PROGRAMA DE DESARROLLO DE LAS CIENCIAS BÁSICAS
Ministerio de Educación y Cultura,
Universidad de la República

Tesis de Maestría
Área Biología
Subárea Neurociencias

One-year old infants control bottom-up saliencies to sustain visual attention

Estudiante: Andrés Méndez Oehninger
Orientador: Leonel Gómez Sena
Facultad de Ciencias, Universidad de la República
Co-orientador: Linda Smith
Department of Psychological and Brain Sciences, Indiana University

TRIBUNAL:
Presidente: Dr. Ángel Caputi
Vocales: Dres. Verónica Ramenzoni y Álvaro Cabana

MARZO 2021

Abstract:

Gaze is naturally attracted to highly visible stimuli. A vast amount of work has focused on how perceivers internally suppress these saliencies to purposely sustain gaze on a target. Infants' purposeful gaze to objects during active play has been interpreted as a marker of the early emergence of inhibitory processes principally because of the strong predictive relations between sustained gaze by infants and the later development of generalized inhibitory control and self-regulation. Here we show that one-year-old infants' sustained gaze during active play is linked to the external rather than internal suppression of competitors. We measured infants' moment-to-moment gaze during active object exploration and also measured the moment-to-moment visual size of objects in the infant field of view, a salience that is well-known to robustly attract gaze. We found that when infants shifted gaze to an object they simultaneously changed the spatial relation of the head to the object such that the attended object was visually larger than competitors. The onset, duration, and offset of the salience advantage coincided with the onset, duration and offset of the look. Longer looks coincided with a larger visual size advantage than shorter looks. These findings show that infants' purposeful gaze, at a time scale simultaneous with the look itself, puts external visibility gains on selected targets. This bottom-up path for top-down control appears fundamental to vision in a freely moving perceiver and raises new questions about the nature of attention and the collaboration of exogenous and endogenous control in the development of self-regulatory processes..

This thesis is the outcome of joint work with Linda Smith and Chen Yu from the Department of Psychological and Brain Science in Indiana University. This document is an extended version of a manuscript in preparation for a peer-reviewed journal. .

CONTENTS

1. INTRODUCTION

> *"Nothing determines me from outside,*
> *not because nothing acts upon me,*
> *but, on the contrary, because I am from the start*
> *outside myself*
> *and open to the world."*
> Mearleau-Ponty (1982)

Every second, in our childhood or during our grown up lives, even when deeply focused on a single important task, our retina is exposed to tens of recognizable entities that surround us. These entities compete, but it is well known not equally. The center of our retina, to which we generally orient the objects we attend to, consists of a densely packed number of receptors. This provides us with high acuity in the center of our visual field (Dowling, 1987; Lee, 1996; Meister & Tessier-Lavigne, 2013). However, moving, novel and visually (or emotionally) salient objects, projected to any point of our retina, can capture our attention at any moment (Abrams & Christ, 2003; Carretié et al, 2012, Carretié, 2014). Little would we achieve in our lives, then, if our attention was at the whim of external processes. In words of Desimone and Duncan (1995): "An attentional system [...] would be of little use if it were entirely dominated by bottom-up biases" (Desimone & Duncan, 1995). What is needed, in any living organism, according to these authors, is a top-down control to bias attention towards relevant information to behavior. Since their influential work, considerable contemporary research has focused on the endogenous networks that place gains on some external signals and inhibit others in response to the context and the goals of the perceiver (van Moorselar & Slagter, 2020; Gaspelin & Luck, 2018). Internal and external influences have been generally conceived as competing factors guided by separate networks, a competition which needs to be internally solved to organize behavior (Desimone & Duncan, 1995; Bowling et al, 2020; Buschman & Miller, 2007, Chica et al 2013).

Adults are much better at avoiding distractors than infants. The development of early visual attention has been traditionally characterized by a shift from exogenous, stimulus-driven orienting of attention (bottom-up) to greater endogenous control (top-down) (Colombo 2001). Infants' purposeful and sustained gaze to objects has been interpreted as a marker of the early emergence of endogenous (top-down) control principally because of the strong predictive relations between sustained gaze by infants and the later development of generalized inhibitory control and self-regulation (Welsh et al, 2010; Reck & Hund, 2011). How does this endogenous control emerge? Internal solutions have generally focused on the study of neural networks that underlie attentional processes in controlled laboratory studies. This has led to the identification of relevant changes in brain organization - cortical maturation, long-range connections - that occur alongside the development of visual attention (see Amso & Scerif 2015 for a review).

The study of internal pathways in such restrained settings has, however, left much unknown about the development of attention. Current reviews show conflicting results on attentional tasks in infants when making changes in few task parameters (Ristic & Enns, 2015a, Ristic & Enns, 2015b), point to difficulties in integrating evidence from multiple technical approaches (genes, environment, behavior, brain) in both typical and atypical populations (Amso & Scerif, 2015), and report a lack of transfer from attention training to other related processes (Amso & Scerif, 2015). There is a need to take a closer look to how the coupling between attention and other relevant processes change through development and influence attention. These additional processes that have been implicated in the development of attention include learning, memory and the perceptual and motor mechanisms that orient and select the stimuli which we attend to. (Scerif, 2010; Amso & Scerif, 2015).

External factors, generally studied as distractors in an ongoing task with distinct properties (generally spatial) which separates them from relevant targets, are not entirely independent in unrestrained settings. Putting gains to targets and inhibiting distractors cannot be achieved without a relatively stable external signal in the first place. Eyes are in heads, and heads are in bodies. Bottom-up processes (from target and competitors) depend on the orientation of the sensory systems in the world. How, then, do infants with immature sensoriomotor systems orient and coordinate their bodies towards the attended stimuli is a relevant matter that can shed light on the underlying mechanisms that contribute to select and sustain attention.

The coupling of sensory and motor processes has been widely studied in animal cognition and neuroethology as a fundamental property of perception (Kelinfeld et al, 2006; Crapse & Sommer, 2008; Simony et al; 2008; Caputi et al, 2004). Organisms move their sensory surfaces towards their target, actively changing the input they receive from the world, to extract relevant information (Kleinfeld et al, 2006, Hoffman et al, 2013, Taub & Yovel, 2020). The role of action has been central in embodied and enactive perspectives to cognition, and the way such perspectives understand cognitive development (Port & Van Gelder, 1995; Thelen & Smith, 1996; Von Hofsten; 2007; Gottlieb; 2007; Barsalou, 2008; Engel et al 2013). A recent methodological advance in developmental science is to use head-mounted cameras and eye trackers to measure and quantify the visual information in children's field of view (Yu & Smith; Smith et al, 2015). The ways in which toddlers interact with objects are not like the ways adults do (Yoshida & Smith, 2008; Smith et al, 2011). Toddlers' bodies, postures, motor behaviors, interests, and goals are very different from adults' and lead to unique patterns of object interactions (Yu & Smith, 2012, Pereira et al, 2014). A study of first person-view of children's environments at different ages shows changing visual properties that might be relevant for development and learning (Smith et al, 2018). Altogether, these studies underscore a key component of the perception-action coupling for development, and for this work. Body orientation and movement define the input structure from which children learn; and therefore the perceptual properties of the objects they attend to and of those that compete for their attention. This opens relevant questions about the possible role input regulation, rather than internal signal inhibition, may play in how infants become able to control their attention.

Here we study looking behavior while infants play with their caregivers. We show a tight coupling between external visual properties and gaze as children play with novel objects and sustain their attention to those objects. We suggest that a key component of how attention is controlled works by reaching out to the world to alter the visual input in ways that might optimize processing of task relevant information (see Byrge et al, 2014 for a proposal on how mutually influential shaping of body, environment and brain is critical for development). Through their whole body and eyes, children restructure the visual input such that objects children sustain their attention to are salient, with decreased visual competition from distractors. We suggest this is a developmentally foundational component of human visual selection in everyday life which needs to be addressed to better understand how visual attention control develops. We do not claim modulation of internal signals is not a relevant component, but that the view of external factors as uncontrolled signals that need to be internally inhibited might be missing a key point in how infants learn to coordinate their attention. We used data from infant head-mounted eye-trackers and coding of real-time behaviors to extract measures of looking, handling and object size. We chose this context as a large part of children's play with objects occurs in the context of social play.

2. THEORETICAL FRAMEWORK

In this section we will i) describe the predominant developmental view on how infants control attention to direct and sustain their gaze on a given target, ii) discuss theories that emphasize the relevance of taking into account how our body and the environment are coupled and might lead to complex behaviors, iii) and iii) present a recent body of work that looks at the body-environment coupling by adopting infant's first person perspective, relevant to the study of attention and to this work.

2.1. From bottom-up to top-down control of attention: the predominant developmental view

Laboratory studies clearly show that adults can covertly attend without moving eyes or heads to the target and without extrinsic salience advantages, resolving the competition among distractors through purely internal means (Carrasco, 2011). Infants' purposeful and sustained gaze to objects has been interpreted as a marker of the early emergence of this endogenous (top-down) control. Research in ecological settings measure the length of an unbroken look to an object as infants play with several objects that compete for their attention (Ruff & Capozzoli, 2003; Kannass & Oakes, 2006). In controlled laboratory studies, a predominant approach to studying how this ability develops has been to adapt attentional tasks used in adults to detect when at the behavioral level this ability appears in infancy and to study the development of the neural pathways in childhood that precede and relate to the activations seen in adults. Both in adults and infants, internal and external signals have been traditionally conceived as competing factors guided by independent networks that need to be solved to organize behavior.

The development of early visual attention has been characterized by steady movement away from more exogenous, stimulus-driven orienting of attention (bottom-up) to greater endogenous control (top-down) (Colombo 2001, Wright and Vlietstra, 1975, Ruff Rothbart, 1996, although with different names). Although active scanning of their visual environment has been described at early stages, internally guided attention slowly progresses from the end of the first semester onwards (Colombo 2001). At the end of the first year infants show the ability to suppress saccadic movements and use strategic shifting in paired-comparison tasks (see Colombo 2001 for review). More purposeful forms of visual attentional control and allocation (Riviere and Falaise, 2011), decrease in distractibility (Kannas, Oakes, Shaddy 2009, Ruff & Capozzoli, 2003), increase in sustained focused attention (Ruff, Capozzoli, & Weissberg, 1998) and integration of endogenous and exogenous cues (Iarocci, Enns, Randolph and Burack, 2009; Ristic and Kingstone, 2009) will fully mature after the first year of life, towards the preschool years and beyond. These latter studies on distractibility and sustained attention are reported in more ecological contexts of infant toy play with multiple objects. Sustained visual attention develops incrementally from late infancy through early childhood and measures of sustained attention are predictive of later cognitive developments (Welsh et al, 2010; Reck & Hund, 2011). Most of these later changes in attention are related to executive control. Conflict resolution and monitoring are considered key attention components needed to acquire more complex executive functions (Amso, 2015).

Multiple neurobiological approaches have addressed the question of how we select a target object, inhibit distractors and sustain attention: single neuron recordings, brain connectivity and brain oscillations (Moran & Desimone, 1985; Jensen & Mazaheri, 2010; Clayton et al, 2015; Gaspelin, 2018; Rosenberg et al, 2016; van Moorselar, 2020). The field was highly influenced by the discovery from single-neuron recordings in monkey visual cortex of neurons whose response are determined by the properties of attended objects (even when other objects are present in their receptive field) (Moran & Desimone, 1985) and evidence from lesions in parietal cortices - also in monkeys - which disrupt covert selective attention (Posner y col, 1984). Several studies have addressed the existence of networks in the brain for the exogenous and endogenous attentional control. The description of more posterior regions (superior colliculus, parietal cortex, visual cortex) associated with bottom-up control and frontal regions (Frontal Eye FIeld, Dorsolateral prefrontal cortex) supporting top-down control (Buschman & Miller, 2007; Bowling et al, 2020), together with the posterior to anterior axis of cortical development (Casey et al, 2005), has provided a coherent framework for the exogenous to endogenous trajectory of attention control with development.

2.2. Taking embodiment seriously

No matter how internal and volitional the allocation of attention is, attention does not exist without a visible target. Vision at every moment selects through behavior directed to the world: by moving our eyes, to direct gaze to a target. In doing so, we change the input at the level of the retina, making the target to which gaze is directed more salient and easier to visually process than the competitors. This occurs both because of the greater contrast sensitivity around the gaze point (the fovea), and the convergence of both eyes that determines what remains in and out of focus. Where one looks determines what one sees, shaping visual input. This looking is not achieved solely by eye movements, but also head, shoulder and body movements organized to reach a certain goal. The position of our head and eyes, and several muscles to control them provides us - and other primates - with a precise and sophisticated system for sensoriomotor exploration (Stryker & Schiller, 1975; Crawford et al, 1999; Tomlinson & Bahra. 1986). Brain studies in monkeys show neck and eye muscles are jointly regulated by the superior colliculus - a relevant region in eye movements - for the coordinated orientation of attention (Cornell et al, 2004). Despite this relation between body orientation, exploration and attention, the role of sensory and motor processes have not been central in the study of endogenous control.

