven attention across the levels of visual saliency (also see Fig. 5).

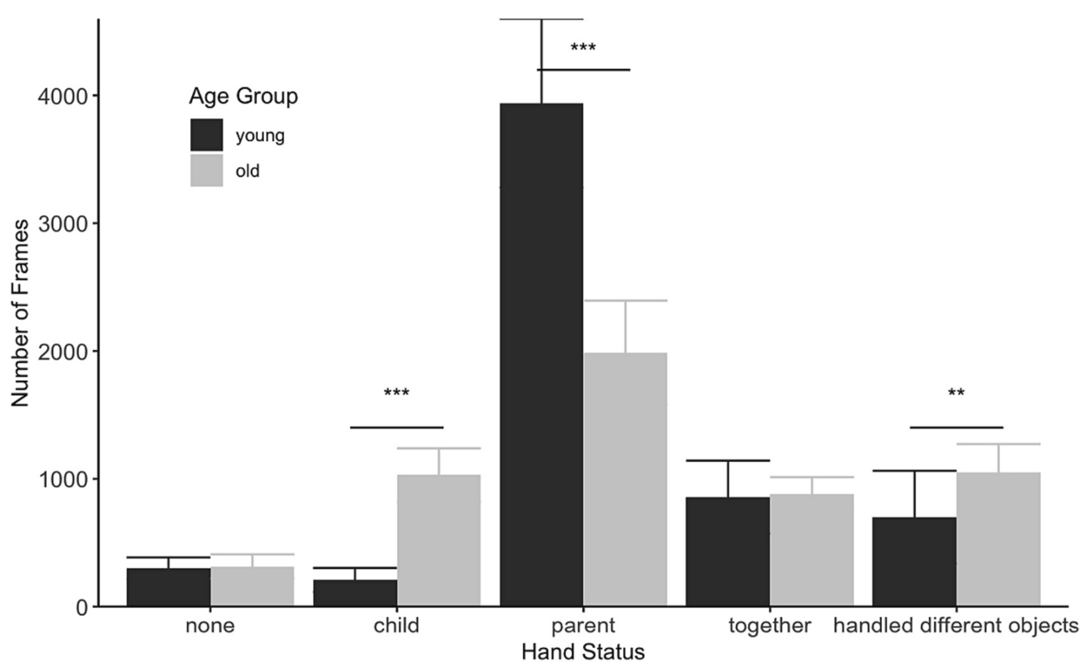

Fig. 3. The distribution of object handling throughout the play session. The error bar represents the standard error. The horizontal line highlights the compared age groups in conditions. The asterisk represents the *p*-value of statistical testing: ** for <.01, *** for <.001.

