RESEARCH ARTICLE

Developmental Science WILEY

Controlling the input: How one-year-old infants sustain visual attention

Andres H. Mendez[1,2] | Chen Yu[3] | Linda B. Smith[4]

[1]CICEA, Universidad de la República, Montevideo, Uruguay

[2]Institut de Neurociencies, Universitat de Barcelona, Barcelona, Spain

[3]Department of Psychology, University of Texas, Austin, Texas, USA

[4]Psychological and Brain Sciences, Indiana Unversity, Bloomington, Indiana, USA

Correspondence
Linda B. Smith, Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA.
Email: smith4@iu.edu

Abstract

Traditionally, the exogenous control of gaze by external saliencies and the endogenous control of gaze by knowledge and context have been viewed as competing systems, with late infancy seen as a period of strengthening top-down control over the vagaries of the input. Here we found that one-year-old infants control sustained attention through head movements that increase the visibility of the attended object. Freely moving one-year-old infants ($n = 45$) wore head-mounted eye trackers and head motion sensors while exploring sets of toys of the same physical size. The visual size of the objects, a well-documented salience, varied naturally with the infant's moment-to-moment posture and head movements. Sustained attention to an object was characterized by the tight control of head movements that created and then stabilized a visual size advantage for the attended object for sustained attention. The findings show collaboration between exogenous and endogenous attentional systems and suggest new hypotheses about the development of sustained visual attention.

KEYWORDS
infancy, salience, sustained attention, top-down control

1 | INTRODUCTION

After their first birthday, infants make considerable strides withstanding distraction (e.g., Rosen et al., 2019; Rothbart et al., 2011). One context used to measure attentional control is active toy play (Ruff & Lawson, 1990; Wass, Noreika, et al., 2018; Yu & Smith, 2016). Like many other everyday contexts, the visual world of toy play is busy with many potential targets. Within this context, *sustained* visual attention, the frequency of looks at a single target that lasts multiple seconds, predicts in-task learning (Yu & Smith, 2012; Yu et al., 2019). Sustained attention has also been related to later developments in attention and self-control (Brandes-Aitken et al., 2019; Fisher, 2019; Johansson et al., 2015; Ruff et al., 1990). By some accounts, infants' everyday visual attention constitutes a training ground for the prefrontal cortex and neural circuitry of executive functions (Rosen et al., 2019). For these reasons, infant attention in active, free-flowing contexts has become an important empirical target for understanding the development of self-regulatory processes (Brandes-Aitken et al., 2019; Thompson & Steinbeis, 2020).

1.1 | Two attentional systems

Visual attention is generally understood as controlled by two competing neural systems: exogenous and endogenous (Baluch & Itti, 2011; Columbo & Cheatham, 2006; Desimone & Duncan, 1995; Wass, Clackson, et al., 2018; Wolfe, 2020). The exogenous system guides gaze direction to targets based on low-level stimulus saliences such as

contrast, luminance, visual size, and motion (Itti et al., 1998; Proulx & Egeth, 2008; Wolfe & Horowitz, 2017). These attention-grabbing saliences are often discussed in terms of their evolutionary value as alerts to possible danger (Itti & Koch, 2001). In contrast, the endogenous system guides gaze direction to targets based on context and the perceiver's goals and is often driven by well-learned cues that rapidly direct attention to context-relevant information (Rossi et al., 2009; Werchan & Amso, 2020). The exogenous system is often referred to as "bottom-up" because saliences in the input itself drive gaze. The endogenous system is referred to as "top-down" because of the role of the prefrontal cortex and experience-based cues in guiding attention (Supporting information).

Much of what is known about these two systems—in infants (e.g., Rosen et al., 2019; Werchan & Amso, 2020) and in adults (e.g., Itti & Koch, 2001; Rossi et al., 2009)—derives from well-controlled experiments in which brief discrete trials pit cued targets (top-down signal) and salient distractors (bottom-up competitor) against each other. In contrast, in the free-flowing and extended temporal contexts of everyday attention, including infant object play, saliences and attentional goals are uncontrolled. They may sometimes compete but also may redundantly organize attention to the same target. Thus, in studies of infant attention during free-flowing object play, the main experimental focus has not been on the details of competition and relative saliences but on the duration of visual attention to a single object; in this context, it is this *sustained* attention, lasting multiple seconds, that predicts learning (Yu et al., 2019) and longer-term outcomes (e.g., Ruff et al., 1990; Yu et al., 2019). The predictive value of sustained attention in active vision makes sense: a long multi-second look at a single toy requires the infant to withstand any emerging distractions.

1.2 | A closed-loop control hypothesis

Active everyday visual attention also differs from typical laboratory experiments in the role of body movements (Luo & Franchak, 2020). Most contemporary experiments present stimuli as 2-dimensional arrays on a screen and constrain body movements other than eye movements (see Fisher, 2019). In everyday attention tasks, perceivers freely move in a dynamic 3-dimensional space. The geometry of the 3D world and the body mean that the perceiver's movements directly influence the image received by the retina. Small head and posture adjustments can have large effects on the image received at the retina (Ellis et al., 1989), particularly on the spatial layout of the objects in the image. Changes in vantage point can bring partially occluded objects into full view or totally obscure competing objects. Moving the head closer to a target of interest makes that target visually larger and can also block the view of a competitor object. In brief, changes in the spatial relation between the perceiver's head and objects in the world directly determine the visual sizes of potential targets for attention.

The link between body movements and the visual sizes of objects in the input image is relevant to sustained attention during object play because visual size increases the visibility of the target. Visual size is also a well-documented bottom-up salience that powerfully attracts

Research Highlights

- One-year-old infants wore head-mounted eye trackers and motion sensors while exploring multiple novel objects of the same physical size.
- Infants systematically moved their heads to make the object of current interest visually larger than competitors and stabilized the head to maintain that advantage during sustained attention.
- The onset, maintenance, and offset of sustained visual attention was time-locked to the onset, maintenance, and offset of a visual size advantage of the attended object over competitors.
- The findings implicate a close-loop control system: Infants control visual attention by behaviorally controlling the visual input.

gaze in laboratory studies of adult attention (e.g., E. M. Anderson et al., 2022; Borji et al., 2013; Wolfe & Horowitz, 2017) and appears to do so in infants as well (Cohen, 1972; Guan & Corbetta, 2012). During free-flowing toy play, uncontrolled infant body movements could create volatile variability in the visual sizes of potential targets making sustained attention more difficult. In contrast, *controlled* body movements by the infant could be used to create and control the visual sizes of targets, giving the selected target a salience advantage over competitors. In brief, *if* infants behaviorally control the spatial relation between the eyes and a selected target, they could behaviorally control the visual sizes of the target and potential competitors and, in so doing, limit distraction in the service of sustaining gaze on the selected target.

This control-the-input-to-control-attention hypothesis aligns with Ruff's (Ruff & Lawson, 1990) seminal observations on focused attention in young children: she noted that focused multi-second concentration on an object was often accompanied by a stilled body and head with the target of interest close to the head (see also, Yu & Smith, 2012). The hypothesis is also consistent with evidence showing that infants and young children shift posture and move their heads as well as eyes when they look at targets of interest (e.g., Borjon et al., 2021; Kretch et al., 2014; Luo & Franchak, 2020). Other research suggests that infants (and their parents) handle objects in ways that may also systematically increase the visual size of the object (e.g., E. M. Anderson et al., 2022; Burling & Yoshida, 2019; Smith et al., 2011; Yu & Smith, 2012). However, there is as yet no evidence that directly links sustained gaze during infant object play to the visual size of the selected target, nor is there evidence that infants behaviorally control the visual sizes of the target and distractors during sustained attention. Here, we provide this evidence.