There are theoretical and methodological reasons why low-level processes have not been included in explanations of higher cognition. The study of any cognitive process requires the answering of specific questions and a decision of where to look for these answers: i) are there rules that guide this process?, ii) where do we look for the elements that make these rules possible?, iii) are these rules pre-existing or created in the moment? Often, the study of cognition is divided into two distinct approaches[1]. Briefly, 'cognitivist' approaches that zoom-out of behavior to capture the abstract properties, computed by the brain, that govern behavior and describe cognition (cognition as rules), and the 'embodied' approaches that zoom-in behavior to describe the ongoing brain-body-environment interactions that support the organization of complex behavior and cognition (cognition as emergent complex behavior). For long, embodied approaches have been downplayed by its inability to provide empirical explanations of higher cognition. True cognition was expected to be amodal, and distinguishable from the physical constraints that make it possible (Fodor, 1975). Moreover, the study of interactions between brain, body and environment present huge challenges to researchers not only because of the diverse nature of the data analyzed but also because of the technical demands of collecting, analysing and storing high-dimensional data. Experimental approaches to cognition, specially in humans, in too many cases have driven sensory-motor processes out of the picture.

Studies of sensory systems in animals with careful control of stimuli and measures of brain and behavior have been an exception, but often considered mechanisms not analogous to human higher cognition. These studies already described several ways in which perception and action are not mere input-output signals but are regulated in precise ways to control stimuli input and to extract information not present in the stimuli itself (Hoffman et al, 2013, Taub & Yovel, 2020). Studies that disrupt the motor-sensory coupling in animals show evidence

[1] Forcing the "Cognitivist" and "Embodied" label to each data driven proposal or author's work would be ill-posed and would be of little help. Reviewing this discussion is out of the scope of this work (for an elegant review of important elements of this discussion see Vernon, 2014)

of impaired cortical development (Attinger et al, 2017). The need to study the way behavior changes sensory input and how this influences back on behavior as fundamental to cognition is by no means new. Powers (1973) already stated: “we know nothing of our own behavior but the feedback effects of our own outputs. To behave is to control perception” (see Ahissar et al, 2016, for a current proposal based on similar principles).

There are strong reasons to put embodied and dynamic perspectives back to the table. First, the growing possibility of running experiments with multimodal data have boosted their explanatory power. Second, cognitive science has not yet provided full explanations of cognition nor how it develops, and there is still much to answer on how best to approach the development of cognition and its disorders. Finally, the difficulty in explaining higher cognition does not always justify leaving dynamic perception-action loops out of the picture. Embodied and more cognitivist explanations need not be antagonists. An understanding of the preserved rules to a process being studied and the relevant ongoing dynamics are both needed. As Esther Thelen and Linda Smith have already shown, even an apparently simple motor behavior like walking is guided by internal patterns but cannot be reduced to the genetics and neurophysiology of central pattern generators in the central nervous system (Thelen and Smith, 1996). The rules of how to walk are distributed in the brain and throughout the body (eg. spring-like properties of the leg). A central pattern generator not only does not possess privileged walking knowledge but also does not explain the expertise even young walkers achieve in multiple surfaces and the ability to adapt walking in the moment.

Advances in i) eye tracking and neural recordings in unrestrained settings, ii) machine learning, and iii) computer vision, all present interesting opportunities for theories that aim to understand how the coupling of perception and action shapes cognition and the brain in humans. These technologies, be it applied to the brain, behavior or the incoming sensory data, have shown multiple sources of regularities and information present at each level potentially important for cognition. The input we receive carries relevant information that guides behavior. The use of computational models in 2D displays shows that bottom-up saliency in images predict what objects people considered most interesting (Elazary & Itti, 2008). This exploration is not, however, unbiased. Image processing applied to scenes shows how the spatial layout of a scene and the different categories of depth related to the observer carry different types of information which affects attention memory and visual search (Castelhano & Krzyś, 2020). This exploration is grounded in bodily processes that go beyond eye movements. Using head-mounted eye-trackers in natural settings Abbot & Faisal (2020) show how the conformation and dynamics of the body predict eye-movements, improving previous models of attention allocation. Moreover, evidence shows the brain is not a passive processor of unaltered images of the environment. Neural recordings in unrestrained animals show a modulation of cortical response during locomotion, suggesting sensory cortices not only react to sensory stimuli but also to ongoing movements and behavioral states (Niell & Stryker, 2010; Schneider, 2020). Although detailed mechanisms are described in animals, the basic principles are likely to apply to humans (see Buzsáki, 2019). Finally, deep neural networks have revolutionized cognitive neuroscience because of its ability to learn, achieving performance similar to humans on specific tasks. The use of such neural networks not only support the existence of brain areas associated with face, body and scene processing, but show that face, body and scene data shape the inner layers of the networks in analogous ways to what happens in the brain (Dobs et al, 2019; Dobs et al, 2020). This opens the possibility that properties of objects explored might play a role in shaping neural circuits, suggesting an important role of experience in learning and shaping the brain. Accordingly, recent work suggests experience can build internal representations that could explain infants' rapid ability to learn in specific domains such as object recognition (Orhan et al, 2020). Cognition then might be better characterized by understanding the complex patterns through which brain, behavior and environment are interconnected.

2.3. Embodied change

The explanations about how things work do not necessarily answer what makes them change, which is central to the study of development. However, identifying the underlying factors to a specific process is necessary to know what needs to be considered when thinking about development. The search for adult-like forms of cognition in infants, although useful, might make the development of our own cognition hard to track, and might be highly misleading. Complex behavior (and cognition) might emerge from multiple interactive processes (Thelen & Smith, 1996). Cognition, in its adult form, might be created through multiple alternative pathways and solutions to the challenges in infants' everyday life (Smith, 2013). For a long time the messy, dynamic and variable nature of sensorimotor processes was seen as noise with little relation to the true rules of cognition (Smith & Sheya, 2010). Advances described above provide us with tools to make sense of the mess, and already suggest lots of possible elements that might be playing a role in how cognition in its adult form is achieved. Development is not restricted to the mere learning of more complex internalized ways to analyze incoming input, but is better described by the interconnected and mutually influential changes between brain, body and behavior. In words of Byrge et al (2014) "brain and behavioral development is a process through which the information that moves the system forward is created probabilistically in interactions that cross time scales and span the brain, the body, and behavior.“ A more unified understanding may emerge then from theory and experiments focused on how knowledge emerges through real time interactions in the world that create the data for learning and that change as the infant changes.

2.4. *A first-person perspective of infants' visual input.*

Studying the way body and environment interact requires both the ability to measure behavior and sensory input. What is the projected stimuli over sensory systems is not a trivial question. Experiments in animal ethology carefully design the environment and stimuli in ways to constrain behavior and control input. Inferring sensory input from less restrained conditions in more ecological settings is far more complex. A recent methodological advance in developmental science is to use head-mounted cameras and eye trackers to measure and quantify the visual information in children's field of view (Smith et al, 2015). While camera images do not represent the actual sensory input, they do reflect relevant spatial and temporal properties of the visual world.

The use of head-cameras at home led to the opportunity of taking a glimpse at what children see in their daily lives at different temporal scales, and has had relevant consequences for our understanding of the development of visual object recognition and word learning (Clerkin et al, 2017; Smith et al, 2018). In the lab, the use of head-cameras and eye trackers provide moment-to-moment information of the visual properties of visual scenes and gaze. Together with motion tracking and coding for speech and touching, this multimodal fine-grained and first person approach has provided alternative explanations of how word-object reference, vocabulary learning and social coordination might emerge so robustly in development (Yu & Smith, 2017; Slone et al, 2019; Suanda et al, 2019).

Studies of infant-parent play show infant´s head camera images have unique visual properties as compared to caregivers. Parents' views generally keep both infants and all objects in view. Parent's faces, however, are not often present in children's head cameras while they play; infant scenes usually show few objects in view such that one dominates the scene. How much of the world around an infant is filtered is a consequence of infants' body properties (small bodies, short arms). During free toy play, infants tend to bring attended objects towards their body's midline and attend to objects with head and eyes aligned, systematically creating images with the attended object at the center (Bambach et al, 2016). This centering not only applies to the objects children interact with. Infants' overall visual experience is shaped by the physical relation between infants' bodies and the locations of objects in the world (Luo & Franchak, 2020). The relation between object dominance and gaze has been suggested but not studied, as detailed gaze analysis was not included in the studies mentioned.

Such unique embodied visual properties are relevant to this work in two important ways. First, body orientation and movement define the input structure from which children learn; and therefore the perceptual properties of the objects they attend to and of those that compete for their attention. The understanding of what mechanisms underlie attention in real time makes these dynamic properties of toddlers' active vision relevant (see discussion in Smith et al, 2011). Second, such unique properties imply adult and infant's attentional systems operate in very different sensory distributions. The adult-like internalized form of attention control might not necessarily resemble the mechanism occurring earlier in development. Altogether, the visual properties of infants scene and its close relation to infants ongoing behavior opens relevant questions about the possible role input regulation may play in how infants become able to control their attention.

3. CURRENT STUDY

Both basic and applied research converge on the need to understand attention in real-time and the perceptual and motor contributions to how infants select and sustain attention. Previous research suggests selection might not be restricted to internal processes but might include the regulation - through the body - of the incoming sensory input. In other words selecting and sustaining attention might be closely related to the external saliency of attention targets. Such finding would question traditional views of how endogenous control develops and suggest that the dynamic coordination of the perceptual and motor processes needed for visual selection plays a fundamental role in the development of attention control, its regulation and more complex forms of executive behavior.

We study looking behavior while infants play with their caregivers, in ages described to still have poor endogenous attentional control (12 to 16 months old). The current work uses data from infant head-mounted eye-trackers and coding of real-time behaviors to extract measures of looking, handling and scene visual properties. We chose this context as a large part of children's play with objects occurs in the context of social play. We measured infants' moment-to-moment gaze during active object exploration and also measured the moment-to-moment visual size of objects in the infant field of view, a salience that is well-known to robustly attract gaze (Borji et al, 2013; Cohen, 1972; Guan & Corbetta, 2012; Sensoy et al, 2020). We expected visual input to have similar properties as described in previous studies - with one object dominating the scene - and that such properties would relate to gaze in ways that supports sustained attention.

Real-time interaction data can be studied at different levels (see **Figure 1**). Frame-level analysis provides detailed information of the dynamic visual properties of scenes from the child's first person perspective and indicates the frame-by-frame target of infants' gaze (or eye ROI). Event level analysis provides information about onset, duration and offset of events. We were interested in the visual properties of infants scenes, how they relate to gaze at the frame level ('Are visual dominance and gaze-target related?') and the event level ('What is the temporal relation of gaze and the visual properties of target objects?').

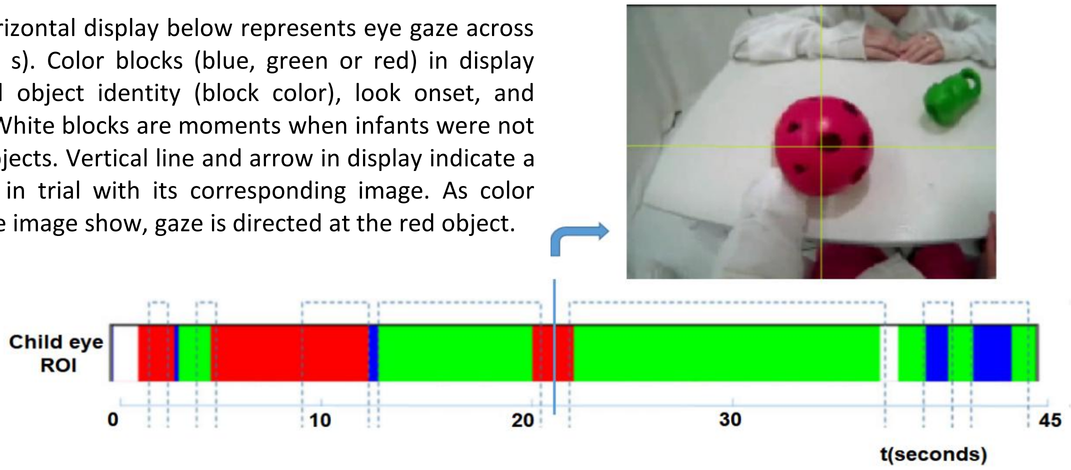


Figure 1. Horizontal display below represents eye gaze across one trial (45 s). Color blocks (blue, green or red) in display show looked object identity (block color), look onset, and look offset. White blocks are moments when infants were not looking at objects. Vertical line and arrow in display indicate a given frame in trial with its corresponding image. As color block and the image show, gaze is directed at the red object.