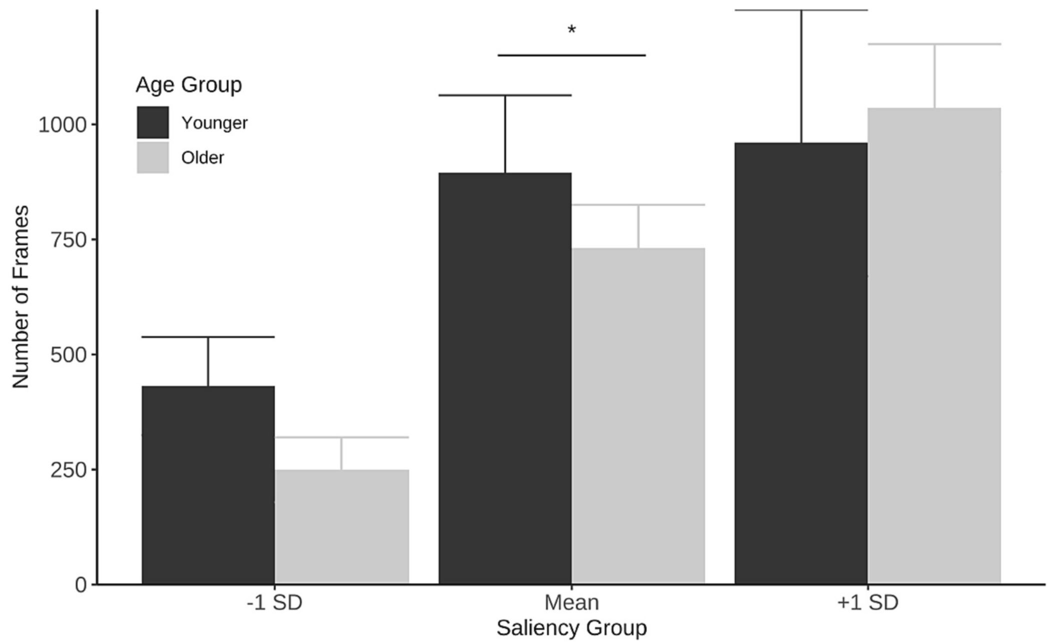

Fig. 4. The distribution of infant object attention accompanying the saliency change of the infant's visual scenes. The error bar represents the standard error. The horizontal line highlights the compared age groups in conditions. The asterisk represents the *p*-value of statistical testing: * for <.05.

Table 2
The development of infant object attention as a function visual saliency and the handled actions by parents and/or infants.

Variable	Estimate	Std. error	Statistic	P-value	Conf. Low	Conf. High
(Intercept)	4.38	0.15	29.21	<.001	4.09	4.67
Saliency	0.06	0.04	1.84	.25	−0.02	0.13
Presenter	−0.79	0.06	−3.98	<.001	−0.91	−0.68
Age group: older	−0.92	0.22	−4.30	<.001	−1.35	−0.50
Saliency × presenter	1.08	0.07	17.01	<.001	0.95	1.22
Saliency × age group: older	1.06	0.06	16.91	<.001	0.93	1.18
Presenter × age group: older	1.36	0.08	14.03	<.001	1.20	1.53
Saliency × presenter × age group: older	−0.93	0.10	−9.03	<.001	−1.12	−0.75

4. Discussion

The present study characterized the dynamics of visual saliency within the infant's scenes in the naturalistic play environment and indicated the establishment of hand-eye coordination between parents and infants and within infants in the first year of their lives. These findings highlighted that saliency-driven attention varied in accordance with object manipulation by parents and/or infants. Specifically, younger infants (3- to 6-month-old) had significantly more object looking with enhanced saliency, primarily engaging under the guidance of parental handled actions. In contrast, older infants (9- to 12-month-old) distributed more attention to the target objects with their self-generated hand actions, and their saliency-driven attention on objects remained only when the salient value was 1 SD above the average.

The development of selective attention is coupled with the rapidly changing sensorimotor system (e.g., Borjon et al., 2021; Burling et al., 2013). When the control of the head and body movements is still developing, the limited motor skills constrain the visual input of the egocentric scenes (Bambach et al., 2013a). Thus, visual information in the infants' field of view is predominantly organized by parents' actions. For instance, over 80 % of infants' visual scenes in the present work were coordinated with object handled actions by parents. The accompanying parental object manipulation is consistently predominant within the naturalistic context of dyadic object play throughout the first year and even later, and it readily acts as a visually robust reference for attention sharing between infants and their social partners (Deák et al., 2014, 2018; Sun & Yoshida, 2022; Yu & Smith, 2017). The impact of parental manual engagement in early visual exploration, as demonstrated across various theoretical perspectives, underscores the significance of multimodal references in facilitating sustained attention and associated learning outcomes (Deák et al., 2018; Suarez-Rivera et al., 2019; Sun & Yoshida, 2022; Yu et al., 2019).

Here, we offer a socially embodied pathway to infants' exogenous attention on objects through parent object handling. In this framework, parents' manual engagement enhances saliency value around the object, ensuring the target object visually stands out from its surroundings. In turn, this increases the likelihood of infants directing their attention toward the handled object by parents. We propose two major contributions of parental manual engagement in facilitating early perceptual experiences within social interactive activities. First, similar to screen-based tasks (e.g., Amso & Johnson, 2006; Kwon et al., 2016), visual saliency features contributed to infant attention in the context of naturalistic viewing (also see Anderson et al., 2022; Kretch & Adolph, 2015). Parents' manual engagement strengthened saliency through heightened motion contrasts, increased visual size, and optimal spatial positioning of target objects, representing an important resource of exogenous features of infants' visual experiences to compel their attention. As the saliency value increased, infants in the present work tended to look more at the salient object. Specifically, younger infants demonstrated more attention attributed to the saliency change along with parents' hand actions. These findings suggest that exogenous processing serves as a significant sensorimotor path for infants to establish attention shift from hands and maintain stable attention on the target of interests through hands-on interaction.