One might ask if a system that controls attention by controlling sensory input counts as a form of top-down control. One might also ask how controlling the input to control attention relates to the development of top-down and covert control of attention, as evident in

adults (Posner, 2008). We will consider these larger theoretical issues in the general discussion in light of the empirical findings. However, we note that the proposed mechanism posits a collaboration between the endogenous and exogenous systems in the form of a closed-looped control system: a circular process through which behavior generates momentary input and behavior is continually controlled to maintain a property in visual input. Such closed-loop solutions to attentional tasks are common in many biological organisms other than humans (Biswas et al., 2018; Hoffman et al., 2013; Kleinfeld et al., 2006; Taub & Yovel, 2020) and have also been shown to be influenced by context and learning which could be interpreted as "top down" control (Bechtel & Bich, 2021; Biswas et al., 2018).

1.3 | Rationale for the empirical approach

We measured sustained attention during toy play in one-year-old infants because this is the age at which sustained attention during toy play is first systematically observed (Ruff & Capazolli, 2003). Our preliminary studies, as well studies by others, indicate that infants at this age show less sustained attention when playing alone than with a known social partner (Wass, Clackson, et al., 2018). Accordingly, the infant's caregiver was present and told to act naturally with their infant during the four test trials. Our preliminary studies also indicated that infants sustained engagement with toys was most frequent when toys were first introduced. Accordingly, we used very brief toy-play trials, with each trial 45 s in duration, the same play period used in prior studies in which infants explored toys one at a time with no competing toys (Pereira et al., 2010). Here we added competitors (3 toys for play) on each trial.

Infants wore a hat outfitted with three head-mounted sensors: a scene camera that captured the image projected from the world to the head, an eye camera that tracked gaze within the head-camera scene, and a motion sensor that captured the movement of the head relative to the world. The toys used for the play trials were all of comparable physical volume. Thus, differences in the visual size of the objects in the scene-camera images were determined solely by the vantage point and proximity of the infant's head to the toys in play. The principal analyses focus on the visual sizes of the toys in the infant point-of-view (POV) image. There are many other bottom-up salience that attract the human gaze to a target (Itti et al., 1998; Wolfe & Horowitz, 2017); however, many of these (e.g., contrast) are not directly controllable by the perceiver through momentary body movements (see, E. M. Anderson et al., 2022). Visual size, directly affected by the momentary spatial relation of the head to the object, provides a tractable and known bottom-up salience. Therefore, we analyzed the visual sizes of objects during sustained attention to provide a first test of a closed-loop control solution to attention in a busy and changing visual world.

The main empirical hypothesis tested is this: If infants behaviorally control the visual size of a selected target to sustain attention, then sustained gaze should co-occur with a behaviorally maintained visual size advantage for the target over competitors. Accordingly, the main analyses focus on onset, duration and offset of increases in visual size as a function of onset, duration, and offset of gaze during sustained attention. The additional critical question is how control of the input during sustained attention is accomplished. Because the spatial relation between the eye and the object is the world determines the visual size of the object in the infant view, the most direct path to creating and then maintaining a visual size advantage for a selected target is through head movements and head stabilization. Accordingly, we also measured the velocity of head movements with respect to the onset, duration, and offset of sustained gaze and the visual size of the target.

Parents often scaffold infant attention to objects, and the degree to which they do so predicts the amount of infant sustained attention in the task (e.g., Suarez-Rivera et al., 2019; Sun & Yoshida, 2022) as well as longer-term outcomes in executive control and self-regulation (see Marciszko et al., 2020; Rosen et al., 2019). We do not directly measure this scaffolding during our brief 45-s play trials. However, we report analyses of parent as well as infant handling of objects in relation to infant sustained attention. In the general discussion, we consider the implications of the present findings for how parent scaffolding may relate to infant control of visual attention through infant control of the input.

2 | METHODS

2.1 | Experimental procedures

Participants. 45 infants (25 male, mean age 14.1 months, SD = 1.4) met the inclusion criteria of contributing head-mounted eye tracking data for all trials. 15 recruited infants contributed no data due to failure to contribute eye-tracking data for all four trials. 30 of 45 infants contributed both eye-tracking and head motion data. Failures to contribute head-mounted eye-tracking data for all 4 trials were due to infant fussiness, removal of the headgear by the infant during the trials, refusal to wear the headgear from the beginning, or equipment failure. Failure to contribute motion data (while contributing eye-tracking data) was based on the experimenter's decision during data collection. Before starting the experiment, the experimenter needed to place the headgear and then adjust the eye camera, head camera, and motion sensor, typically in that temporal order. Multiple adjustments of the headgear invited infant hand actions directed to the headgear in the moment or caused later removal of the gear by the infant. Accordingly, the goal was to adjust the motion sensors in one or two attempts. If the proper adjustment of the motion sensor was not easily accomplished, the experiment proceeded with the collection of eye-tracking data but not head motion data. The infants were recruited from an opt-in database for developmental research. Participants in the database broadly represent the demographics of Monroe County Indiana but over-sample minority groups (68% European American, 10% African American, 7% Asian American, 7% Latino, 7% Mixed race) and consist predominantly of working- and middle-class families. Recruitment and procedures were approved by the Institutional Review Board of Indiana University (protocol number 0808000094).
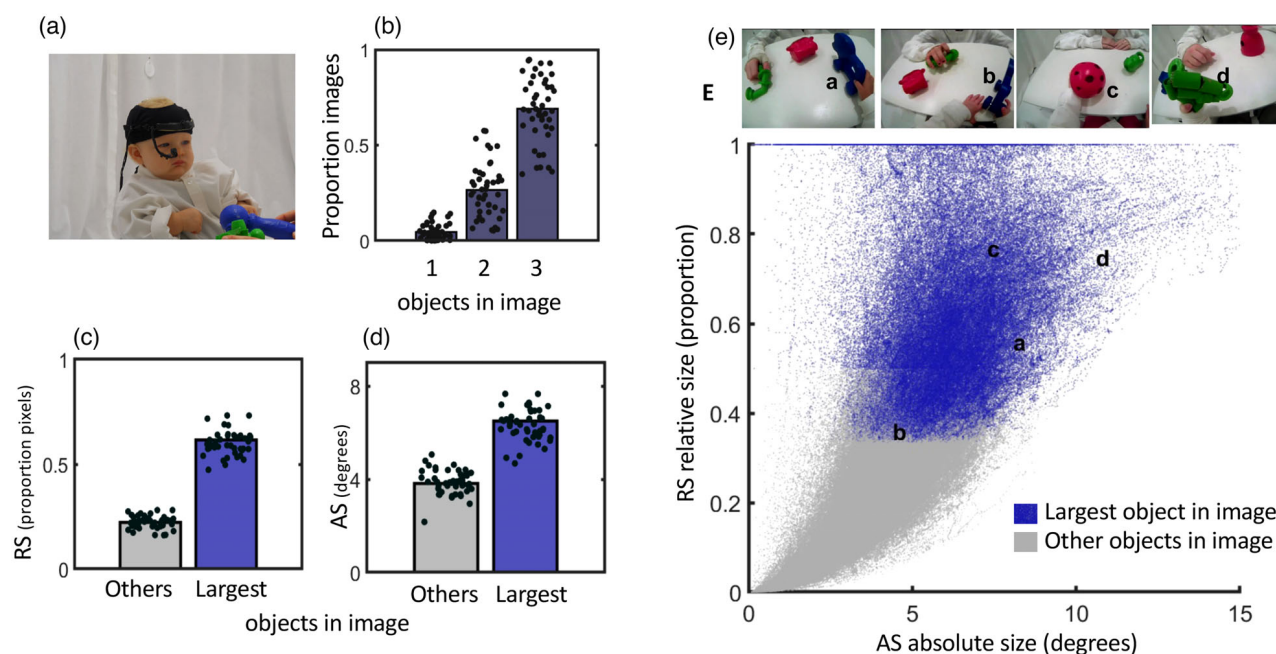
FIGURE 1 The visual size of objects in the infant view. (a) Infant in head-mounted eye-tracker. (b) Proportion of all frames in the corpus with 1, 2 or 3 of the toys present (at least 1 pixel) in the image. Dots show the proportion of images with 1, 2 or 3 toys present for individual infants. (c) Corpus mean relative visual size (RS) of the largest toy and the two other toys in the image; dots show individual participant means. (d) Corpus mean of absolute visual size (AS) for the largest and two other toys; dots show individual participant means. (e) The relative (RS) and absolute (AS) visual size of all toys present (420,325) in the corpus of (on-task) frames. The largest object in 4 example head-camera images is labeled near its RS by AS coordinates in the scatterplot.