4. METHOD

4.1. Participants

The final sample consisted of 45 parent-infant dyads with the infants ranging in age from 11 to 16 months - an age group considered to have poor endogenous control - while they played with age-appropiate toys. The data included is already-collected data, collected throughout the years with the same experimental settings and procedures. Fifteen additional dyads began the study but only infants that provided enough data (see below) throughout the whole session were included. Incomplete sessions occurred when infants refused to wear the measuring equipment. Dyads were recruited from a population of working and middle class families in a Midwestern town in the USA.

4.2. Experimental Setup

Infants and parents sat across from each other at a small table (61cm × 91cm × 64cm). Infants sat on a chair in front of their parents. Parents were asked to sit

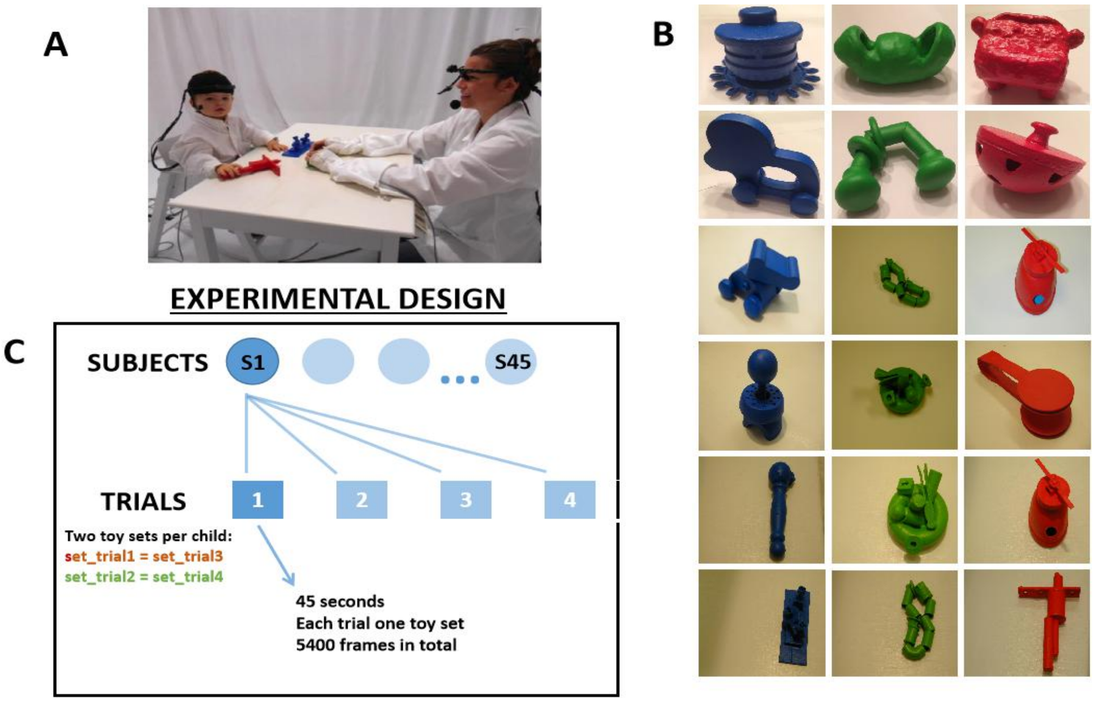


Figure 2. Experimental setup. **A**. Dyads sat across a white table in a white room while playing with colored objects. **B**. In each trial each dyad played with a blue, green and red object. **C**. Diagram shows the multi-level structure of the analysed data.

on the floor so that their eyes and heads were at the same distance from the tabletop as those of infants. The table, floor and walls were white and both participants wore white smocks which, together with the toy color selection, supported computer object recognition (**see Figure 2A and B)**.

Both participants wore head-mounted eye trackers from Positive Science, LLC (Franchak, Kretch, Soska, Babcock, & Adolph, 2010; Yu & Smith, 2013). Each eye-tracking system includes an infrared camera—mounted on the head and pointed to the right eye of the participant that records eye images, and a scene camera (see in **Figure 4A**) capturing the first person view from the participant's perspective. Each eye tracking system recorded both the egocentric-view video and gaze direction (x and y) in that view, with a sampling rate of 30 Hz. Three additional cameras recorded the interaction from third-person views.

4.3. Stimuli

Each child played with two sets of six novel toys, made from multiple and often moveable parts and constructed in the lab to be interesting and engaging for infants (**see Figure 2B**). Each set consisted of three toys (average size 288 cm^3) of unique uniform color (blue, green and red).

4.4. Procedure

Three experimenters worked together during the experiment. One experimenter played with the infant while another placed the eye tracking gear low on the forehead of the infant at a moment when the child was engaged with a toy used only for this phase of the experiment. The third experimenter controlled the computer to ensure data recording. To collect calibration points for eye tracking, the first experimenter directed the infant's attention toward an attractive toy used only for calibration while the second experimenter recorded the attended moment that was used in later eye tracking calibration. This procedure was repeated 15 times with the calibration toy placed in various locations on the tabletop. Parents were told that the goal of the experiment was to study how parents and infants interacted with objects during play and therefore they were asked to engage their infants with the toys and to do so as naturally as possible. For each age group, each of the two sets of toys was played with twice (**see Figure 2C**). Order of sets (ABAB or BABA) was counterbalanced across dyads.

4.5. Data processing and coding

The quality of eye-tracking video for each dyad was checked to ensure calibration quality. Re-calibration was conducted if necessary. The eye-tracker collected data at a rate of 30 frames per s for approximately 360 s (four trials with 1.5 min per trial) of interaction, yielding potentially 10,800 data points per measure for each participant. Of this total possible, “missing” frames include eye blinks and periods when the infant was off-task (e.g., looking around the room rather than at the objects or parent). To analyse similar data across children, for each of the 45 subjects we considered the first 45 seconds of each of the four trials (children with one trial that lasted less than 45 seconds were excluded from analysis). As a result each child provided 3 minutes of data, 5400 frames (**see Figure 2C**). As we were interested in looking behavior to objects, the frame level analysis only contains frames where children were directing their gaze to an object (on task data).

4.5.1. Scene visual properties

Measures of the visual properties of object were taken for each of the three play objects using a custom image-analysis software (see Yu et al., 2009). The image size **(IS)** of each of the three objects - measured by proportion of object pixels in the image - was extracted for each of the approximately 5400 frames contributed by each participant. Relative size **(RS)** for each object was calculated as the proportion of pixels that corresponded to an object from all the object pixels in that frame (**see Figure 3**).

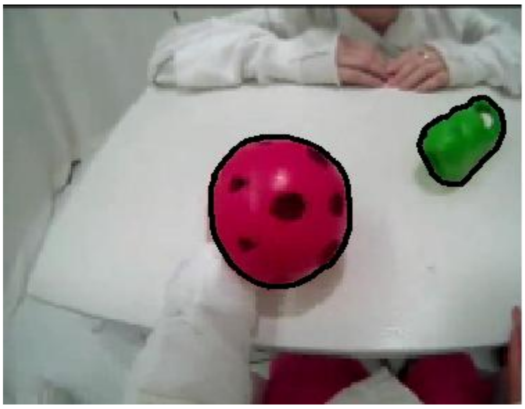


Figure 3. Frame from infant head camera showing two objects in view. Image size (IS) calculated by extracting colored pixels of each object. Red object IS = 6.74 % of all image pixels. Relative size is calculated only including all pixels that correspond to objects (red and green in this frame) in the scene. Red object RS = 0.77 of all object pixels).

Namely, through this work, both image size and relative size, although related to actual object size, refer to the size of the image as projected to the infant's camera and in relation to the other objects in the scene, respectively.

4.5.2. Looking and sustained attention

Looks to each of three objects was coded from eye-tracking data. Coders were highly trained and naïve to the specific hypotheses or experimental questions of this study (for details on coding see Yu & Smith, 2016). A second coder independently coded a randomly selected 10% of the frames with the inter-coder reliability ranged from 82% to 95% (Cohen's kappa = 0.81). Looking to an object was considered as sustained attention (SA) or long when infants directed their gaze to that object for more than 3 s without any looks elsewhere. This 3 s threshold is the average time a 1-year-old infant attends to a single toy in active play (Ruff & Lawson, 1990).

4.5.3. Infant and parent hand contact

Hand behavior, although not central to this work, has been suggested to be relevant for attention and therefore was included for analyses. Infant and parent manual contact with an object was coded frame-by-frame from images captured by the overhead camera and the other two third-person cameras. A custom coding program was used to allow coders to access three views simultaneously to determine which object was manually handled frame by frame. Coders made frame-by-frame yes/no decisions that a parent hand was in contact with an object. A second coder also independently coded randomly selected 25% of the frames of five parents and obtained intercoder reliability assessed by Cohen's kappa of .90 (range 0.76 –0.96).

4.6. Data analyses

Data was analysed at the frame and event level, with means reported at the event, subject or corpus level. Analyses was restricted to ontask data (158.296 of 243.000 total frames, 65%), i.e, moments when infants were engaged with toys. Data processing and analyses was mostly done in Matlab 2017b. Linear mixed-effect logistic regression model was conducted using the lme4 package in R (Version 3.6.1; Bates, Mächler, Bolker, & Walker, 2014). Models used object and subject as random effects and relative size (RS) as fixed effect[2]. At the frame-level, these models assess if RS predicts the largest object in a frame and if RS predicts which objects are looked at. At the event-level, models assess if RS predicts look duration (long or short) (see SI section 9.3 for model summaries).

5. RESULTS

Unconstrained infants wore a head-mounted eye tracker as they freely explored three objects of the same physical size. The relative visual sizes of the objects depend on the objects spatial relations to the perceiver as the objects may be nearer or farther and may overlap (and thus be partially occluded) in the perceiver's line of sight. As a consequence, the relative visual sizes of the objects

[2] As will be mentioned later RS will be the reported measure representing object visual properties in the results section. RS and IS are highly correlated and IS results are reported in the supplmentary material

in play varied moment to moment with small as well as large body movements that alter the angle of the line of sight and the distances of the objects to the head. Frame-by-frame, the image size (IS) of each object was measured in terms of the percentage of total pixels in the head-camera image that belonged to that object; also frame-by-frame, the relative image size (RS) of each object was measured as the proportion of total object pixels in the image (sum of all three objects) that belonged to the object. In the analyzed corpus, the variability of individual object IS and RS was considerable and positively correlated (**Figure 4I**, r^2= 0.40).) In virtually all images (99.4%), one object was visually larger than the other two and often considerably so. The mean IS of the largest object in each frame was 5.7%, SD =3.7%–, a highly visible image size **(Figure 4A-H**). As for relative size, for each frame we considered the largest object RS and the RS of the sum of the competing objects. The maximum RS was 1.0 indicating only one of the three play objects was in the infant view, although this rarely was the case (see **Figure S1**). Logistic mixed effects regression indicated that the largest object RS ($M_{largest}$ = 0.62, $SE_{largest}$ = 0.0013) was on average greater in the proportion of in-view pixels than the sum of the two competitors ($M_{competitors}$ = 0.38; $SE_{competitors}$ = 0.0013; β = 11.05; z = 237.9; p < 0.001) (**see Figure 4J**) .