Second, object manipulation can act as a training ground for refining visual selection and developing advanced cognitive control over attention. The visual system is notably attuned to spatial and temporal changes within the motor system, which is constantly informed about the layout of one's surroundings (Barrett et al., 2005; Bassili, 1976;

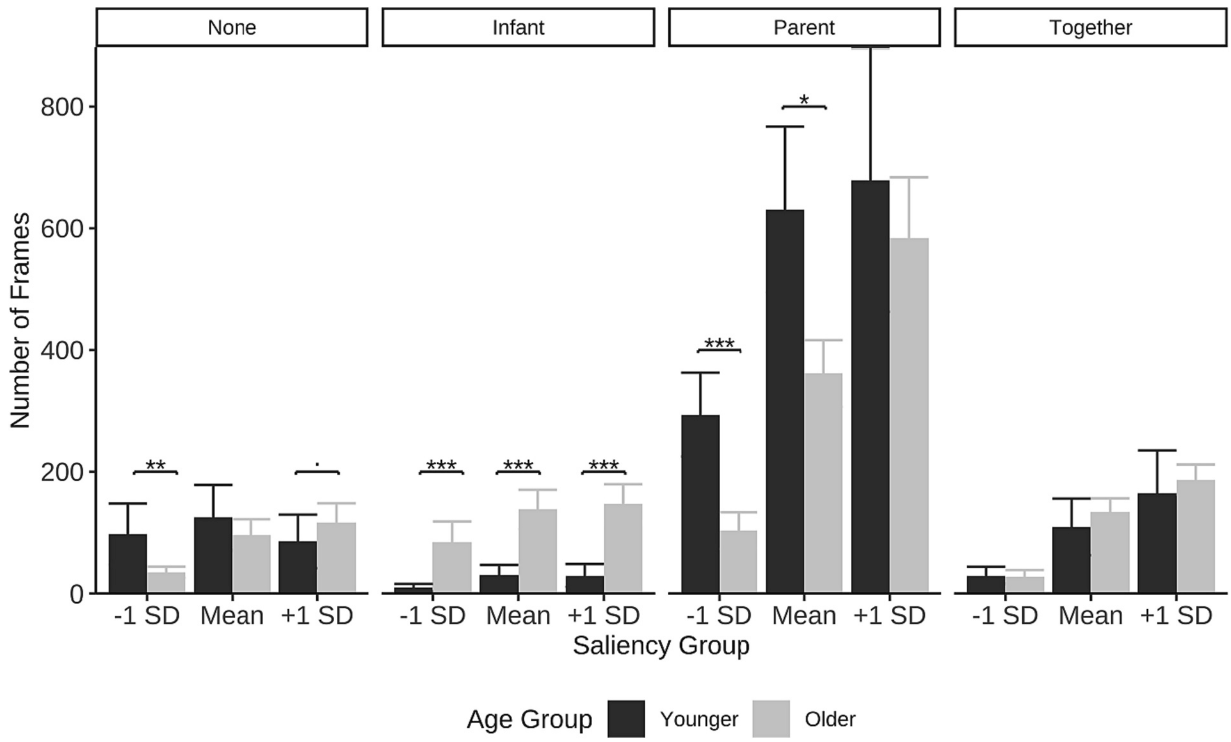

Fig. 5. The distribution of Infant object attention by object handling and saliency. The error bar represents the standard error. The horizontal line highlights the compared age groups in conditions. The asterisk represents the p-value of statistical testing: for =.05, * for <.05, ** for <.01, *** for <.001.

Pashler, 1991). Parents' hand actions can initiate the visual spotlight around the target, triggering and maintaining infants' interest in objects and sequentially adapting attention to moving items in visually cluttered environments. This phenomenon echoes the effect of hand proximity in the attention literature, leading to faster orienting toward objects when they are near hands (Abrams et al., 2008; Brown et al., 2009; Reed et al., 2006). Adult studies have also shown that hand-center processing is associated with activations in the inferior motor cortex (Rizzolatti et al., 1997), and nearby-hand produces faster perceptual processing and enhances attentional engagement with visuo-tactile stimuli (Thomas & Sunny, 2017; Tseng et al., 2012). The alignment between manual actions and the enhanced saliency upon the object can reinforce the visual selection processes, preparing infants to maintain attentional focus with the target while inhibiting irrelevant distractors in the environment (Wass et al., 2018). By adapting to the patterns of visual selection shaped by parents' hand actions, children can prepare to integrate incoming visual information through self-generated exploration and accumulated knowledge as they become capable of endogenous control over their attention.