Recording Equipment. The head-mounted eye tracker (Positive Science LLC, http://www.positivescience.com) included an infrared camera mounted on the head, pointed to the right eye of the infant, that recorded eye images, and a scene camera that captured the events in the world from the infant's point of view (Figure 1a). The scene camera's visual field was 108° diagonal. Together the eye and scene cameras yielded gaze within the scene location (x and y) at a sampling rate of 30 Hz. A wired motion capture sensor (Polhemus Liberty) was affixed to the hat on the right temple of the infant's head and collected rotational position data (roll, pitch, and yaw) at 60 Hz. There was also an overhead birds-eye camera that provided good views of hand actions by the infant and parent.

Play context and toys. Infants and parents sat across from each other at a small table (61 cm × 91 cm × 64 cm). The infant sat alone on a chair that allowed leaning and rotation of the body and torso. The infant (and parent) wore a white smock, and the walls and table were painted white. This setup served two purposes: first, the all-white background dampened in-room distractors beyond the toys themselves; second, the white background aided the algorithmic measures used to determine the presence and visual sizes of the toys in the infant head camera images.

Each infant played with three toys on each of the trials. The three toys for each trial were a different uniform color: blue, green, or red. The toys were novel, designed by the experimenters, and constructed of various materials and moveable parts to engage 1-year-old infants. The volume of each toy was 300 cm³. Different infants played with different sets of 6 toys. In total, across

the 45 infants, 18 different toys were used. See Figure 1e, for examples.

Procedure. The headgear was placed on the infant and adjusted while an experimenter and parent kept the infant occupied with a toy with a spinning light. This attention-grabbing toy was selected to calibrate the eye-tracking system and was not used in the experiment. On average, fifteen calibration points at different locations on the play table were collected; the experimenter directed the infant's attention toward the toy while a technician recorded the attended moment used in later off-line eye-tracking calibration.

Parents were told that the goal of the experiment was to study how infants explored novel objects and to interact with their infant as they normally would. Each infant received two unique sets of 3 toys twice in the four trials in the order of ABAB, with the individual sets of 3 toys for each infant randomly assigned as set A or B. The inter-trial onset was about 1.5 min apart; this made for smoother transitions in play from one trial to the next. Onset was defined as the timepoint at which all three objects for that trial were on the play table. Only the first 45 sec after onset was considered the trial proper and included as data.

2.2 | Data processing and coding

Objects in the image. The presence and image size of the objects in the infant head camera image were algorithmically determined using computer vision techniques (Smith et al., 2011; Yu et al., 2009). The algorithm delivered the number of pixels present for each toy object

in the image. The validity of the automatic coding results was assessed by asking two human coders to annotate a small proportion of the data (~ 1200 frames); the comparison of hand coding with the image processing results yielded 91% frame-by-frame agreement across all measures.

Measures of visual size. Because the objects were the same physical size, the visual size of the objects in the head-camera images depended on occlusions and distance to the head camera. The *absolute visual size* (AS) for each object present in a frame was calculated as the visual angle, that is, as an estimation of the angle the in-view objects subtends at the eye. The diameter of the fovea (adult) is typically estimated at 5.2 degrees (Jonas et al., 2015). The head camera videos are 480 pixels in height by 640 pixels in width, which at 72 dpi, translates to 22 cm x 16.9 cm, respectively. The average distance of the eye to the table center for infants sitting on the chair was 44.5 cm. The degrees in a single pixel of the head camera were therefore calculated as:

$$\theta_{pix} = \frac{\arctan\left(0.5 \times 16.9/44.5\right)}{0.5 \times 480}$$

An object's relative visual size (RS) was calculated directly from the number of image pixels belonging to each object in each frame as the proportion of all object pixels in the frame that belonged to that object. For both AS and RS, an in-play object with no pixels in the head camera image (was out of view) was entered as 0 in size. Because RS measures the proportion of all pixels belonging to one object divided by the sum of all pixels of the 3 objects in play, the expected value of RS, if all three objects were in an image and the same visual size, is 0.33. If the infant's vantage point is such that only one object is in the image, the RS is 1.00 (since 100% of all object pixels in the image belong to that object). Thus, for each toy in play on a given trial, the RS could vary from 0 (the toy has no pixels in the image) to 1.00 (the toy is the only in-view toy in that head-camera frame). The AS of an object could vary from 0 to in principle 90º if the infant positioned the object and head such the object was so close, its pixels filled the field of view of the camera. However, in the collected data the largest visual size of a single object was 28º.

Gaze data. Frame-by-frame gaze was determined for each image with respect to three regions-of-interest (ROIs) defined precisely as the pixels belonging to each of the three objects in play. Trained coders, naive to the hypotheses and goals of this study, indicated when the gaze crosshair fell on a pixel belonging to a toy object. Because the three toys were three different primary colors and differed from skin tones and the white background, this could be done with accuracy. Reliability was computed between two independent coders on eleven dyads that were randomly selected. Coders coded 25% of each dyad's frames making judgments on 2,790 frames per dyad on average. The inter-coder reliability of eye-gaze coding performed by these highly trained coders ranged from 82% to 95% for each individual subject with an overall Cohen's kappa of 0.75.

Looks. A look to an object was defined as a continuous stream of frames in which gaze was directed to pixels belonging to the same object. For look duration and stability measures, frames with gaze directed to the same object were combined into the same look if separated by no more than one frame (.033 sec).

Head motion. We measured the velocity of head movements from the 3 rotational coordinates delivered (in degrees) by the motion sensor (Borjon et al., 2021; Richards & Hunter, 1997; Rosander & Von Hofsten, 2000). Rotational velocity was calculated by taking the difference in the 3 rotational coordinates at each 60 Hz sample divided by the change in time. The key experimental question with respect to the control-loop hypothesis is how much head movement (and not specific directions of movement) because all varieties of directional movements may alter proximity, vantage point, and occlusions depending on the 3-dimensional composition of the world scene. Further, the motion sensor could not be quickly attached and aligned to match XYZ coordinates of the world. Accordingly, we used an aggregated measure of momentary rotational velocity, the calculated Euclidean distance across the 3 directional velocities. Momentary rotational velocities at or above the 99th percentile for each subject were excluded.

Infant and parent handling of objects. Because infant and parent handling of objects could systematically affect the visual sizes of the objects in play, human coders determined, frame-by-frame if a hand, and whose hand, was in contact with a toy. Coders had simultaneous access to both the infant perspective view and the view from the overhead camera but primarily relied on the overhead camera for coding hands in contact with an object. A second coder independently coded a randomly selected 25% of the frames of five dyads; agreement ranged from 0.76 to 0.90 for individual dyads, with an overall Cohen's kappa of 0.90.

Images analyzed. Each infant contributed 5400 frames (45 s* 4 trials* 30 Hz), a total corpus of 243,000 frames.

Statistical analyses. Analyses were conducted at the corpus level. The dependent measures were frame-by-frame measures of visual size of all objects in the images, gaze directed to those objects, durations of gaze to the same object, and head velocity. From these data, we determined the onsets, durations, and offsets of increases in the visual size of objects, of continuous gaze to a single object, and head velocity in relation to gaze onset, sustained gaze and gaze offset.

Mixed-effect models were conducted using the lme4 package in R (Version 3.6.1; Bates et al., 2014) with measures of object size as the fixed effect predicting the duration of looks. The nlme package in R (Version 3.1-160, Pinheiro et al., 2017) was used to study temporal changes in head speed. Corpus level error bars were calculated as per Campbell (2017). Individual infants and the specific toy objects were random variables in all analyses.

Open Data. The de-identified data for all reported analyses, the individual comparisons in the time-series analyses, and the supplementary analyses of looks below our threshold criterion for sustained attention and are available at https://osf.io/tcd35/.