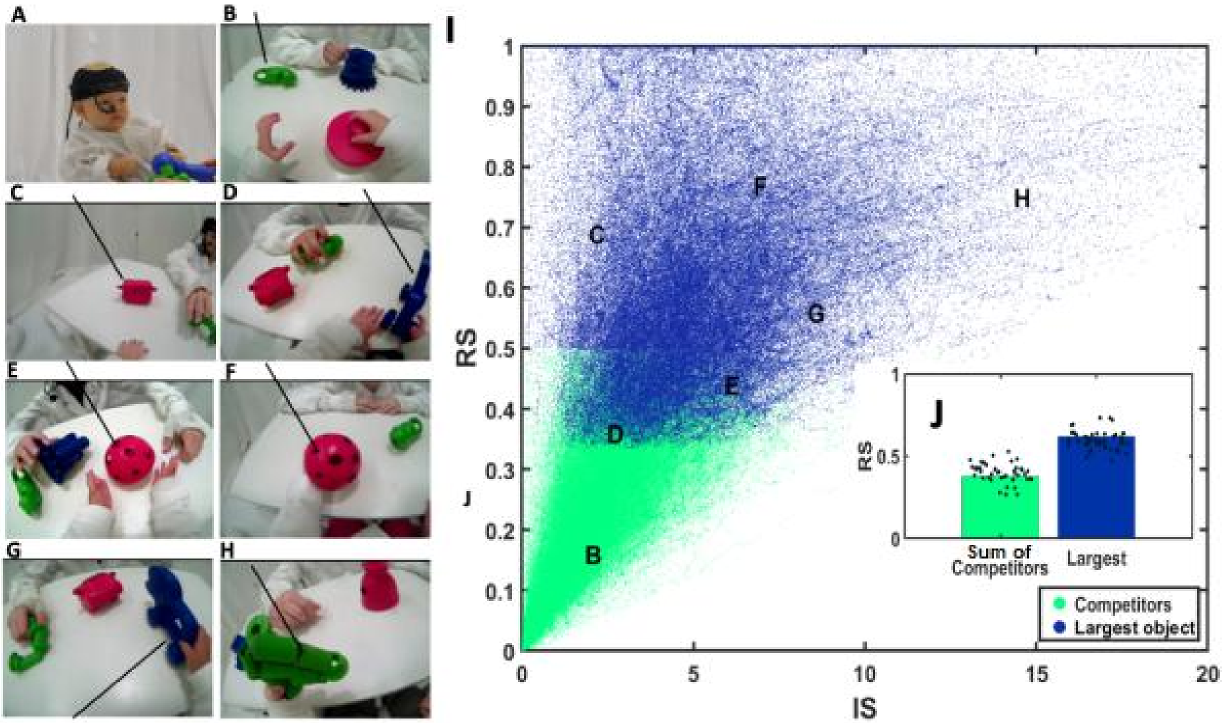


Figure 4. ***A*:** Infant head-mounted eye tracker consisting of the RGB camera (recording scene) and an infrared camera (directed towards infant eye). ***B-H:*** Infant scenes showing multiple object

RS and IS. Black lines indicate reported object. ***I:*** Corpus level IS and RS from objects in view, both for largest object and competitors. Capital letters show the ISxRS of the objects in A-H. ***J:*** RS of largest object VS sum of competitors. Bar represents corpus mean and dots represent subject means

In brief, freely moving infants during the active exploration of objects create scenes in which one object is often considerably larger and thereby more visible than the others. Further analyses will be applied to object RS (results are the same for IS, see SI **Figures S3-S5**).

Greater visibility attracts gaze and did so in the present study (**Figure 5**). The distributions of RS for gaze-directed and competitor objects differed reliably (Two-sample Kolmogorov-Smirnov test, RS $p < .001$) (**Figure 5a**). The gaze-directed object RS ($M_{\text{gaze-directed}}$ = 0.52) was larger than competitors across participants ($M_{\text{competitors}}$= 0.26; β = 5.24; z = 257.7; $p < 0.001$). These results show, as expected, that visual size is a saliency that attracts infant gaze. This is true independent of the number of objects in view (see **Figure S2**). In laboratory studies, the stimuli are images constructed by experimenters and presented on a screen to a constrained infant. In such a context a pattern of results similar to those observed here might be interpreted as gaze driven primarily by exogenous saliencies rather than endogenous control. However, as analyses at the event level will show that, for freely moving infants, the visual sizes of targets and competitors is influenced by the infants own purposeful behaviors which alter the spatial relations of the targets and competitors to the infant body and eyes.

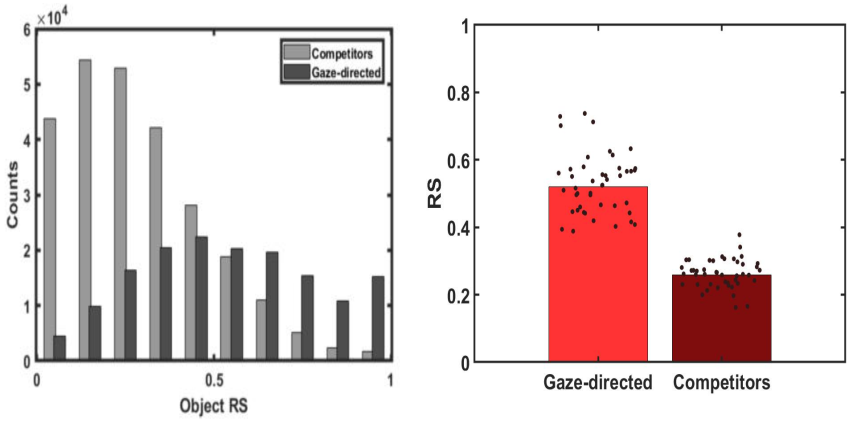


Figure 5. Left: Corpus level distribution of gaze-directed and competitors object RS. Each contributes 3 datapoints when 3 objects in view. ***Right:*** Bars represent corpus level medians and dots subject-medians

We defined a look as continuous unbroken gaze to an object. The frequency corpus distribution of look durations was extremely skewed such that most look durations were very brief but such that there was also a very long tail of looks that last several to many seconds (**Figure 6**, top-left). Short (< 3s) and long (> 3s) looks were analysed separately. We first determined the stability of RS for the looked to object using the measure of the proportion of look duration for which the looked to object was the visually largest in the image. For short looks (81% of all looks), the distributions of these proportions was bimodal (**Figure 6**, bottom-left): the object looked to either remained the largest in image for nearly the entire look duration or remained not the visually largest for the near entirety of the look to that object. Long looks, on the other hand, made up 19% of all looks. It is the frequency of these so-defined long looks in late infancy that specifically predict self-regulation in later development. Critically, proportion of long looks during which the looked-to object was the visually largest in the field of view was unimodal, with the salience advantage of the looked-to object typically maintained throughout the entire look (**Figure 6**, bottom-right). Finally, longer looks coincided with a larger visual size advantage (M_{long} = 0.55) than shorter looks (M_{short} = 0.48; $\beta = 1.37$; $z = 5.89$; $p < 0.001$). In sum, these results show that infant sustained gaze during toy play, considered a marker of endogenously controlled attention, is strongly associated with targets that are more visible than competitors.

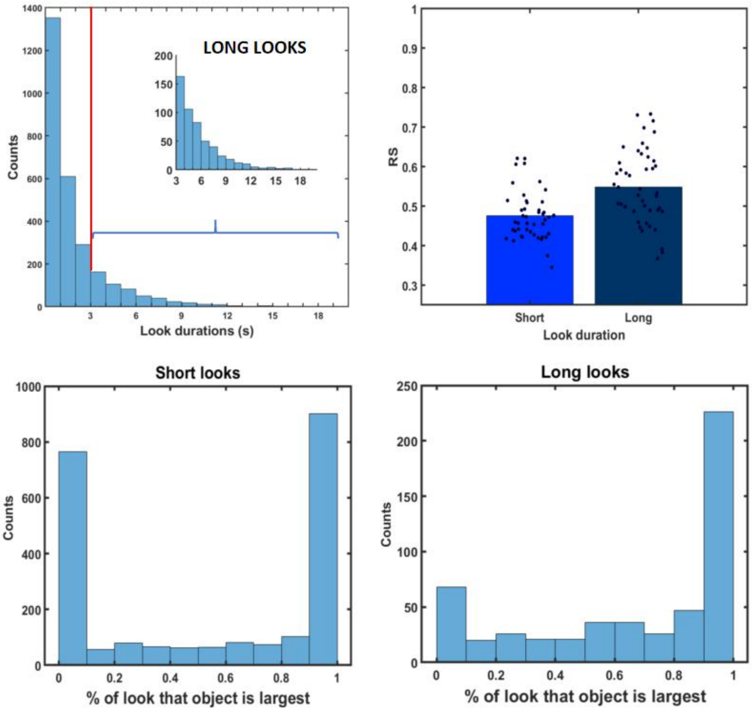


Figure 6. ***Top-left:*** Distribution of corpus look durations. Inset plot shows durations of long looks. **Top-right:** Mean RS during short and long looks. Each dot shows the mean across all look mean RS for each subject. ***Bottom-left:*** For short looks, corpus distribution of the percentage the looked at object is the largest. ***Bottom-right*:** For long looks, corpus distribution of the percentage the looked at object is the largest.

The salience advantage of the looked-to object not only lasted throughout the duration of a look but also began and ended nearly simultaneously with the onset of the look, and did so for both short and long looks (**Figure 7**). We calculated the corpus object RS baseline (whether looked to or not) by randomly selecting an object for each ontask frame and taking a mean of all the selected objects. To determine when in relation to onset and offset of a look, the looked-to target object RS diverged from the baseline of RS for all objects in the corpus we determined the first significant difference in a series of ordered pairwise t-tests from 500 msec before the onset and offset (Allopenna, Magnuson, & Tannenhaus, 1998). By this measure, RS of the looked-to object increased at 100 msec before the look onset and decreased at 200 msec after the offset. In brief, the increased visibility of the looked-to target changed nearly simultaneously with the onset and offset of the look. Many different body movements by the infant as well as external events in the world could alter the visual size and visibility of the potential targets for gaze. In the present context, these include the handling and moving of the objects by the infant and also by the parent who was also present. However, the temporal properties of the increase, as well as the temporal relation of hand actions respect to gaze (see **Figure 8**), suggest that head movements are principally responsible for the increase and decrease RS at the onset and offset of gaze, respectively. As the eyes move to direct the gaze point to the selected object, the head also

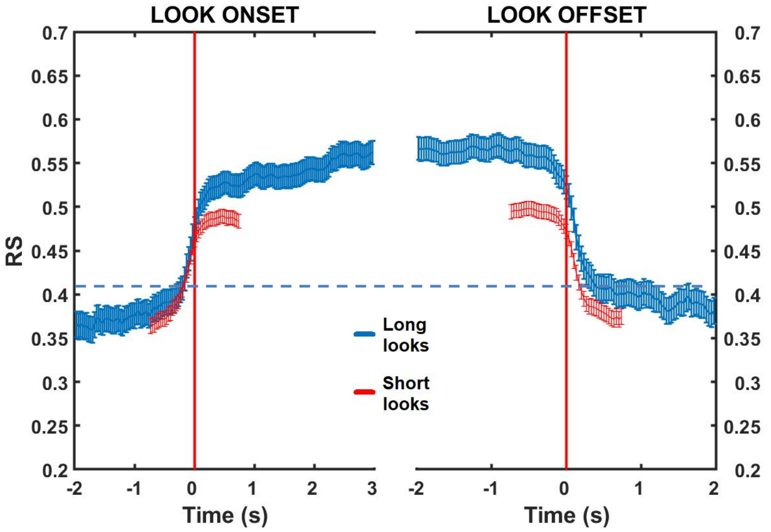


Figure 7. Moment-to-moment RS of the looked at object at the onset and offset. Data shows means and standard error of the aligned RS of long and short looks. Increase and decrease in RS occurs nearly simultaneously with look onset and offset, respectively.

moves to increase and maintain the greater visibility of selected object until the offset of the look. Looking, for infants, involves the whole body which creates a salience advantage over competitors, an advantage that appears particularly crucial to infant's sustained gaze to an object.

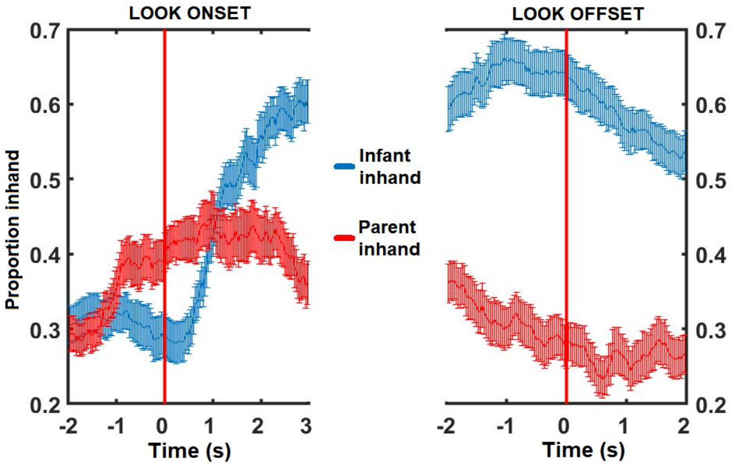


Figure 8. Data indicates if looked object is in infant's or parent's hands at long-look onset and offset. Proportions are calculated at the subject level. Plot describes moment-to-moment mean and error bars. Neither parent's nor infant's hands align with look onset or offset.