Moreover, parents' manual engagement not only scaffolds infant attention but also serves as an essential social reference for learning and communication. In the present work, we specified the episodes in infants' attended objects that were handled by parents and infants together and found that infant attention to the jointly handled object elicited a robust signal of sharing that was little affected by the strength of the visual saliency. One possible explanation is that joint manual engagement stabilizes the location of the viewed object and provides opportunities for attention sharing, thereby reinforcing the reciprocal flow of information between parents and children (Burling & Yoshida, 2019; Deák et al., 2014; Mundy & Newell, 2007). The variabilities in hand-eye coordination provide numerous opportunities to boost their interest in self-exploration and gradually acquire the social meaningfulness of actions when proceeding with their surroundings more efficiently (Capone & McGregor, 2004; Goodwyn et al., 2000; Sun et al., 2022; Yu & Smith, 2013).

The present study primarily focused on the exogenous effect of visual selection during early free-flowing viewing, but this is half of the story. Visual attention comprises two integrated systems organized by exogenous and endogenous processing together. Considering the visual resources in an individual's field are dynamic and complex on a real-time scale, it is infeasible to manipulate the exogenous and endogenous input and assess their effects on object looking separately under naturalistic circumstances. Not only do parental manual actions enrich the kinematic properties of moving objects, but the child's head and body movements aligned with the attention shifting can also contribute to the saliency of one's egocentric visual scenes.

4.1. *Limitations*

The saliency computation in the present work was estimated by the features of visual scenes rather than based on object-centered detection alone. As a result, a parent's face and other elements obscured in the background, such as the edge of the table, might also be considered in the saliency calculation, influencing the saliency estimation of objects. Additionally, the potential competition between infants' ongoing motivation and the physical properties of the objects was not addressed. According to the biased competition model (Duncan, 1996; Duncan et al., 1997), infants can be biased to maintain attention when involving multiple novel distractors. Also, it is notable that the present work focused on the interactive loop between parents' and infants' manual engagements with objects without assessing individuals' gross and fine motor skills. Thus, the degree of manual engagement during play might not reflect infants' motor capacities or their internal interest in objects. Future work will concern the role of endogenous processing in directing attention through existing knowledge and investigate the microstructure of infants' responsiveness, including gaze shifting, vocalizations, or other motor responses, to parents' object labeling across diverse social contexts. These future findings could shed more definite light on the development of sustained attention as an integration of endogenous and exogenous processing.

4.2. Conclusion

The present study highlighted that infants' naturalistic viewing is tightly coordinated with momentary perceptual experiences and corresponding motor capacities. It is not doubted that acting on objects, whether by parents and/or infants, is critical for visual selection. Here, we provide a new perspective to examine the significance of hand actions in infant object attention by assessing the momentary saliency change from infants' scenes. Infants, especially in the younger group, were exclusively sensitive to the salient features of the attended object in line with handled actions. Specifically, parent object handling, as an essential supplement of the saliency input, optimizes one's perceptual experiences during social interactions with others, supporting the initiation of self-manipulation and gradually nurturing the coordination between visual and sensorimotor development. The impact of manual engagement may account for the variabilities of individual growth in attention selection skills and has broader implications for communication and learning in various social contexts, including one-on-one interactions and homes and child care settings.

CRediT authorship contribution statement

Lichao Sun: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **David J. Francis:** Software, Supervision. **Yukie Nagai:** Resources, Software, Supervision. **Hanako Yoshida:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

All the authors have no conflict of interest or relationship, financial or otherwise.