3 | RESULTS

The input. For the first set of analyses, we analyzed frames (68% of the total frames) in which the infant gaze was directed to one of the toys
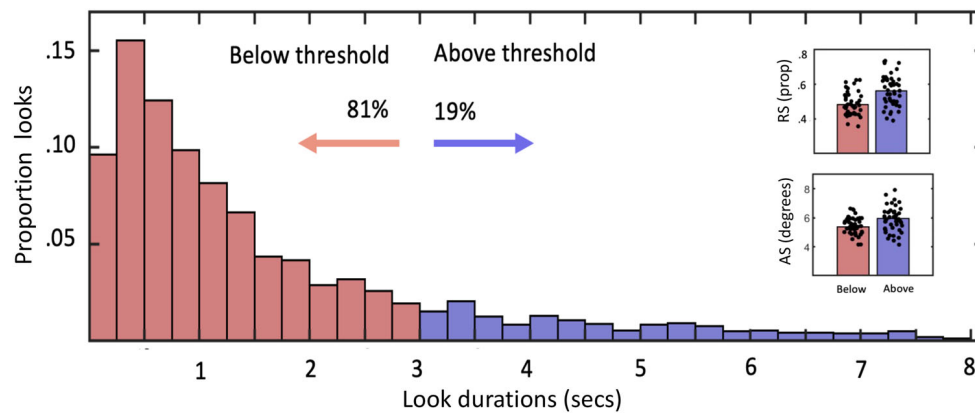
FIGURE 2 Sustained attention. (a) Histogram (frequency normalized as proportion of all looks to an object) of durations of unbroken looks to a single toy object. Bin size is 250 msec. The insets show the corpus means of RS and AS for Looks below and above for the 3-s threshold for sustained attention; dots show participant means. Although looks maintained 3 s or longer to a single object are only 19% of all individual looks, they constitute 66% of all the total time that infants direct gaze at a play object.

(excluding gaze to parent face and to targets in the room that were not toys). The goal of this first analysis is to provide a description of the stimulus, that is of the visual sizes of potential targets in the images in front of infant faces. Figure 1b,c and d show summary statistics of the corpus means (bars), and participant means (dots) of the proportion of frames with 1, 2 or 3 objects in the image (Figure 1b), RS and AS measures of the visual size of objects with at least one pixel in the image (Figure 1c,d). Figure 1e shows the main result: the RS and AS for each object present (at least one pixel) in an image with the blue dots indicating the objects that were the visually largest in each image and the grey dots indicating the remaining objects. As is apparent, the visual sizes of potential targets vary markedly. The AS of objects varied from under 1º to 28º; in 99% of cases, the AS was equal to or less than e than 15º, the maximum value shown in Figure 1e. The RS of the individual objects varied from 0.00 (the object was not in the image) to 1.00 (the object was the only one in the image). The visually largest object in a frame ranged from an RS of 0.35 (slightly larger than the expected size if all three objects were present and equal in visual size) to 1.00. On average, the largest object in an image had an RS that accounted for 0.63 of all object pixels (SE = 0.17) putting it at just less than twice the visual size of the *sum* of the pixels of the other two objects. The visually largest object in an image had an average absolute visual size of 4.8º (SE = 1.8). Figure 1b,e use letter labels to align the RS and AS coordinates of 4 individual objects that were the largest in 4 different infant-egocentric images. AS and RS are both based on pixel size and thus necessarily correlated ($r^2 = 0.51$) However, as is apparent in Figure 1e, objects with the same AS can differ markedly in their relative size (RS) to the other objects in the image. In sum, in the uncontrolled but natural context of exploring toys, the stimulus arrays in front of infant faces present highly variable within-image visual-size saliences and thus potential bottom-up competition among potential targets and distractors.

Defining sustained attention. Figure 2 shows the frequency distribution of all infant look durations with their well-known skew (Borjon et al., 2021; Suarez-Rivera et al., 2019). Most looks are very brief,

glances that last a fraction of a second. But there is a long tail of much longer multi-second looks to a single object. In the following analyses, we concentrate on the long tail of very long looks to a single object that last 3 s or longer. Although these looks longer than 3 s constitute only 19% of individual looks, they constitute 66% of the total duration of time that the toddlers were looking at an object. That is, the clear majority of looking time to objects occurs within looks that are 3 s or longer. We focus on these very long multi-second looks at a single object because they are characteristic of infant-looking behavior during object exploration because these sustained looks predict learning in the task (Pereira et al., 2014; Schroer & Yu, 2022), and because they predict longer-term developmental outcomes in visual attention, self-regulation, and executive function (e.g., Brandes-Aiken et al., 2019; Frick et al., 2018, 2019; Ruff et al., 1990). It has been suggested that briefer and sustained looks result from different underlying processes (Richards, 2010); however, there is limited evidence on this issue and no established empirical criteria for determining different kinds of looks or where the breakpoint between different kinds of looks might lie. Therefore, we took a conservative but established approach (Suarez-Rivera et al., 2019; Wass, Noreika, et al., 2018; Yu et al., 2019), using a 3-s threshold.

On face value, sustained looking at a single object for a duration longer than 3 s should be difficult if there are other competing saliences. The insets in Figure 2 show preliminary evidence that salience advantages may be associated with sustained looks, advantages that could make attention to a target easier to sustain. The means for both RS (proportion of all object pixels in the image belonging to the gazed-to-object) and AS (absolute size in degrees of gazed-to-object) are greater for looks longer than the 3 s threshold than for looks below that threshold: RS for above threshold long looks, $M_{long} = 0.56$; $SE_{long} = 0.009$, for below threshold shorter looks, $M_{short} = 0.48$; $SE_{short} = 0.005$; $\beta = 1.59$; $z = 6.39$; $p < 0.001$; AS for above threshold long looks, $M_{long} = 5.96$; $SE_{long} = 0.075$, below threshold shorter looks, $M_{short} = 5.36$; $SE_{short} = 0.037$; $\beta = 0.14$; $z = 4.90$; $p < 0.001$. However, the critical questions for the control hypothesis are about
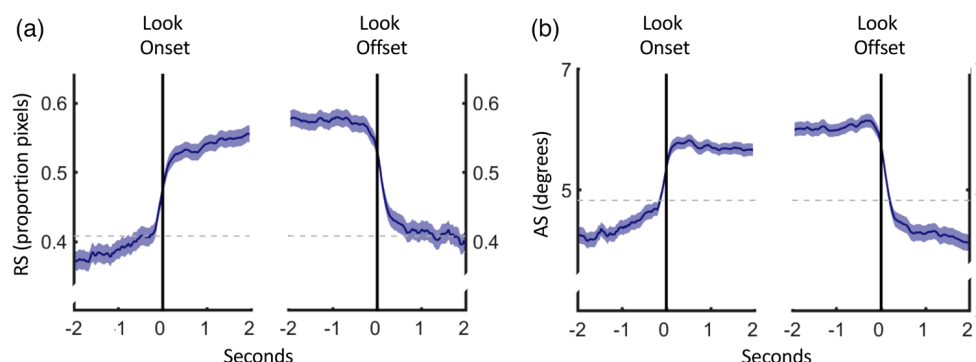
FIGURE 3 Onset and offset of visual size advantages relative to the onset and offset of sustained attention. (a) Mean corpus (calculated frame by frame) relative size (RS –proportion of object pixels in a frame corresponding to the attended object) from 2 s before and after the onset and offset of all looks 3 s or longer. (b) Mean corpus absolute size (AS—degrees subtended) for the attended object from 2 s before and after the offset and onset of a looks 3 s or longer. In both a and b, the dashed line indicates the expected baseline if looking is independent of the RS or AS of the attended object. The shaded area indicates the standard error of the corpus mean.

the dynamics of looking behavior. Do infants behaviorally create and control (that is, maintain) the observed salience advantage associated with sustained gaze to an object? We answer this question through a series of analyses of all looks to a single object that lasted 3 s or longer. We make no claim that there are categorical differences between looks that fall above and below the selected threshold. We provide supplementary analyses of the dynamics of visual size for looks lasting less than 3 s at https://osf.io/tcd35/. These shorter looks are a mixture of very different durations –from fractions of seconds to minutes. Our preliminary analyses suggest that they may also consist of different kinds of looks. Determining the criteria for separating different kinds of looks measured via head-mounted eye-tracking in freely moving perceivers remains an open and critical question for the field..