6. GENERAL DISCUSSION

A core motivation in vision is to see clearly. Directing gaze to a target selectively increases the visibility of the target over other information because the retinal area around the gaze point captures a higher resolution image than does the periphery (Dowling, 1987; Lee, 1996; Meister & Tessier-Lavigne, 2013). As shown here, body movements that coincide with shifting gaze create *external* visibility gains on the gaze-selected target. Infant sustained attention to a target thus begins with a bottom-up advantage in the projected image to the retina that is enhanced at the retina by the greater acuity around the gaze point. For one-year-old infants, the bottom-up salience in the image, dependent in part on the infant's own bodily behavior, appears central to maintaining a purposeful look. For these infants, goal-directed attention is not a competition between exogenous influences and endogenous control but a collaboration.

6.1. Outside-in pathways to the endogenous control of attention

Laboratory studies, as already mentioned, clearly show that adults can covertly attend through purely internal means without moving eyes or heads to the target and without extrinsic salience advantages resolving the competition among distractors. Here we propose that extending internal explanations of such adult-like abilities to younger ages misses a key element of early attention control. The external properties generated by how we move - in both adults and infants - are amplified in children because of their unique motosensory exploration (Yu & Smith, 2013). Present results do not imply internal inhibition and gains do not occur at early ages, but underscore how show such enhancement and suppression of signals occurs in how the world projects to infants retina's.

The external regulation of stimuli provides an efficient mechanism that could offload the computational requirements to control attention. That is, the incoming stimuli already holds a self-generated bias towards the attended object. Head movements, as eye movements, play a key role in the control of object visibility and stability in the visual field. Motor processes are then both associated with output and input control. Accordingly, important internal pathways to attention control are also pre-motor areas that control eye movements - Frontal Eye Field, Superior Colliculus -, such that the un-executed plan to localize gaze leads to gains on internal signals emanating from that location (Corbetta et al, 1998; Ignashcenkova et al, 2004; Müller et al, 2005). Current research shows these same areas also play a role in head and hand movements towards a goal (Gandhi & Katnani, 2011; Chen, 2006; Stryker & Schiller, 1975; Mark; Walton et al, 2007). The role of planned but not executed head movements in covert attention has not been systematically studied, although findings suggest that planned head movements could play a similar role (Cicchini et al, 2008; Corneil & Munoz, 2014). In this way, infants´ purposeful control of external saliencies through body movements could build –from the outside-in – circuitry that supports internal control. Sustained gaze requires not just a decision as to where to look but moment-to-moment resolution of the conflict to maintain gaze or shift to a different target.

6.2. From low-level decisions to complex executive behavior, through the body

Current views of development no longer picture infants as quiescent and reactive, waiting for cortical development - specifically prefrontal areas - to take place and control behavior (Dehaene-Lambertz & Spelke, 2015; Werchan & Amso, 2017). Prefrontal cortex is active very early to support early cognition. Infancy is characterized by important brain-related structural and functional connectivity changes, the latter being crucial for the development of executive functions (Werchan & Amso, 2017). Critically, executive function development shows multiple sensitivity periods (Thompson & Steinbeis, 2020) and changes in the development of the prefrontal cortex are driven both by processes of neural adaptation and niche construction (Werchan & Amso, 2017). Recent theoretical analysis has proposed that sustained visual attention in infancy plays a key role in the development of the broader class of executive functions mediated by the prefrontal cortex precisely because sustained attention requires these moment-to-moment resolutions of competing pulls on gaze and thus strengthens key feed-forward and feedback loops (Amso & Scerif, 2015; Rosen et al, 2019). The purposeful creation of external saliencies that align with internal goals may play a direct role in the development of more complex forms of cognitive control.

Infants have immature motor systems and spend a lot of time trying to dominate each early acquisition. There is a trade-off between overall posture control and infant's ability to engage in cognitive activities (Berger et al, 2019). A large older literature on infant sustained visual attention (Ruff & Capozzoli, 2003; Ruff, Capozzoli, & Weissberg, 1998; Ruff & Rothbart, 1996) found that long looks by toddlers during object play were associated with a stilled head. The present results suggest that these long looks depend on external salience that is stabilized for the duration of the look, a stabilization that may require controlling head movements. Head stabilization is difficult for infants and toddlers, is characterized by large individual differences, and has been implicated as a marker of difficulties in attentional control in older children (Teicher et al, 1996; Friedman et al, 2005). If, as the present results suggest, the developmental path from sustained gaze in infancy to the self-regulation of attention later in life goes through creating and stabilizing salience advantages for selected targets, individual differences in sensory-motor development may play a role in the development of the internal control of attention.

6.3. An empirical approach to attention

Results here suggest the classification of exogenous and endogenous as independent processes can be misleading to how attention develops, and questions their underlying computations. Controversies on what attention does is by no means new (Allport, 1987; Anderson, 2011; Hommel et al, 2019). Attention is generally used to describe a broad range of abilities which hardly share overlapping explanations (selective attention, visual search, divided attention, selection for action, feature integration, sustained attention, goal-centered attention) (Hommel et al, 2019). What are the computational basis of attention - as of other cognitive processes and cognition itself - is a matter of intense debate. Debates on perception offer interesting insights to our current findings on attention control (Purves et al, 2015). Perception is extensely recognised as a feature extraction process (Hubel & Wiesel, 2005). The retina's 2D display holds strong similarities to how light is projected towards cameras, which intuitively directs pixel-to-pixel thinking of visual perception. The computations underlying perception, however, get more complicated with every new study. More than a dozen morphologically distinguishable ganglion cell types are described throughout the retina (Masri et al, 2019) and relevant stimuli processing occurs as a consequence of temporal properties at the photoreceptor layer (Masland, 2017; Jovanovic & Mamassian, 2020). Microsaccades, sometimes conceived as thermodynamic noise to be corrected, are now associated with more efficient ways of encoding information in a temporal dimension (Lowet et al, 2018). Alternatively, empirical explanations to perception propose that only by understanding the perceptual niche in which organisms evolve and develop will we understand the computations underlying perception (Purves et al, 2014; Purves et al, 2015; Yang & Purves 2003). Projected retinal images do not hold an unambiguous link about the physical source that generates them. This does not imply representations in the brain do not resemble stimuli properties, but that the implemented solutions obey adaptive roles and not the extraction of the 'true' properties of a visual stimuli. Perception has to be good enough to survive. The specificity of perception will then depend on the needs of an organism and also on the current and changing properties of the environment.

A recent theoretical review on attention proposes a philogenetical and comparative approach to recognize which, of all processes defined under the ‘attention’ umbrella, are necessary to understand its core computational properties (Hommel et al, 2019). According to these authors, selection is the key behavior from which attention derives. Behaviors of approach and withdrawal appear in very primitive animals: when two stimuli are present, approach behavior cannot be solved by the vectorial sum of stimuli but requires a winner-take-all strategy (see **Figure 9**). Conserved brain regions across vertebrates participate in body orientation and the regulation of approach-withdrawal behaviors (tectum in fish reptiles and birds, and superior colliculus in mammals).

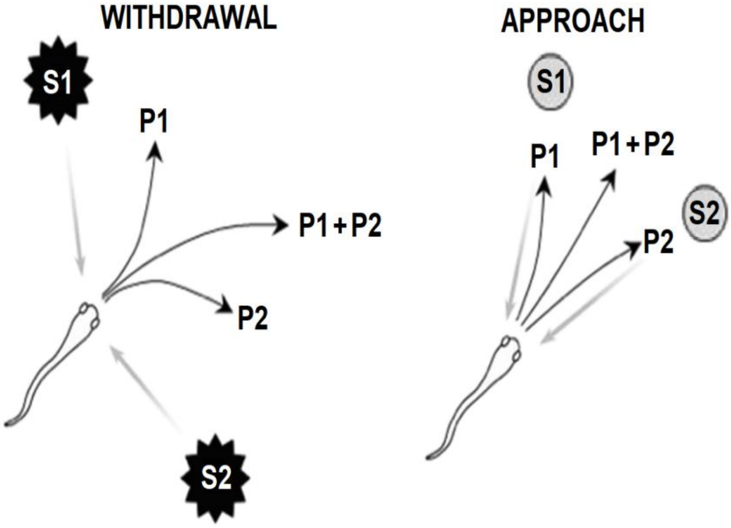


Figure 9. Selection for withdrawal (left) and approach (right). Images show the consequences of making linear stimuli integration for the selection of a response. Such method would result in appropriate behaviors for withdrawal but not approach. Adapted from Hommel et al (2019).

Studies in larval Zebrafish (ZF) offer an interesting example of how evolution implemented stimuli selection (Förster et al, 2020). Size is an ecologically relevant feature that informs about behaviorally relevant stimuli properties (as ZF feeds on small organisms). Prey approach consists of a series of ordered steps that starts with a small projected retinal image that gets displaced (from nasal to temporal retina) and becomes larger as fish coordinates their swimming to capture it. This is realized by a sophisticated circuitry in which retinoganglion cells in different retinal positions (nasal-temporal) send projections into separate layers of the tectum - which as a result show selective neuronal response to stimuli of different sizes - and to motor areas to successfully orient the body towards the target. Interestingly, although often considered a reactive animal, a study in a virtual environment in ZF shows approach behavior requires the ongoing visibility of the prey, which immediately disappears if the stimulus also does.

How do ZF select their prey? Larval ZF knowledge of food identity is distributed throughout their whole body, both in the perceptual and motor system structure and connectivity. Millions of years of evolution selected fish bodies that detect the relevant information in the environment to successfully approach their prey. Such success is the results of a meticulous tuning of stmuli projections towards separate areas in the retina and tectum, and to motor areas that coordinate appropriate orientation responses. In words of the authors: “the local statistics of the sensory environment, which changes dynamically as the animal interacts with the outside world, shape the topographic specializations of higher-order sensory and sensorimotor circuitry” (Förster et al, 2020). More specifically, local stastistics that have meaning for species survival will be key players in body and behavior evolution. As an example, looming, on top of size (and which relates to size) is another strong relevant feature that reorients attention, is informative about the environment and has also been described in other animal models (Mysore et al, 2011).

Pezzulo & Nolfi (2019) elegantly present two distinct ways to solving a perceptually challenging problem. One solution is to cognitively enrich the organism to internally represent past information of the environment and using it in appropriate ways. An alternative solution, as already mentioned in ZF, occurs through behavior, and is referred to as “sensorimotor enrichment”. The rationale of the latter is that the projection of stimuli towards an agent is shaped by the agent’s own action (which play a role in the next action selection), and that action selection and how it affects ongoing stimuli can be exploited to solve the perceptual problem without an internal representation of the environment (Pezzulo & Nolfi, 2019). Our results are consistent with a sensorimotor enrichment of how infants solve selection problems. Selective sustained attention onset coincides with a reorganization of the body-object relation which generates larger RS image projections to the system - and therefore decreased competition in the transduced internal signals - which promotes further selection.

Humans are adapted to a very different ecological niche than ZF. As a consequence, throughout the millions of years of our separate evolutionary history, different events of body-stimuli coordination have shaped very different visual circuitry for selection. Object size - rather than being informative of prey - indicates relevant information for sensorimotor behavior (Castelhano & Krzyś, 2020). The projection of larger retina images - which in itself is a strong exogenous cue - also informs about proximity, and is therefore relevant for action selection. Studies in primates show the tuning of neuron in premotor areas to events happening specifically in the animal's peripersonal space (Caggiano et al, 2009). Moreover, close and proximal objects generate different projected image changes when displaced (see **Figure 10).** Although our results do not disambiguate between size and proximity effects, regulating both kinds of external cues might be contributing to our results and to selective and sustained attention.