Data availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Acknowledgement

This research was supported in part by a National Institutes of Health grant awarded to the University of Houston (R01HD058620-01 & R21 HD108746-01), and by the University of Houston Small Grant Program and Research Progress Grant. Parts of analyses and findings were presented at ICIS and SRCD Conferences. We thank the parents and children in our community who supported the research and participated in the study, as well as the research assistants in the Cognitive Development Lab for the collection and annotation of the data. We also thank Dr. Martha E. Arterberry, the Section Editor of Acta Psychologica, for the invaluable support and guidance throughout the review process.

References

Abrams, R. A., Davoli, C. C., Du, F., Knapp, W. H., III, & Paull, D. (2008). Altered vision near the hands. *Cognition, 107*(3), 1035–1047.

Açık, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Frontiers in Psychology, 1*, 207.

Adolph, K. E. (1997). *Learning in the development of infant locomotion*. University of Chicago Press.

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

Anderson, E. M., Seemiller, E. S., & Smith, L. B. (2022). Scene saliencies in egocentric vision and their creation by parents and infants. *Cognition, 229*, 105256.

Axelsson, E. L., Perry, L. K., Scott, E. J., & Horst, J. S. (2016). Near or far: The effect of spatial distance and vocabulary knowledge on word learning. *Acta Psychologica, 163*, 81–87.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development, 55*(4), 1278. https://doi.org/10.2307/1129997

Bambach, S., Crandall, D. J., & Yu, C. (2013a). Understanding embodied visual attention in child-parent interaction. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1–6).

Bambach, S., Crandall, D. J., & Yu, C. (2013b). Understanding embodied visual attention in child-parent interaction. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1–6). https://doi.org/10.1109/DevLrn.2013.6652555

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 (ICDL-EpiRob)* (pp. 132–137).

Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior, 26*, 313–331.

Basilio, M., & Rodríguez, C. (2017). How toddlers think with their hands: Social and private gestures as evidence of cognitive self-regulation in guided play with objects. *Early Child Development and Care, 187*(12), 1971–1986. https://doi.org/10.1080/03004430.2016.1202944

Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology, 33*, 680–685.

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

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science, 5*(1), 72–83.

Borji, & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 185–207.

Borjon, J. I., Abney, D. H., Yu, C., & Smith, L. B. (2021). Head and eyes: Looking behavior in 12- to 24-month-old infants. *Journal of Vision, 21*(8), 18. https://doi.org/10.1167/jov.21.8.18

Brown, L. E., Morrissey, B. F., & Goodale, M. A. (2009). Vision in the palm of your hand. *Neuropsychologia, 47*(6), 1621–1626.

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

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. https://doi.org/10.1111/cdev.13201

Burling, J. M., Yoshida, H., & Nagai, Y. (2013). The significance of social input, early motion experiences, and attentional selection. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1–2). https://doi.org/10.1109/DevLrn.2013.6652556

Capone, N. C., & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research, 47*(1), 173–186. https://doi.org/10.1044/1092-4388(2004/015)

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology, 41*(5–8), 342–361. https://doi.org/10.1080/87565641.2016.1274313

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220. https://doi.org/10.1037/h0026256

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology, 52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. In *, Vol. 34. Advances in child development and behavior* (pp. 283–322). Elsevier Science & Technology. https://doi.org/10.1016/S0065-2407(06)80010-8.

Corbetta, D., Thelen, E., & Johnson, K. (2000). Motor constraints on the development of perception-action matching in infant reaching. *Infant Behavior and Development, 23*(3–4), 351–374.

Csibra, G., Tucker, L. A., & Johnson, M. H. (2001). Differential frontal cortex activation before anticipatory and reactive saccades in infants. *Infancy, 2*(2), 159–174. https://doi.org/10.1207/S15327078IN0202_3

Curtindale, L. M., Bahrick, L. E., Lickliter, R., & Colombo, J. (2019). Effects of multimodal synchrony on infant attention and heart rate during events with social and nonsocial stimuli. *Journal of Experimental Child Psychology, 178*, 283–294. https://doi.org/10.1016/j.jecp.2018.10.006

Dannemiller, J. L. (1998). A competition model of exogenous orienting in 3.5-month-old infants. *Journal of Experimental Child Psychology, 68*(3), 169–201. https://doi.org/10.1006/jecp.1997.2426

Dannemiller, J. L. (2000). Competition in early exogenous orienting between 7 and 21 weeks. *Journal of Experimental Child Psychology, 76*(4), 253–274. https://doi.org/10.1006/jecp.1999.2551

Dannemiller, J. L. (2005). Motion popout in selective visual orienting at 4.5 but not at 2 months in human infants. *Infancy, 8*(3), 201–216. https://doi.org/10.1207/s15327078in0803_1

Datavyu Team. (2014). *Datavyu: A video coding tool*. Databrary Project, New York University. URL http://Datavyu.Org.