The dynamics of visual size and sustained attention. We computed the RS and AS of the attended object from 2 s before and 2 s after the onset and offset of each sustained look and compared those values to baselines for RS and AS respectively. The baselines were computed to instantiate the null hypothesis that looking at each moment is random with respect to the visual sizes of the objects in the image projected to the eye. Specifically, the moment-to-moment baseline RS and AS were determined individually for *each* sustained look by randomly selecting one in-view object and its visual size from the ordered series of within-look frames. The expected baseline (if object visual size does not matter) was then calculated frame by frame across all looks as an expected mean visual size from 2 s before to 2 s after onset and also offset. Figure 3a,b show the marked increases in RS and AS respectively at the onset of a sustained look and marked decreases at the offset.

We used the time-series approach of a series of t-tests first introduced by Allopenna et al. (1998; see also Yu & Smith, 2012) to answer the statistical question about onset as to *when* in a series of comparisons, a measure *first* differs from the baseline *and remains above the baseline* for a pre-specified duration. Although the Type 1 error for each comparison in the series is set at 0.05, the overall Type 1 error rate for a rise that exists for N comparisons after the first rise is much smaller

(approximated by 0.05^{N}). This approach can also answer the statistical question, critical with respect to the offset of a look, as to when in a series of t-tests, a measure *that differs from baseline* for N comparisons, first returns to and *remains* at baseline for a pre-specified duration. We computed, across the corpus of all sustained looks, frame-by-frame pairwise comparisons (30 per sec), comparing RS to its baseline and AS to its baseline from 2 s before the onset to 2 s after the onset and for the window 2 s before to 2 s after the offset with individual looks as the random variable. The RS of the attended object, Figure 3a, first reliably increased from baseline and stayed above baseline at 100 msec before look onset. Thus, the increase in RS for the attended object was nearly in synchronous with the onset of the look itself. The advantage remained reliably different for 2 s ($p < 0.00001$ for the overall measure of onset of increased RS). The observed RS of the target object first reliably decreased to baseline and then remained at baseline from 200 msecs after offset ($p < 0.0001$ for the overall measure). Thus, onset and offset of the relative size advantage for selected target were temporally coordinated with the onset and offset of sustained attention to the target.

AS, the absolute size of the target, provides direct evidence of the spatial relation of the attended object to the infant's head. Using the same statistical analysis approach as for RS, the AS of the attended object first reliably differed from baseline at 66 msecs before look onset and did not return to base for the following 2 s of comparisons; the AS of the target first decreased and then remained at baseline at 166 msecs after offset, $p < 0.0001$ for both overall measures of onset and offset. In sum, the target's absolute visual size (AS) increases just ahead of the onset of the look and remains stable until just after the offset of the look. This fact indicates that the distance and vantage point of the attended object relative to the infant's head changed at the start of sustained attention and was maintained throughout the duration of the look.

A visual size advantage (as well as other saliences) could cause infant selection of a target for sustained attention. However, the very small lead in the onset of the increase in visual size with respect to the onset of gaze makes it unlikely that a visual size advantage for
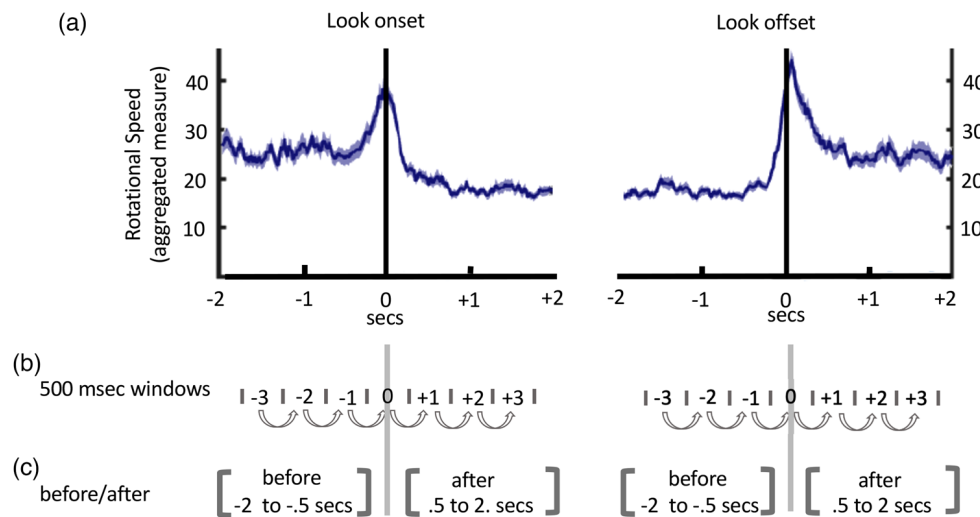
FIGURE 4 Head movements at the onset, offset and during sustained attention. (a) Mean corpus aggregated rotational velocity computed frame by frame from 2 s before and 2 s after the onset and offset of sustained attention (look 3 s or longer in duration). The shaded area indicates the standard error of the corpus mean. (b) The seven 500 msec windows centered on the onset and offset of the looks used for analyses of the precision of head movements at onset and offset. (c) The temporal windows before and after look onset and look offset used for analyses of head velocity within sustained attention as compared to the period before onset and the period after look offset.

the to-be-attended object was a systematic cause of the initial shift of gaze to the object. If the onset of the visual size advantage of the target were considered the stimulus onset, the stimulus onset to gaze onset (100 msec) would fall well below the estimated minimum (250 to 500 msec depending on the complexity of the array) for infants to execute a shift of gaze to a target (e.g., Marchman & Fernald, 2008; Oakes & Luck, 2013; Yu et al., 2012) as determined in controlled experimental studies. Instead, the brief lead of the visual size increase relative to the gaze shifts likely reflects the well-documented fact that infants and children (unlike adults) often shift the head before their eyes when re-directing gaze (e.g., Borjon et al., 2021; Luo & Franchak, 2020; Nakagawa & Sukigara, 2013; Schmitow et al., 2013).

Head movements. Head movements provide a direct path for creating and controlling a visual size advantage for targets of interest. If infants increase the visual size of an object at the start of sustained attention and then maintain that advantage for the duration of the look by controlling the spatial relation of their head to the object, there should be a rapid head movement near the start of gaze onset and then a stilling of the head within the look. Figure 4a shows the mean momentary aggregate velocity for the 30 infants who contributed head motion data for the period from 2 s before to 2 s after look onset and for the period from 2 s before to 2 s after look offset. As is apparent, infant heads move markedly around the onset and offset of a look. To statistically quantify the reliability of this increase in head movements at onset and offset of sustained attention, we calculated the mean rotational velocity of the head in 500 msec windows (a window size around gaze onset that corresponds to previous reported temporal relations between coordinated head movements and gaze shifts by toddlers, Borjon et al., 2021; Luo & Franchak, 2020; Nakagawa & Sukigara, 2013; Schmitow et al., 2013). As shown in Figure 4b, one 500 msec

window was centered on look onset. This creates 7 windows from 1.75 s before [window numbers: −3, −2, −1] to after [window numbers: +1, +2,+3] onset [0] and likewise 7 windows from 1.75 s to 1.75 after the offset of a Long look. We compared head velocity in each window to the successive next window with a preset alpha of 0.001 to correct for the individual 6 comparisons. For onset, the only reliable comparisons were from the window −1 to 0, and from 0 to +1: that is, head velocity during the window centered on Look onset differed from the just preceding window ($p < 0.0001$) and from the just following window ($p < 0.0001$). No other comparisons in the series of 6 comparisons were reliable. The head moves at the start of a look and then stills. For offset, head velocity at the window centered at offset increased reliably from the just preceding window (window −1 to 0, $p < 0.0001$) and the just following window (0 to +1, $p < 0.0001$). No other adjacent comparisons approached significance. The head moves rapidly at the onset and offset of the look to a target and is stilled during the look.