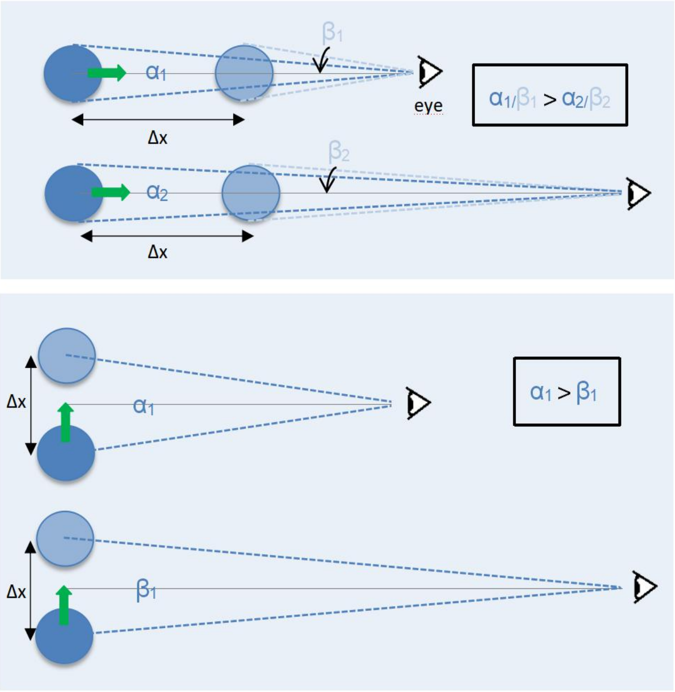


Figure 10. Proximal objects not only project larger images in retina but also cause distinct projected displacements when eye or objects move. ***Above:*** two examples of object moving the same distance towards the eye. The projected angle towards the observer shows bigger changes in closer objects (top). ***Below:*** two example of objects moving without changing its distance to perceiver. The same displacement, generates larger displacement angles in close objects (top)

6.4. Developmental sensorimotor scaffolding

Infancy is frequently described as a rapidly changing phase characterized by relevant motor milestones. Perfection requires serious practice. Infants spend a lot of their time rehearsing their achievements and are constantly generalizing their new abilities to new targets and combining them for more complex behaviors. These changes, however, are not mere motor skills, but reorganization of perceptual and motor mutually influentiable events that are relevant for every achievement - walking, reaching, approaching, etc. (Anderson et al, 2013). As mentioned before, our results suggest these environment-body-brain feedforward and feedback loop organization might also extend to attention and more executive forms of behavior.

Human infant's plastic and big brains have been related to the immense repertoire of actions they can display (Heldstab et al, 2016; Heldstab et al, 2020). Our hands are highly sophisticated machines that can be combined with body and head movements with enormous degrees of freedom - we could say the same for oral facial and vocal behavior. More complex behaviors - executive behaviors - generally require the use and coordination of distinct muscles and body parts with particular spatial and temporal profiles. Every movement will also receive a very different combination of perceptual feedback. That is, we can feel we are walking forward by the propioceptive sense, through the visual flow, by the changing auditory landscape, by the feel of the air in our skin, by the vestibular information on the changes caused by displacement. These phylogenetically and ontogenetically fundamental properties tune and select bodies fit for survival. As an example, three year old infants are more probable to display stepping behavior as a result of terrestrial (forward) than rotating or static optic-flow (Barbu-Roth et al, 2009). We are deeply embedded in the environment around us in ways that are necessary and fundamental to what we do and how we do it, and for development.

What occurs in early development is a drastic change in body-body and body-world coupling (Smith, 2013). Every new acquisition implies precise and unique motosensory coordination that spans different spatial and temporal scales. The reduced infant workspace, already suggested to be important for object recognition and word learning, scaffolds attention control. As proposed here, attention control emerges then through the coordination of appropriate sensorimotor solutions for selecting relevant outcomes (which imply both appropriate stimuli and actions). It is the materiality of these sensorimotor processes - i) the selection of appropriate muscles, ii) the self-projected spatial and temporal consequences of behavior and iii) how it leads to failed, partial or complete positive outcomes - that shape internal circuits and sensorimotor coordination in ways that scaffold more abstract, complex forms of attention and cognition. The term materiality has been used to underscore the physical nature underlying apparently abstract changes in language acquisition (Clark, 2006). We bring this term here to underscore the relevance of understanding the fine-grained, dynamic study of behavior - its physical embodied nature and historicity (Thelen & Smith, 1996; Gomez-Marin & Ghazanfar, 2019) - for a more integrated and coherent model of developmental change.

6.5. Limitations

Current results should be generalized to other settings, activities and demographic populations. Although unrestrained, the lab is still a structured task occurring in an unnatural environment and our data represent a small fraction of children´s daily life. More ecological forms of perceptual feedback and noisy environments, in home studies, could provide with other input properties and regularities that might be relevant for how attention control develops. In our settings, selection was restricted to the three toys infants were presented with. In their own environments, attention might be attracted towards very different kind of objects and events. We believe, however, that the strong coupling of behavior with input visual properties shown here in a three minute play is just a glimpse of powerful mechanisms that operate on a daily basis shaping and selecting infants ongoing development, and that lead to more complex forms of behaviors across different cultural and geographic contexts.

6.6. Future directions

The fine-grained dynamics of body and head coordination are key components which will shed further light in how infants organize more complex behaviors. Computer vision and motion tracking both present reasonable solutions to track body movement in space and the head´s yaw, pitch and roll as they play with objects. Such detailed analyses would provide data on what are - from the infant´s perspective - predicatble and unpredictable input changes taking into account the consequences of their own actions. Understanding how expectations based on previous selection history (previous recent experience), saliency and goal-directed behaviors are related throughout infancy in real-time behavior is necessary to overcome current conceptual inconsistencies (Macaluso & Doricchi, 2013; Hommel et al, 2019). Moreover, the different ways in which children solve head-hand-eye coordination could describe alternative pathways to stabilizing perception and identify ways in which impairments in such coordination might relate to attention or executive disorders.

The use of head cameras at home to analyse children´s environments would let us build statistics of the children 3D environments. As discussed, size is a property that correlates with relevant spatial information like proximity. Studying infants everyday visual regularities and relating it to sensorimotor strategies and attentional outcomes will shed light on what might be pervasive and relevant motosensory properties of how attention develops. Moreover, carefully planned laboratory studies could help disambiguate how correlated variables - such as size and proximity - might differently contribute to attention control.

Finally, parental scaffolding has been pointed out as relevant to infants self-regulation (Yu & Smith, 2016; Suarez-Rivera et al, 2019). Our results show parent's behavior do not create the unique visual properties. However, parents pervasively act on children´s peripersonal space, creating and influencing infant´s visual workspace. Current result suggest that the way parent´s interact with children and relate to infant´s self-created motosensory contingencies might be at the basis of how this scaffolding takes place.

7. CONCLUSION

Having a body, and more critically a body that moves creates rules that depend on the relation between the organism and the world - i.e: views are always asymmetric (close=big), movement of objects in retina depend on how we move and how close they are to us, etc. The relevance of measuring body-world coupling has been extensely documented as important for information seeking (sensory systems create information out of movements that was not there on stimulus before), for learning (relevance of structure of data created) and cognitive processing (tools become part of system to offload cognition). Attention control, however, is traditionally conceived as an internal process emerging from internal memory and goals. We show, then, yet another reason to take a very close look at body-world interrelations. We suggest attention in infancy is an online process closely related to ongoing motosensory behavior. This way, visual input (which depends both of physical laws and how stimuli project to infant's sensory systems) and motor output (organized by brain but also by sensory information) are connected in time, mutually shape each other to select and sustain attention and might underly the reorganization of behavior (or 'sensorimotor scaffolding') into more complex forms of endogenous and executive attentional control.

8. REFERENCES

Abbott, W. W., Harston, J. A., & Faisal, A. A. (2020). Linear Embodied Saliency: a Model of Full-Body Kinematics-based Visual Attention. bioRxiv.

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. Psychological Science, 14(5), 427-432.

Ahissar, E., & Assa, E. (2016). Perception as a closed-loop convergence process. Elife, 5, e12830.

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.

Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. Perspectives on perception and action, 15, 395-419.

Amso, D., & Scerif, G. (2015). The attentive brain: insights from developmental cognitive neuroscience. Nature Reviews Neuroscience, 16(10), 606-619.

Anderson, B. (2011). There is no such thing as attention. Frontiers in psychology, 2, 246.

Anderson, D. I., Campos, J. J., Witherington, D. C., Dahl, A., Rivera, M., He, M., ... & Barbu-Roth, M. (2013). The role of locomotion in psychological development. Frontiers in psychology, 4, 440.

Attinger, A., Wang, B., & Keller, G. B. (2017). Visuomotor coupling shapes the functional development of mouse visual cortex. Cell, 169(7), 1291-1302.

Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. (2016, September). 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). IEEE.

Barbu‐Roth, M., Anderson, D. I., Desprès, A., Provasi, J., Cabrol, D., & Campos, J. J. (2009). Neonatal stepping in relation to terrestrial optic flow. Child development, 80(1), 8-14.

Barsalou, L. W. (2008). Grounded cognition. Annu. Rev. Psychol., 59, 617-645.

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

Berger, S. E., Harbourne, R. T., Arman, F., & Sonsini, J. (2019). Balancing act (ion): Attentional and postural control strategies predict extent of infants' perseveration in a sitting and reaching task. Cognitive Development, 50, 13-21.

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.

Bowling, J. T., Friston, K. J., & Hopfinger, J. B. (2020). Top‐down versus bottom‐up attention differentially modulate frontal–parietal connectivity. Human brain mapping, 41(4), 928-942.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science, 315(5820), 1860-1862.

Buzsáki, G. (2019). The brain from inside out. Oxford University Press.

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.

Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., & Casile, A. (2009). Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. Science, 324(5925), 403-406.

Caputi, A. A. (2004). Contributions of electric fish to the understanding sensory processing by reafferent systems. Journal of Physiology-Paris, 98(1-3), 81-97.

Carrasco, M. (2011). Visual attention: The past 25 years. Vision research, 51(13), 1484-1525.

Carretié, L., Ríos, M., Periáñez, J. A., Kessel, D., & Álvarez-Linera, J. (2012). The role of low and high spatial frequencies in exogenous attention to biologically salient stimuli. PloS one, 7(5), e37082.

Carretié, L. (2014). Exogenous (automatic) attention to emotional stimuli: a review. Cognitive, Affective, & Behavioral Neuroscience, 14(4), 1228-1258.

Castelhano, M. S., & Krzyś, K. (2020). Rethinking Space: A Review of Perception, Attention, and Memory in Scene Processing. Annual Review of Vision Science, 6.

Chica, A. B., Bartolomeo, P., & Lupiáñez, J. (2013). Two cognitive and neural systems for endogenous and exogenous spatial attention. Behavioural brain research, 237, 107-123.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development?. Trends in cognitive sciences, 9(3), 104-110.

Chen, L. L. (2006). Head movements evoked by electrical stimulation in the frontal eye field of the monkey: evidence for independent eye and head control. Journal of neurophysiology, 95(6), 3528-3542.

Cicchini, G. M., Valsecchi, M., & De'Sperati, C. (2008). Head movements modulate visual responsiveness in the absence of gaze shifts. Neuroreport, 19(8), 831-834.

Clark, A. (2006). Language, embodiment, and the cognitive niche. Trends in cognitive sciences, 10(8), 370-374.

Clayton, M. S., Yeung, N., & Kadosh, R. C. (2015). The roles of cortical oscillations in sustained attention. Trends in cognitive sciences, 19(4), 188-195.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711), 20160055.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. Child Development, 869-879.

Colombo, J. (2001). The development of visual attention in infancy. Annual review of psychology, 52(1), 337-367.

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., ... & Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. Neuron, 21(4), 761-773.

Corneil, B. D., Olivier, E., & Munoz, D. P. (2004). Visual responses on neck muscles reveal selective gating that prevents express saccades. Neuron, 42(5), 831-841.

Corneil, B. D., & Munoz, D. P. (2014). Overt responses during covert orienting. Neuron, 82(6), 1230-1243

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. Nature Reviews Neuroscience, 9(8), 587-600.

Crawford, J. D., Ceylan, M. Z., Klier, E. M., & Guitton, D. (1999). Three-dimensional eye-head coordination during gaze saccades in the primate. Journal of Neurophysiology, 81(4), 1760-1782.

Dehaene-Lambertz, G., & Spelke, E. S. (2015). The infancy of the human brain. Neuron, 88(1), 93-109.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual review of neuroscience, 18(1), 193-222.

Dobs, K., Kell, A., Palmer, I., Cohen, M., & Kanwisher, N. (2019). Why Are Face and Object Processing Segregated in the Human Brain? Testing Computational Hypotheses with Deep Convolutional Neural Networks. Oral presentation at Cognitive Computational Neuroscience Conference, Berlin, Germany.

Dobs, K., Kell, A. J., Martinez, J., Cohen, M., & Kanwisher, N. (2020). Using task-optimized neural networks to understand why brains have specialized processing for faces. Journal of Vision, 20(11), 660-660.

Dowling, J. E. (1987). The retina: an approachable part of the brain. Harvard University Press.

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. Journal of vision, 8(3), 3-3.

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. Trends in cognitive sciences, 17(5), 202-209.

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.

Fodor, J.A. (1975). The language of thought. Harvard University Press.

Förster, D., Helmbrecht, T. O., Mearns, D. S., Jordan, L., Mokayes, N., & Baier, H. (2020). Retinotectal circuitry of larval zebrafish is adapted to detection and pursuit of prey. Elife, 9, e58596.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010, March). Head-mounted eye-tracking of infants' natural interactions: a new method. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (pp. 21-27).