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy, 23*(1), 4–28. https://doi.org/10.1111/infa.12204

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science, 17*(2), 270–281. https://doi.org/10.1111/desc.12122

Duncan, J. (1996). In N. T. Inui, & J.l. McClelland (Eds.), *Cooperating brain systems in selective perception and action* (pp. 549–578). Cambrideg, MA: MIT Press.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7*(2), 255–261. https://doi.org/10.1016/S0959-4388(97)80014-1

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48*(1), 269–297.

Elmlinger, S. L., & Yu, C. (2019, August). Toddlers' hands organize parent-toddler attention across different social contexts. In *2019 ICDL: 20th IEEE International Conference on Dieletric Liquids, Roma, Italy*. https://doi.org/10.1109/DEVLRN.2019.8850682

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics, 12*(2), 201–204.

Fantz, R. L., & Miranda, S. B. (1975). Newborn infant attention to form of contour. *Child Development, 46*(1), 224–228. https://doi.org/10.2307/1128853

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

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Reply: Measuring variability in early child language: Don't shoot the messenger. *Child Development, 71*(2), 323–328. https://doi.org/10.1111/1467-8624.00147

Fenson, L., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). *Variability in early communicative development*. Society for Research in Child Development.

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science, 21*(4). https://doi.org/10.1111/desc.12626. e12626-n/a.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking: Head-mounted eye tracking. *Child Development, 82*(6), 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Franchak, Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free Viewing Gaze Behavior in Infants and Adults. *Infancy, 21*(3), 262–287.

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.

Goodwyn, S. W., Acredolo, L. P., & Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior, 24*(2), 81–103. https://doi.org/10.1023/A:1006653828895

Group, W. M. G. R. S., & de Onis, M. (2006). WHO Motor Development Study: Windows of achievement for six gross motor development milestones. *Acta Paediatrica, 95* (S450), 86–95. https://doi.org/10.1111/j.1651-2227.2006.tb02379.x

Harbourne, R. T., Lobo, M. A., Karst, G. M., & Galloway, J. C. (2013). Sit happens: Does sitting development perturb reaching development, or vice versa? *Infant Behavior & Development, 36*(3), 438–450. https://doi.org/10.1016/j.infbeh.2013.03.011

Heiman, C. M., Cole, W. G., Lee, D. K., & Adolph, K. E. (2019). Object interaction and walking: Integration of old and new skills in infant development. *Infancy, 24*(4), 547–569. https://doi.org/10.1111/infa.12289

Hietanen, J. K. (2002). Social attention orienting integrates visual information from head and body orientation. *Psychological Research, 66*(3), 174–179. https://doi.org/10.1007/s00426-002-0091-8

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., & Koch, C. (2001). Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging, 10*(1), 161–169.

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. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37*(2), 229–261. https://doi.org/10.1017/S0305000909990432

Iverson, J. M. (2022). Developing language in a developing body, revisited: The cascading effects of motor development on the acquisition of language. Wiley interdisciplinary reviews. *Cognitive Science, 13*(6). https://doi.org/10.1002/wcs.1626. e1626-n/a.

Kannass, K. N., & Oakes, L. M. (2008). The development of attention and its relations to language in infancy and toddlerhood. *Journal of Cognition and Development, 9*(2), 222–246. https://doi.org/10.1080/15248370802022696

Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior and Development, 32*(4), 437–444.

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review, 27*(3), 283–348.

Kreft, I. G., De Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*(1), 1–21.

Kretch, K. S., & Adolph, K. E. (2015). Active vision in passive locomotion: Real-world free viewing in infants and adults. *Developmental Science, 18*(5), 736–750. https://doi.org/10.1111/desc.12251

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development, 85*(4), 1503–1518. https://doi.org/10.1111/cdev.12206

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.

Madden, D. J., Whiting, W. L., Cabeza, R., & Huettel, S. A. (2004). Age-related preservation of top-down attentional guidance during visual search. *Psychology and Aging, 19*(2), 304.