Is the head stilled more during the look than prior to the look? If infants inhibit head movements *during* sustained attention to maintain the visual size advantage, the head should be more stilled during the look than in the period before the head movement at the start of a look. To test the hypothesis, illustrated in Figure 4c, we compared head movements for the 2 s before onset and for 2 after onset during the look excluding the 500 msec window of rapid head movements at onset. We did the same for the 2 s during look just before offset and for the 2 s after look onset, again excluding the 500 msec window around look offset. The key question is whether head movements during the look are less than those in the period outside of the sustained look. Head movement in the 2 s window before onset M = 26.25, SE = 1.11 are faster than the head movements in the 2 s after onset window, that is, *during* the look, M = 18.72, SE = 0.822, B = 7.64, t

(a) Infant Handling of Looked to Object

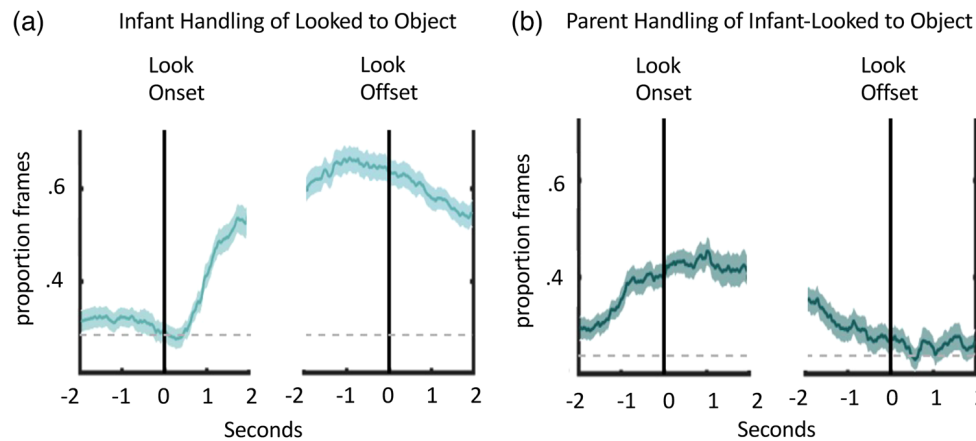(b) Parent Handling of Infant-Looked to Object

FIGURE 5 Handling of objects at onset and offset of sustained attention. (a) Mean handling by infant (determined frame by frame) of the attended object for 2 s before and after the onset and 2 s before and after the offset looks 3 s or longer. The dashed line indicates the expected baseline if looking is independent of infant handling. (b) Mean handling by parent (determined frame by frame) of the attended object for 2 s before and after the onset and 2 s before and after the offset of looks longer 3 s or longer. The dashed line indicates the expected baseline if looking is independent of infant handling. For handling, the shaded area indicates the standard error of the mean proportion of participants handling of the looked-to object for each frame.

$(145) = 7.54$, $p < 0001$. Likewise, head movements in the 2 s before offset, that is during the look, M = 17.47, SE = 0.62, were slower than in the 2 s after, M = 25.29, SE = 0.97, B = 7.87, t(146) = 5.75, $p < 0.001$. After one-year-old infants move their head to direct gaze to an object for a Long look, they still the head during the look to a greater degree than before or after the head movements that initiated and ended the look. This result implicates control of head movements in the service of sustaining attention.

Object handling. Infants (and their parents) explored the objects by handling them. These activities could create and sustain visual size advantages if they systematically brought handled objects close to the infant head or if handling was the systematic cause of the infant's leaning in for a closer look at the start of sustained attention. We determined for the 2-s window before and after the onset and offset of each sustained look, the frames in which the looked-to object was in contact with the hands of the parent or infant. We calculated the expected baselines of infant and parent object handling by randomly selecting one object from the corpus of frames. The baseline thus instantiates the null hypothesis that object handling was independent of the object and infant sustained attention. We used the same time series analysis (Allopenna et al., 1998) to determine changes in visual size at the onset and offset of sustained attention. As shown in Figure 5, the dynamics of hand contact by both infants and caregiver are not systematically related to the onset and offset of infant sustained gaze to an object. The first significant increase in infant handling of the attended object occurred at 733 msecs *after* the onset of sustained gaze. Infant handling did not decrease to baseline within the +/- 2 s window around look offset. Thus, infant handling of an attended object was common for some portion of the time that gaze is directed to the object but was not systematically associated with the onset and offset of the look and thus not systematically associated with the onset and offset of the visual size advantages nor the head movements at the onset and offset of sustained looks to an object. Onset and offset of parent handling

were also not coordinated with the onset and offset of the infant's look to the object. The likelihood of parent handling of an object to which an infant sustained gaze was reliably above baseline for the entire +/- 2-s window around look onset but returned to baseline and reliably stayed at baseline from 766 msecs forward. In brief, although handling and moving objects occurred before, during, and after sustained attention to the handled object, handling is not a systematic proximal cause of the increased visual size at look onset, the maintenance of that salience advantage, the breaking of the visual size advantage that occurs at look offset, nor the head movements that occur at the start and end of a sustained gaze to an object.

4 | DISCUSSION

Moving heads and moving objects during infant toy play create a context of highly variable input with respect to the visual sizes of potential targets for attention. However, during a sustained look to an object, the visual size of the target is tightly controlled by infant behavior. Infants control the visual size of the attended object in the input during sustained attention by controlling their head movements and thus the spatial relation between the head and the object. Infants control the input by systematically moving the head at the start of sustained attention in a way that makes the target visually larger than competitors, they inhibit head movements during the look and thus maintain the salience advantage, and they break the look and the salience advantage by moving the head. This solution to sustained attention has not been explicitly considered in the developmental literature. But as Gibson (1979) argued, the behavior of freely moving perceivers creates the visual input, defines the perceptual task that must be solved, and provides the solution. Perception in the context of freely moving perceivers is also the environment of evolutionarily constrained developmental process. In the following discussion, we consider the

implications of infants' behavioral control of the input for developing top-down control of attention.

4.1 | Collaboration

Is infant closed-loop behavioral control of sustained attention a form of bottom-up or top-down control? Head movements that increase the visual size of the target occur at the same time as the infant directs gaze to the object. Thus, whatever factors determine the shift in gaze to what will become the target of sustained attention also determine the synchronous head movements that change the spatial relation of the head to the target making the target larger than competitors in the infant field of view. We do not know the factors that cause the infants to shift gaze to the target. Those factors could include other bottom-up saliences, memories of past experiences, or other momentary goals of the infant. Nonetheless, the head movements that create and sustain a visual size advantage for targets of sustained attention emerge at the same moment as the decision to look. These observations suggest that there may not be an easy partition of infant attention into distinct top-down versus bottom-up processes. In a theoretical paper on infant visual attention, Rosen et al. (2019) argued that in everyday attention, exogenous and endogenous systems were collaborative and not separable as moment-to-moment influences on attention: bottom-up saliences alert, initiate interest and create top-down attentional goals. Here we show bottom-up salience is also behaviorally controlled by the infant when the infant displays sustained visual interested in an object.

The closed-loop hypothesis and the present findings provide new insights into several previous results. Ruff's seminal work (Ruff & Lawson, 1990) used not only the duration of looks as defining of focused attention but also a stilled head and a stilled body. The present results suggest that controlled body movements may be critical because they stabilize a *visual* advantage of the target over distractors. Two studies (Pereira et al., 2014; Yu & Smith, 2012) of infant point-of view experiences have revealed a visual signature of effective parent naming moments for learning object names. Pereira et al. (2014) found that 18-month-old infants were more likely to learn name-object associations when the parent naming event co-occurred with a visual experience in which the referent was visually large and centered in the infant head-centered field of view for an extended duration of 3–5 s around the heard name, a duration that meets the literature's currently used threshold for sustained attention. Yu and Smith (2012) showed that infants were more likely to learn object names when their head was not moving during parent naming events. In brief, the signature sensory-motor properties for infant learning of object names are also signature sensory-motor properties of sustained attention (see Yu et al., 2019).