Gaspelin, N., & Luck, S. J. (2018). The role of inhibition in avoiding distraction by salient stimuli. Trends in cognitive sciences, 22(1), 79-92.

Gandhi, N. J., & Katnani, H. A. (2011). Motor functions of the superior colliculus. Annual review of neuroscience, 34, 205-231.

Gomez-Marin, A., & Ghazanfar, A. A. (2019). The life of behavior. Neuron, 104(1), 25-36.

Gottlieb, J. (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron, 53(1), 9-16.

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.

Heldstab, S. A., Kosonen, Z. K., Koski, S. E., Burkart, J. M., Van Schaik, C. P., & Isler, K. (2016). Manipulation complexity in primates coevolved with brain size and terrestriality. Scientific reports, 6(1), 1-9.

Heldstab, S. A., Isler, K., Schuppli, C., & van Schaik, C. P. (2020). When ontogeny recapitulates phylogeny: Fixed neurodevelopmental sequence of manipulative skills among primates. Science advances, 6(30), eabb4685.

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.

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. Attention, Perception, & Psychophysics, 81(7), 2288-2303.

Hubel, D. H., and Wiesel, T. (2005). Brain and Visual Perception. A story of a 25-year Collaboration. New York: Oxford University Press.

Iarocci, G., Enns, J. T., Randolph, B., & Burack, J. A. (2009). The modulation of visual orienting reflexes across the lifespan. Developmental Science, 12(5), 715-724.

Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. Nature neuroscience, 7(1), 56-64.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Frontiers in human neuroscience, 4, 186.

Jovanovic, L., & Mamassian, P. (2020). Events are perceived earlier in peripheral vision. Current Biology, 30(21), R1299-R1300.

Kannass, K. N., Oakes, L. M., & Shaddy, D. J. (2006). A longitudinal investigation of the development of attention and distractibility. Journal of Cognition and Development, 7(3), 381-409.

Kleinfeld D, Ahissar E, Diamond ME. 2006. Active sensation: Insights from the rodent vibrissa sensorimotor system. Current Opinion in Neurobiology 16:435–444. doi:

Lee, B. B. (1996). Receptive field structure in the primate retina. Vision research, 36(5), 631-644.

Lowet, E., Gomes, B., Srinivasan, K., Zhou, H., Schafer, R. J., & Desimone, R. (2018). Enhanced neural processing by covert attention only during microsaccades directed toward the attended stimulus. Neuron, 99(1), 207-214.

Luo, C., & Franchak, J. M. (2020). Head and body structure infants’ visual experiences during mobile, naturalistic play. Plos one, 15(11), e0242009.

Macaluso, E., & Doricchi, F. (2013). Attention and predictions: control of spatial attention beyond the endogenous-exogenous dichotomy. Frontiers in human neuroscience, 7, 685.

Masland, R. H. (2017). Vision: two Speeds in the Retina. Current Biology, 27(8), R303-R305.

Masri, R. A., Percival, K. A., Koizumi, A., Martin, P. R., & Grünert, U. (2019). Survey of retinal ganglion cell morphology in marmoset. Journal of Comparative Neurology, 527(1), 236-258

Mearns, D. S., Donovan, J. C., Fernandes, A. M., Semmelhack, J. L., & Baier, H. (2020). Deconstructing hunting behavior reveals a tightly coupled stimulus-response loop. Current Biology, 30(1), 54-69.

Meister, M., & Tessier-Lavigne, M. (2013). Low-level visual processing: the retina. Principles of neural science, 5, 577-601.

Merleau-Ponty, M. (1982). Phenomenology of perception. Routledge.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science, 229(4715), 782-784.

Müller, J. R., Philiastides, M. G., & Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. Proceedings of the National Academy of Sciences, 102(3), 524-529.

Mysore, S. P., Asadollahi, A., & Knudsen, E. I. (2011). Signaling of the strongest stimulus in the owl optic tectum. Journal of Neuroscience, 31(14), 5186-5196

Orhan, E., Gupta, V., & Lake, B. M. (2020). Self-supervised learning through the eyes of a child. Advances in Neural Information Processing Systems, 33.

Niell, C. M., & Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. Neuron, 65(4), 472-479.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. Psychonomic bulletin & review, 21(1), 178-185.

Pezzulo, G., & Nolfi, S. (2019). Making the Environment an Informative Place: A Conceptual Analysis of Epistemic Policies and Sensorimotor Coordination. Entropy, 21(4), 350.

Port, R. F., & Van Gelder, T. (Eds.). (1995). Mind as motion: Explorations in the dynamics of cognition. MIT press.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. Journal of neuroscience, 4(7), 1863-1874.

Powers, W. T. (1973). Feedback: Beyond Behaviorism: Stimulus-response laws are wholly predictable within a control-system model of behavioral organization. Science, 179(4071), 351-356.

Purves, D., Monson, B. B., Sundararajan, J., & Wojtach, W. T. (2014). How biological vision succeeds in the physical world. Proceedings of the National Academy of Sciences, 111(13), 4750-4755.

Purves, D., Morgenstern, Y., & Wojtach, W. T. (2015). Perception and reality: why a wholly empirical paradigm is needed to understand vision. Frontiers in systems neuroscience, 9, 156.

Reck, S. G., & Hund, A. M. (2011). Sustained attention and age predict inhibitory control during early childhood. Journal of experimental child psychology, 108(3), 504-512.

Ristic, J., & Kingstone, A. (2009). Rethinking attentional development: reflexive and volitional orienting in children and adults. Developmental science, 12(2), 289-296.

Ristic, J., & Enns, J. T. (2015a). The changing face of attentional development. Current Directions in Psychological Science, 24(1), 24-31.

Ristic, J., & Enns, J. T. (2015b). Attentional development. Handbook of child psychology and developmental science, 1-45.

Rivière, J., & Falaise, A. (2011). What comes first? How selective attentional processes regulate the activation of a motor routine in a manual search task. Developmental psychology, 47(4), 969.

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.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. Nature neuroscience, 19(1), 165-171.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. Developmental psychology, 26(1), 85.

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. Developmental psychology, 39(5), 877.

Ruff, H. A., Capozzoli, M., & Weissberg, R. (1998). Age, individuality, and context as factors in sustained visual attention during the preschool years. Developmental psychology, 34(3), 454.

Ruff, H. A., & Rothbart, M. K. (2001). Attention in early development: Themes and variations. Oxford University Press.

Scerif, G. (2010). Attention trajectories, mechanisms and outcomes: at the interface between developing cognition and environment. Developmental Science, 13(6), 805-812.

Schneider, D. M. (2020). Reflections of action in sensory cortex. Current Opinion in Neurobiology, 64, 53-59.

Sensoy, Ö., Culham, J. C., & Schwarzer, G. (2020). Do infants show knowledge of the familiar size of everyday objects?. Journal of Experimental Child Psychology, 195, 104848.

Simony, E, Saraf-Sinik I, Golomb D, Ahissar E. 2008. Sensation-targeted motor control: Every spike counts? Focus on: "whisker movements evoked by stimulation of single motor neurons in the facial nucleus of the rat". Journal of Neurophysiology 99:2757–2759

Slone, L. K., Smith, L. B., & Yu, C. (2019). Self‐generated variability in object images predicts vocabulary growth. Developmental science, 22(6), e12816.

Smith, L. B., & Sheya, A. (2010). Is cognition enough to explain cognitive development?. Topics in Cognitive Science, 2(4), 725-735.

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

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.

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. American Psychologist, 68(8), 618.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. Trends in cognitive sciences, 22(4), 325-336.

Stryker, M. P., & Schiller, P. H. (1975). Eye and head movements evoked by electrical stimulation of monkey superior colliculus. Experimental Brain Research, 23(1), 103-112.

Suanda, S. H., Barnhart, M., Smith, L. B., & Yu, C. (2019). The signal in the noise: the visual ecology of parents' object naming. Infancy, 24(3), 455-476.

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.

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

Thelen, E., & Smith, L. B. (1996). A dynamic systems approach to the development of cognition and action. MIT press.

Thompson, A., & Steinbeis, N. (2020). Sensitive periods in executive function development. Current Opinion in Behavioral Sciences, 36, 98-105.

Tomlinson, R. D., & Bahra, P. S. (1986). Combined eye-head gaze shifts in the primate. I. Metrics. Journal of Neurophysiology, 56(6), 1542-1557.

Teicher, M. H., Ito, Y., Glod, C. A., & Barber, N. I. (1996). Objective measurement of hyperactivity and attentional problems in ADHD. Journal of the American Academy of Child & Adolescent Psychiatry, 35(3), 334-342

van Moorselaar, D., & Slagter, H. A. (2020). Inhibition in selective attention. Annals of the New York Academy of Sciences, 1464(1), 204.

Vernon, D. (2014). Artificial cognitive systems: A primer. MIT Press.

Von Hofsten, C. (2007). Action in development. Developmental science, 10(1), 54-60.

Walton, M. M., Bechara, B., & Gandhi, N. J. (2007). Role of the primate superior colliculus in the control of head movements. Journal of neurophysiology, 98(4), 2022-2037.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. Journal of Educational Psychology, 102(1), 43.

Werchan, D. M., & Amso, D. (2017). A novel ecological account of prefrontal cortex functional development. Psychological review, 124(6), 72

Wright, J. C., & Vlietstra, A. G. (1975). The development of selective attention: From perceptual exploration to logical search. In H. Reese (Ed.), Advances in child development and behavior, 10,195-239. New York: Academic

Yang, Z., & Purves, D. (2003). A statistical explanation of visual space. Nature neuroscience, 6(6), 632-640.

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.

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

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PloSone, 8(11), e79659.

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

Yu, C., & Smith, L. B. (2017). Multiple sensory‐motor pathways lead to coordinated visual attention. Cognitive science, 41, 5-31.

9. SUPPLEMENTARY INFORMATION

9.1. Number of objects in view

Although infants play with three objects, they tend to do it one at a time. Despite using wide-angle cameras, this type of play together with play dynamics - objects coming in and out of sight, being handled, etc. - could provoke large variability in the number of objects in view. Results show, however, three objects are generally in view across children (see **Figure S1**).

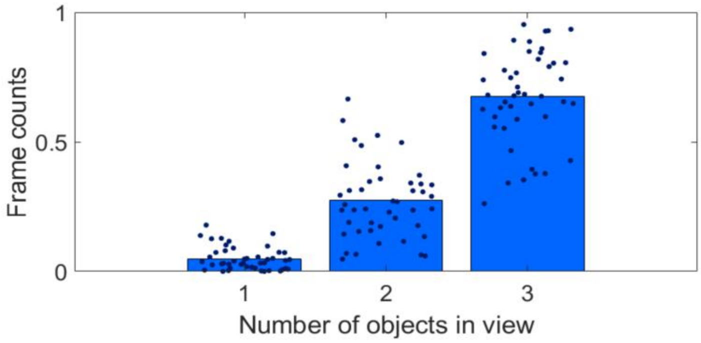


Figure S1. Corpus level number of objects in view.

The number of objects in view also affects the probability of selecting an object to look at. Choosing the largest object is inevitable when one object in view, very probable when two objects in view and somewhat less probable when all objects are in view. An analysis of the probability of looking to the largest object taking this into account show that infants look at the largest object no matter the number of objects in view (see **Figure S.2**).

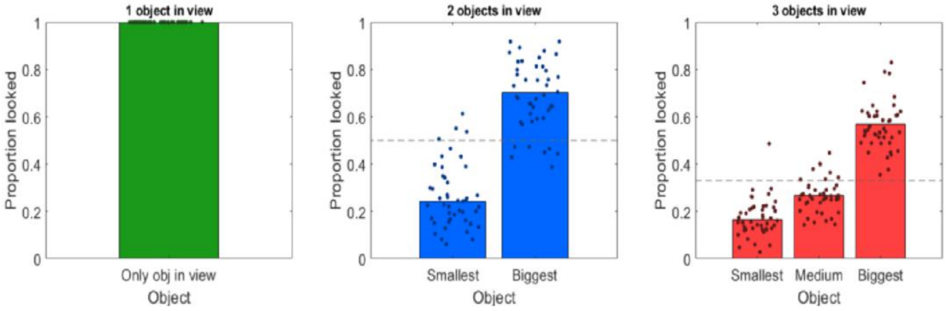


Figure S2. Proportion of objects looked at according to the relative size of objects in view.