Mahdi, A., Su, M., Schlesinger, M., & Qin, J. (2018). A comparison study of saliency models for fixation prediction on infants and adults. *IEEE Transactions on Cognitive and Developmental Systems, 10*(3), 485–498. https://doi.org/10.1109/TCDS.2017.2696439

Marat, S., Ho Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., & Guérin-Dugué, A. (2009). Modelling spatio-temporal saliency to predict gaze direction for short videos. *International Journal of Computer Vision, 82*(3), 231–243.

McFarland, D. H., Fortin, A. J., & Polka, L. (2020). Physiological measures of mother–infant interactional synchrony. *Developmental Psychobiology, 62*(1), 50–61. https://doi.org/10.1002/dev.21913

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Mendez, A. H., Yu, C., & Smith, L. B. (2023). Controlling the input: How one-year-old infants sustain visual attention. *Developmental Science*. , Article e13445. https://doi.org/10.1111/desc.13445

Morse, A. F., Benitez, V. L., Belpaeme, T., Cangelosi, A., & Smith, L. B. (2015). Posture affects how robots and infants map words to objects. *PLoS One, 10*(3), Article e0116012. https://doi.org/10.1371/journal.pone.0116012

Mundy, P., Card, J., & Fox, N. (2000). EEG correlates of the development of infant joint attention skills. *Developmental Psychobiology, 36*(4), 325–338. https://doi.org/10.1002/(SICI)1098-2302(200005)36:4<325::AID-DEV7>3.0.CO;2-F

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science, 16*(5), 269–274.

Nagai, Y., & Rohlfing, K. J. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development, 1*(1), 44–54. https://doi.org/10.1109/TAMD.2009.2021090

Oakes, L. M., Kannass, K. N., & Shaddy, D. J. (2002). Developmental changes in endogenous control of attention: The role of target familiarity on infants' distraction latency. *Child Development, 73*(6), 1644–1655. https://doi.org/10.1111/1467-8624.00496

Pashler, H. (1991). Shifting visual attention and selecting motor responses: Distinct attentional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 17*(4), 1023.

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 (Charlottesville, Va.), 10*(11), 22. https://doi.org/10.1167/10.11.22

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review, 21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Perone, S., Madole, K. L., Ross-Sheehy, S., Carey, M., & Oakes, L. M. (2008). The relation between Infants' activity with objects and attention to object appearance. *Developmental Psychology, 44*(5), 1242–1248. https://doi.org/10.1037/0012-1649.44.5.1242

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25. https://doi.org/10.1080/00335558008248231

R Core Team. (2022). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rachwani, J., Santamaria, V., Saavedra, S. L., & Woollacott, M. H. (2015). The development of trunk control and its relation to reaching in infancy: A longitudinal study. *Frontiers in Human Neuroscience, 8*. https://www.frontiersin.org/articles/10.3389/fnhum.2015.00094.

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: Attentional prioritization of space near the hand. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 166.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science, 277*(5323), 190–191.

Robertson, S. S., Bacher, L. F., & Huntington, N. L. (2001). The integration of body movement and attention in young infants. *Psychological Science, 12*(6), 523–526. https://doi.org/10.1111/1467-9280.00396

Robertson, S. S., & Johnson, S. L. (2009). Embodied infant attention. *Developmental Science, 12*(2), 297–304. https://doi.org/10.1111/j.1467-7687.2008.00766.x

Rochat, P. (1989). Object manipulation and exploration in 2- to 5-month-old infants. *Developmental Psychology, 25*(6), 871–884. https://doi.org/10.1037//0012-1649.25.6.871

Rochat, P., & Goubet, N. (1995). Development of sitting and reaching in 5- to 6-month-old infants. *Infant Behavior & Development, 18*(1), 53–68. https://doi.org/10.1016/0163-6383(95)90007-1

Ruff, H. A. (1996). *Attention in early development: Themes and variations*. Oxford University Press.

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS One, 6*(12), Article e28095.

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* (2104–2019).

Schlesinger, M., Amso, D., & Johnson, S. P. (2007). The neural basis for visual selective attention in young infants: A computational account. *Adaptive Behavior, 15*(2), 135–148.