The present findings may also increase understanding of the developmental paths from "sticky" gaze in infants younger than 4 months of age (Columbo & Cheatham, 2006; Kulke et al., 2015) to sustained attention as observed in infants 9 months of age and older. On the surface, the two phenomena appear similar: long gaze to a continuing target that is not easily disrupted by a competitor. However, in young infants, re-orienting gaze to a new peripheral target is a positive predictor of later cognitive developments (e.g., Columbo, 1997; Columbo & Cheatham, 2006); in infants older than 9 months of age, sustaining gaze on a target is a positive predictor of later cognitive outcomes (e.g., Fisher, 2019; Rosen et al., 2019; see also Geeraerts et al., 2019). When young infants successfully "unstick" gaze to re-orient to a new target, they often move their heads before the gaze shift and these movements have been proposed (Robertson et al., 2001) to unlock the gaze on the prior target. The one-year-old infants in the present study also moved their heads to end sustained gaze in a way that may function to disrupt the helpful salience for sustaining gaze in order to look elsewhere. Together, these observations suggest that the increasing autonomy and control of motor systems in the first two years of life may play a critical role in the development of cognitive control systems. This hypothesis (see also Gottwald et al., 2016; Thompson & Steinbeis, 2020) is consistent with the use of atypical sensory-motor behaviors (including difficulty in head-stabilization) as biomarkers of attentional deficits in children (Berger et al., 2019; Friedman et al., 2005). The hypothesis is also consistent with research on the neural underpinnings of attention that reveals overlaps between brain networks that plan motor behaviors and those that control the spatial direction of attention (e.g., Miller & Cohen, 2001; Van Ede et al., 2019).

4.2 | A developmental hypothesis

Figure 6 illustrates a closed-loop system connecting visual input, brain activation, and behavior (see Byrge et al., 2014, Chiel & Beer, 1997). The input at each moment perturbs ongoing brain activations across multiple networks bottom-up and top-down, activating memories, goals, and behaviors. These activations within the infant brain can affect the world—the next moment of input—through the infant's own behavior. Momentary inputs and the brain activations elicit, incrementally support, tune, and train neural circuitry (Byrge et al., 2014) and likely do so for those relevant to the development of self-regulatory processes (Rosen et al., 2019). By adulthood, internal mechanisms of top-down attentional control are sufficiently strong that attentional tasks can be solved covertly involving no supporting external behaviors, not even shifts in eye gaze (Posner, 1980). However, for adults and infants in most everyday life tasks, attention may typically exploit the closed loop of brain-behavior-input (Foulsham et al., 2011). Considerable research shows that adults also look with their whole bodies and purposely alter the sensory input to support attention and the extraction of task relevant information (e.g., E. M. Anderson et al., 2022; Clark, 2008; Risko & Gilbert, 2016). Given these observations, we propose a developmental hypothesis with two parts. First, behavioral control of the visual input emerges as infants begin to better control motor systems (Adolph & Franchak, 2017). Second, this emergent period of behavioral control of the input is plays a key role in the development of executive function.

Critically and as illustrated in Figure 5, mature social partners through their behaviors also directly influence the visual input to infants and thus the infant's looking behaviors. The closed-loop system illustrated in Figure 5 provides a pathway through which parent
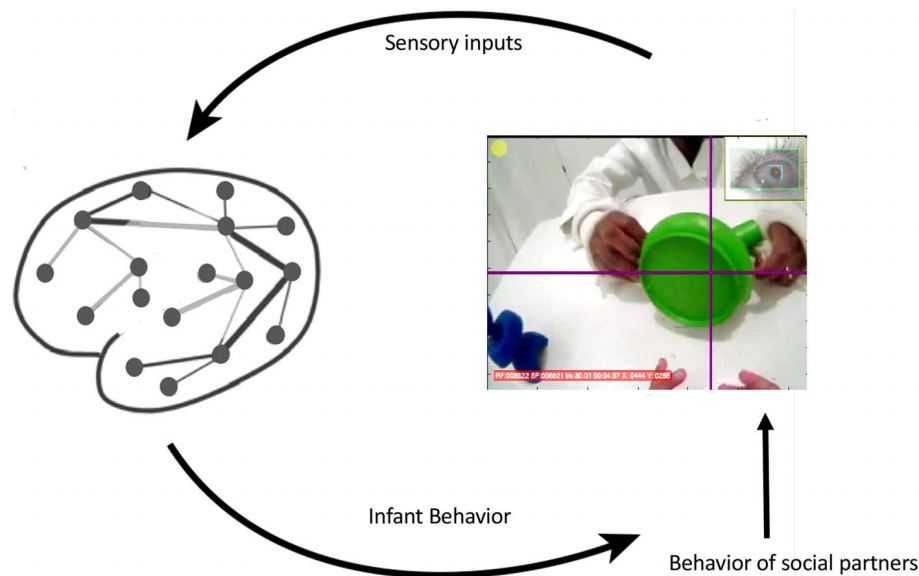
FIGURE 6 Illustration of a hypothesized closed-loop mechanism for sustained visual attention and their possible role in brain networks responsible for the development of self-regulatory processes. At each moment visual input perturbs ongoing brain activations across multiple networks, activating memories, generating context specific goals, and behavior. The generated behavior selects and structures the spatial layout of visual information at the next time step. The behavior of social partners also selects and structure the visual input for the infant. Momentary input incrementally supports, tunes, and trains neural circuitry relevant to sustained attention during infant play and for other self-regulatory processes.

behavior may scaffolds infant visual attention and sustained attention. Past work shows that the key components of parent behavior are responsiveness to the child's own interests during play with potent effects of parent looks to, talk about, and touches of the object to which the infant is already attending (Fay-Stammbach et al., 2014; Suarez-Rivera et al., 2019; Yu & Smith, 2016). Thus, parent behavior that is responsive and coordinated with infant behavior may be particularly potent as a booster of ongoing infant momentary interest in objects, triggering and helping the infant sustain the behaviors (inhibiting head movements) that sustain a visibility advantage for objects during sustained attention. The present finding that infants control their body to control the input to control visual attention raises multiple empirically testable hypotheses about the development of sustained attention in infancy during the period in which individual differences in attention begin to predict later developments in executive control.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

DATA AVAILABILITY STATEMENT

The de-identified data, code for analyses, and supplementary analyses are available on the Open Science Framework https://osf.io/tcd35/.

ORCID

Linda B. Smith https://orcid.org/0000-0001-7163-8181

REFERENCES

Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2), e1430. https://doi.org/10.1002/wcs.1430

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

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. https://doi.org/10.1016/j.cognition.2022.105256

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in neurosciences*, *34*(4), 210–224. https://doi.org/10.1016/j.tins.2011.02.003

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bechtel, W., & Bich, L. (2021). Grounding cognition: heterarchical control mechanisms in biology. *Philosophical Transactions of the Royal Society B*, *376*(1820), 20190751. https://doi.org/10.1098/rstb.2019.0751

Berger, S. E., Harbourne, R. T., & Guallpa Lliguichuzhca, C. L. (2019). Sit still and pay attention! Trunk movement and attentional resources in infants with typical and delayed development. *Physical and Occupational Therapy in Pediatrics*, *39*(1), 48–59. https://doi.org/10.1080/01942638.2018.1432005

Biswas, D., Arend, L. A., Stamper, S. A., Vágvölgyi, B. P., Fortune, E. S., & Cowan, N. J. (2018). Closed-loop control of active sensing movements regulates sensory slip. *Current Biology*, 28(24), 4029–4036. https://doi.org/10.1016/j.cub.2018.11.002

Brandes-Aitken, A., Braren, S., Swingler, M., Voegtline, K., & Blair, C. (2019). Sustained attention in infancy: A foundation for the development of multiple aspects of self-regulation for children in poverty. *Journal of experimental child psychology*, 184, 192–209. https://doi.org/10.1016/j.jecp.2019.04.006

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., 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). https://doi.org/10.1167/jov.21.8.18

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

Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain–body–behavior networks. *Trends in cognitive sciences*, 18(8), 395–403. https://doi.org/10.1016/j.tics.2014.04.010

Campbell, R., (2017). raacampbell/shadedErrorBar, https://github.com/raacampbell/shadedErrorBar

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body, and environment. *Trends in neurosciences*, 20(12), 553–557. https://doi.org/10.1016/S0166-2236(97)01149-1

Clark, A. (2008). Where brain, body and world collide. In *Material agency* (pp. 1–18). Springer.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, 869–879. https://doi.org/10.2307/1127638

Colombo, J. (1997). Individual differences in infant cognition: Methods, measures, and models. In *Developing Brain Behaviour* (pp. 339–385). Academic Press.