9.2. IS and RS

Figure 4 shows a high correlation between IS and RS for objects in view. However, the relative size and image size of objects might vary considerably (as **Figure 4** shows), being both relevant for attention. RS analyses were also done for IS without significant differences in results (**Figure S.3**).

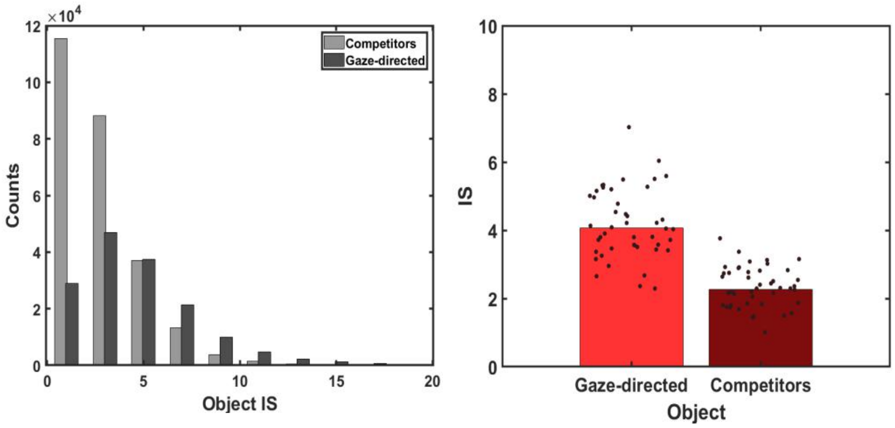


Figure S3. ***Left:*** Corpus level distribution of gaze-directed and competitors object IS. Each contributes 3 datapoints when 3 objects in view. ***Right:*** Subject level means

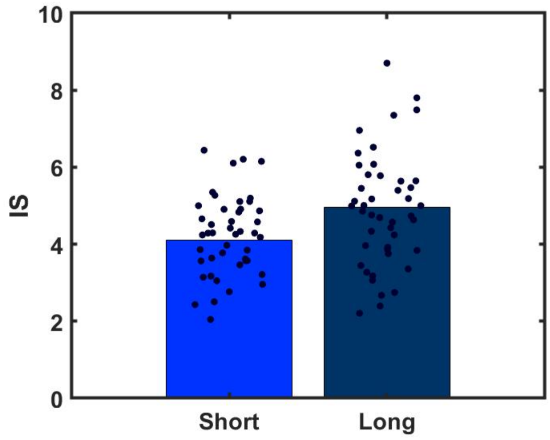


Figure S4. Bars show corpus level means for each look duration. Dots represent each subject’s mean.

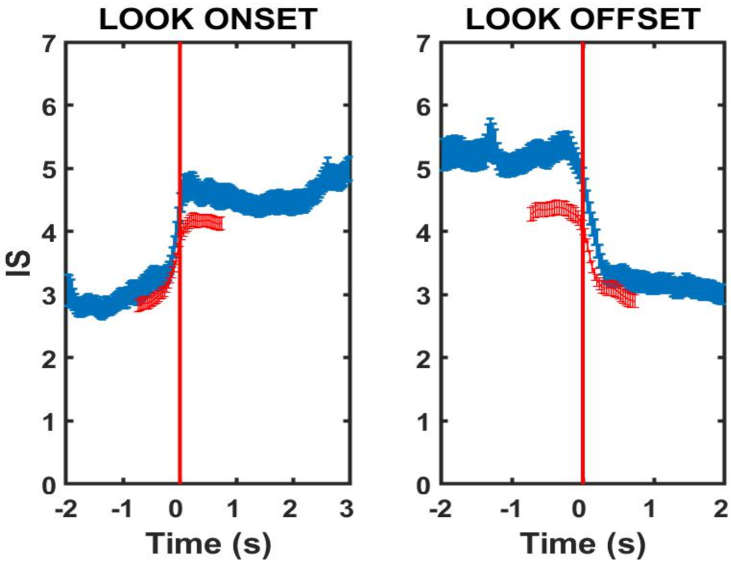
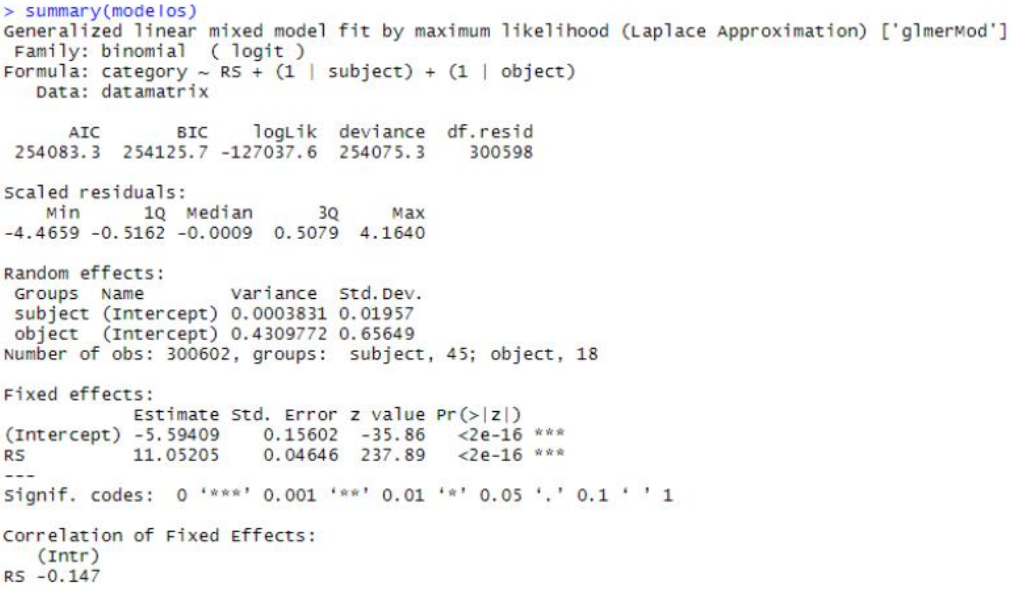


Figure S5. Plot shows the IS of the looked at object at long look onset and offset. Each datapoint is calculated by aligning all long looks and taking the moment-to-moment mean and standard error.

9.3. Logistic mixed modes

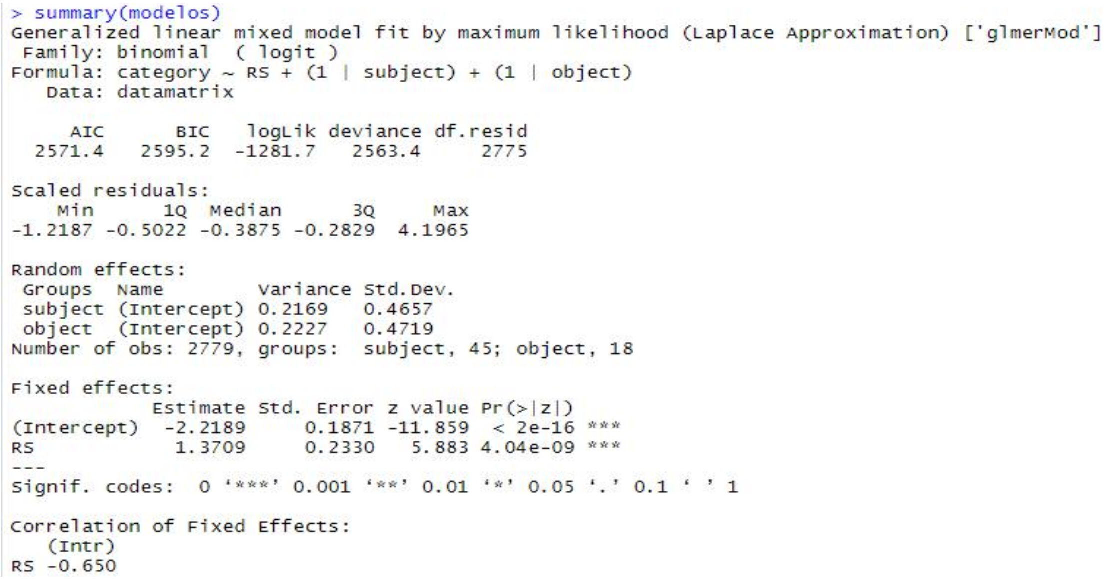
We submitted binary scores/values to a logistic mixed effects model using lme4 package in R. The model predicted three different likelihoods - largest object VS sum of competitors, gaze-directed object VS competitors, short looks VS long looks - from RS as fixed effect. Random intercepts were specified for individual infants and for the specific target objects. Below are R summaries of the three models.

MODEL 1. Largest VS Sum of competitors.



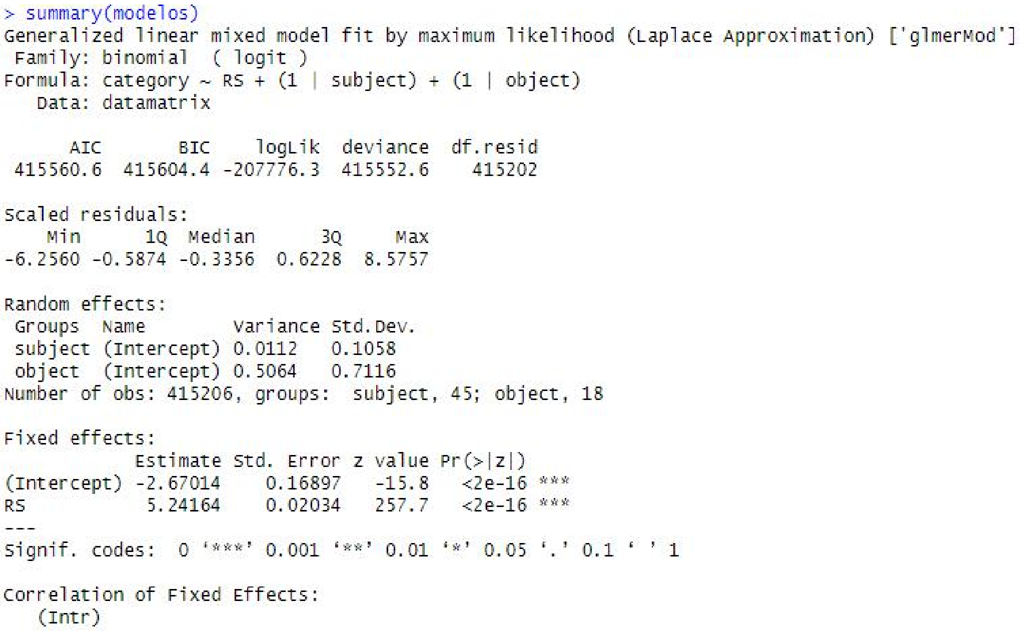
```
> summary(modelos)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: category ~ RS + (1 | subject) + (1 | object)
   Data: datamatrix

      AIC       BIC    logLik  deviance  df.resid
 254083.3  254125.7 -127037.6  254075.3    300598

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4659 -0.5162 -0.0009  0.5079  4.1640

Random effects:
 Groups  Name        Variance  Std.Dev.
 subject (Intercept) 0.0003831 0.01957
 object  (Intercept) 0.4309772 0.65649
Number of obs: 300602, groups:  subject, 45; object, 18

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.59409    0.15602  -35.86   <2e-16 ***
RS          11.05205    0.04646  237.89   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
   (Intr)
RS -0.147
```

MODEL 2. Gaze-directed VS Competitors



```
> summary(modelos)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: category ~ RS + (1 | subject) + (1 | object)
   Data: datamatrix

     AIC      BIC   logLik deviance df.resid
  2571.4   2595.2  -1281.7   2563.4     2775

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.2187 -0.5022 -0.3875 -0.2829  4.1965

Random effects:
 Groups  Name        Variance Std.Dev.
 subject (Intercept) 0.2169   0.4657
 object  (Intercept) 0.2227   0.4719
Number of obs: 2779, groups:  subject, 45; object, 18

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.2189     0.1871 -11.859  < 2e-16 ***
RS            1.3709     0.2330   5.883 4.04e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
   (Intr)
RS -0.650
```

MODEL 3. Short VS Long



```
> summary(modelos)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: category ~ RS + (1 | subject) + (1 | object)
   Data: datamatrix

      AIC       BIC    logLik  deviance  df.resid
 415560.6  415604.4 -207776.3  415552.6    415202

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.2560 -0.5874 -0.3356  0.6228  8.5757

Random effects:
 Groups  Name        Variance Std.Dev.
 subject (Intercept) 0.0112   0.1058
 object  (Intercept) 0.5064   0.7116
Number of obs: 415206, groups:  subject, 45; object, 18

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.67014    0.16897   -15.8   <2e-16 ***
RS           5.24164    0.02034   257.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
   (Intr)
```