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (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. https://doi.org/10.1080/15248372.2014.933430

Sosa, A. V. (2015). Association of the type of toy used during play with the quantity and quality of parent-infant communication. *JAMA Pediatrics, 170*(2), 1–6. https://doi.org/10.1001/jamapediatrics.2015.3753

Soska, K. C., & Adolph, K. E. (2014). Postural position constrains multimodal object exploration in infants. *Infancy, 19*(2), 138–161. https://doi.org/10.1111/infa.12039

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology, 46*, 129–138. https://doi.org/10.1037/a0014618

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. https://doi.org/10.1037/dev0000628

Sun, L. (2022). *Why parent object handling matters: The relations among social references, infant attention, and visual saliency* (Doctoral dissertation). University of Houston.

Sun, L., Griep, C. D., & Yoshida, H. (2022). Shared multimodal input through social coordination: Infants with monolingual and bilingual learning experiences. *Frontiers in Psychology, 13*, Article 745904. https://doi.org/10.3389/fpsyg.2022.745904

Sun, L., & Yoshida, H. (2022). Why the parent's gaze is so powerful in organizing the infant's gaze: The relationship between parental referential cues and infant object looking. *Infancy, 27*(4), 780–808. https://doi.org/10.1111/infa.12475

Tellinghuisen, D. J., Oakes, L. M., & Tjebkes, T. L. (1999). The influence of attentional state and stimulus characteristics on infant distractibility. *Cognitive Development, 14*(2), 199–213. https://doi.org/10.1016/S0885-2014(99)00002-7

Theeuwes, J. (1993). Visual selective attention: A theoretical analysis. *Acta Psychologica, 83*(2), 93–154.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20*(4), 799.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J. P., Schneider, K., & Zernicke, R. F. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. *Child Development, 64*(4), 1058–1098. https://doi.org/10.1111/j.1467-8624.1993.tb04188.x

Thomas, T., & Sunny, M. M. (2017). Slower attentional disengagement but faster perceptual processing near the hand. *Acta Psychologica, 174*, 40–47.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454. https://doi.org/10.2307/1130423

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Tseng, P., Bridgeman, B., & Juan, C. H. (2012). Take the matter into your own hands: A brief review of the effect of nearby-hands on visual processing. *Vision Research, 72*, 74–77.

U.S. Census Bureau. (2020). 2015–2019 American Community Survey (ACS) 5-year estimates. https://www.census.gov/quickfacts/fact/table/houstoncitytexas/INC110219.

Ungerer, J. A., & Sigman, M. (1984). The relation of play and sensorimotor behavior to language in the second year. *Child Development, 55*(4), 1448–1455. https://doi.org/10.2307/1130014

Wass, S. V., Noreika, V., Georgieva, S., Clackson, K., Brightman, L., Nutbrown, R., … Leong, V. (2018). Parental neural responsivity to infants' visual attention: How mature brains influence immature brains during social interaction. *PLoS Biology, 16*(12), e2006328. https://doi.org/10.1371/journal.pbio.2006328

Wentworth, N., Haith, M. M., & Hood, R. (2002). Spatiotemporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy, 3*(3), 303–321.

Yantis. (1993). Stimulus-Driven Attentional Capture and Attentional Control Settings. Journal of Experimental Psychology. *Human Perception and Performance, 19*(3), 676–681. https://doi.org/10.1037/0096-1523.19.3.676

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 25*(3), 661. https://doi.org/10.1037/0096-1523.25.3.661

Yoshida, H., Cirino, P., Mire, S. S., Burling, J. M., & Lee, S. (2020). Parents' gesture adaptations to children with autism spectrum disorder. *Journal of Child Language, 47*(1), 205–224. https://doi.org/10.1017/S0305000919000497

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy, 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. https://doi.org/10.1016/j.cognition.2012.06.016

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), Article e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2015). Linking joint attention with hand-eye coordination – A sensorimotor approach to understanding child-parent social interaction. In *CogSci ... Annual Conference of the Cognitive Science Society. Cognitive Science Society (U.S.). Conference, 2015* (pp. 2763–2768).

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology, 26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development, 88*(6), 2060–2078. https://doi.org/10.1111/cdev.12730

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science, 22*(1), Article e12735. https://doi.org/10.1111/desc.12735