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

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of object constancy. *The Quarterly Journal of Experimental Psychology Section A*, 41(4), 775–796. https://doi.org/10.1080/14640748908402393

Fay-Stammbach, T., Hawes, D. J., & Meredith, P. (2014). Parenting influences on executive function in early childhood: A review. *Child development perspectives*, 8(4), 258–264. https://doi.org/10.1111/cdep.12095

Fisher, A. V. (2019). Selective sustained attention: A developmental foundation for cognition. *Current opinion in psychology*, 29, 248–253. https://doi.org/10.1016/j.copsyc.2019.06.002

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

Frick, M. A., Forslund, T., Fransson, M., Johansson, M., Bohlin, G., & Brocki, K. C. (2018). The role of sustained attention, maternal sensitivity, and infant temperament in the development of early self-regulation. *British Journal of Psychology*, 109(2), 277–298. https://doi.org/10.1111/bjop.12266

Friedman, A. H., Watamura, S. E., & Robertson, S. S. (2005). Movement-attention coupling in infancy and attention problems in childhood. *Developmental Medicine and Child Neurology*, 47(10), 660–665. https://doi.org/10.1017/S0012162205001350

Geeraerts, S. B., Hessels, R. S., Van der Stigchel, S., Huijding, J., Endendijk, J. J., Van den Boomen, C., & Deković, M. (2019). Individual differences in visual attention and self-regulation: A multimethod longitudinal study from infancy to toddlerhood. *Journal of experimental child psychology*, 180, 104–112. https://doi.org/10.1016/j.jecp.2018.11.012

Gibson, J. J. (1979) *The ecological approach to visual perception*. Houghton Mifflin.

Gottwald, J. M., Achermann, S., Marciszko, C., Lindskog, M., & Gredebäck, G. (2016). An embodied account of early executive-function development: prospective motor control in infancy is related to inhibition and working memory. *Psychological science*, 27(12), 1600–1610. https://doi.org/10.1177/0956797616667447

Guan, Y., & Corbetta, D. (2012). What grasps and holds 8-month-old infants' looking attention? The effects of object size and depth cues. *Child Development Research*, 2012, https://doi.org/10.1155/2012/439618

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. https://doi.org/10.1242/jeb.082420

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3), 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

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. https://doi.org/10.1016/j.infbeh.2015.07.001

Jonas, R. A., Wang, Y. X., Yang, H., Li, J. J., Xu, L., Panda-Jonas, S., & Jonas, J. B. (2015). Optic disc-fovea angle: The Beijing Eye Study 2011. *PLoS One*, 10(11), e0141771. https://doi.org/10.1371/journal.pone.0141771

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

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. https://doi.org/10.1016/j.conb.2006.06.009

Kulke, L., Atkinson, J., & Braddick, O. (2015). Automatic detection of attention shifts in infancy: Eye tracking in the fixation shift paradigm. *PloS one*, 10(12), e0142505.d https://doi.org/10.1371/journal.pone.0142505

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

Nakagawa, A., & Sukigara, M. (2013). Variable coordination of eye and head movements during the early development of attention: A longitudinal study of infants aged 12–36 months. *Infant Behavior and Development*, 36(4), 517–525. https://doi.org/10.1016/j.infbeh.2013.04.002

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, 11(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Marciszko, C., Forssman, L., Kenward, B., Lindskog, M., Fransson, M., & Gredebäck, G. (2020). The social foundation of executive function. *Developmental Science*, 23(3), e12924. https://doi.org/10.1111/desc.12924

=Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Oakes, L. M., & Luck, S. J. (2013). Short-term memory in infancy. *The Wiley handbook on the development of children's memory*, 2, 151–180.

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), 22–22. 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

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package 'nlme'. *Linear and nonlinear mixed effects models, version, 3*(1), 274.

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

Posner, M. I. (2008). 18 Neuropsychology of attention. *Neuropsychological research: A review, New York: Psychology Press*; 2008. pp. 331–348.

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-

Richards, J. E. (2010). The development of attention to simple and complex visual stimuli in infants: Behavioral and psychophysiological measures. *Developmental Review, 30*(2), 203–219. https://doi.org/10.1016/j.dr.2010.03.005

Richards, J. E., & Hunter, S. K. (1997). Peripheral stimulus localization by infants with eye and head movements during visual attention. *Vision Research, 37*(21), 30 https://doi.org/10.1016/S0042-6989(97)00082-5

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Rosander, K., & von Hofsten, C. (2000). Visual-vestibular interaction in early infancy. *Experimental Brain Research, 133*(3), 321–333. https://doi.org/10.1007/s002210000413

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

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*, 100699. https://doi.org/10.1016/j.dcn.2019.100699

Rossi, A. F., Pessoa, L., Desimone, R., & Ungerleider, L. G. (2009). The prefrontal cortex and the executive control of attention. *Experimental brain research, 192*(3), 489–497. https://doi.org/10.1007/s00221-008-1642-z

Rothbart, M. K., Sheese, B. E., Rueda, M. R., & Posner, M. I. (2011). Developing mechanisms of self-regulation in early life. *Emotion review, 3*(2), 207–213. https://doi.org/10.1177/1754073910387943

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

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental psychology, 26*(1), 85. https://doi.org/10.1037/0012-1649.26.1.85

Ruff, H. A., Lawson, K. R., Parrinello, R., & Weissberg, R. (1990). Long-term stability of individual differences in sustained attention in the early years. *Child development, 61*(1), 60–75. https://doi.org/10.2307/1131047

Schmitow, C., Stenberg, G., Billard, A., & von Hofsten, C. (2013). Using a head-mounted camera to infer attention direction. *International Journal of Behavioral Development, 37*(5), 468–474. https://doi.org/10.1177/0165025413495749

Schroer, S. E., & Yu, C. (2022). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*, e13290.

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*, https://doi.org/10.1111/infa.12475

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. https://doi.org/10.1111/j.1467-7687.2009.00947.x

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

Taub, M., & Yovel, Y. (2020). Segregating signal from noise through movement in echolocating bats. *Scientific Reports, 10*(1), 1–10. https://doi.org/10.1038/s41598-019-57346-2

Thompson, A., & Steinbeis, N. (2020). Sensitive periods in executive function development. *Current Opinion in Behavioral Sciences, 36*, 98–105. https://doi.org/10.1016/j.cobeha.2020.08.001

Van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience, 22*(3), 477–483. https://doi.org/10.1038/s41593-018-0335-6

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

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), e12667. https://doi.org/10.1111/desc.12667

Wass, S. V., Whitehorn, M., Haresign, I. M., Phillips, E., & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in cognitive sciences, 24*(4), 329–342. https://doi.org/10.1016/j.tics.2020.01.006

Werchan, D. M., & Amso, D. (2020). Top-down knowledge rapidly acquired through abstract rule learning biases subsequent visual attention in 9-month-old infants. *Developmental Cognitive Neuroscience, 42*, 100761.2001d https://doi.org/10.1016/j.dcn.2020.100761

Wolfe, J. M. (2020). Visual search: How do we find what we are looking for. *Annual review of vision science, 6*(1), 539–562. https://doi.org/10.1146/annurev-vision-091718-015048

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 1–8. https://doi.org/10.1038/s41562-017-0058

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. (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., Shen, H., Pereira, A. F., & Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE transactions on autonomous mental development, 1*(2), 141–151.

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), e12735. https://doi.org/10.1111/desc.12735

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.