UNIVERSITY OF CALIFORNIA
RIVERSIDE

# Developmental Changes in Prioritization of Visual Attention to Features and Goal-Directed Manual Actions

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Kellan Kadooka

September 2022

Dissertation Committee:

Dr. John M. Franchak, Chairperson
Dr. Chandra Reynolds
Dr. Rebekah Richert

The Dissertation of Kellan Kadooka is approved:

Committee Chairperson

University of California, Riverside

# Acknowledgments

I would like to express my gratitude to my advisor, John Franchak. This would not be possible without your wisdom, support, patience, and pragmatism. Thank you.

Dedicated to my family and friends who supported me.

To my parents: Words cannot express my gratitude for your encouragement, unconditional love, and all the lessons you've taught me along the way.

To my friends: Thank you for pushing me to the finish line, especially Juan and Jaz.

To Koa: You are the best boy and we did it!

# ABSTRACT OF THE DISSERTATION

## Developmental Changes in Prioritization of Visual Attention to Features and Goal-Directed Manual Actions

by

Kellan Kadooka

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, September 2022
Dr. John M. Franchak, Chairperson

Across three studies, this dissertation evaluates how different influences on visual attention in dynamic scenes develop over infancy and childhood. Features of a visual scene are typically considered either bottom-up (visually salient factors) or top-down (meaningful semantic factors) influences. The first study (Chapter 2) tested whether influences of visual attention developmentally shift from bottom-up to top-down. Attention to visually salient locations and faces was measured across a wide age range and a wide set of video stimuli to operationalize bottom-up and top-down influences. Results indicate that attention does not shift from primarily bottom-up to top-down; attention to salient areas and faces were similar across ages for most stimulus videos. In considering the dynamic nature of attention, the second study (Chapter 3) measured developmental change in attention to hand and hand-object actions. By measuring attention in ways that are sensitive to movement of features in the scene over time, I found age-related increases in looking to hands and hand-object actions. Age differences suggest attention develops by more mature observers increasingly prioritizing meaningful information from moment to moment. The final study (Chapter 4) investigated the role of comprehension of actions while 4-year-olds viewed a video of a person performing a sequence of goal-directed movements. I manipulated children's prior visual experience with a novel action to assess whether comprehension of manual action changes visual attention, hypothesizing that previous experience viewing the action sequence would bolster children's comprehension. However, results showed no differences in attention to hand-object actions regardless of prior visual experience. Influences of visual attention are complex, but refining our perspective and methodological approach is important for characterizing the development of visual attention.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This dissertation examines the development of visual attention in the context of viewing dynamic visual scenes. Visual attention is cognitive process involved in the selection of visual information from the environment (Oakes & Amso, 2018). I will begin by reviewing two broad categories of attentional influences that are defined by theories of visual attention: low-level *bottom-up factors* and semantically-related *top-down factors*. Some accounts of developmental change in visual attention suggest that there is an age-related shift from bottom-up to top-down factors, whereby young infants' attention is primarily influenced by bottom-up factors but later in development becomes primarily influenced by top-down factors. However, alternative accounts suggest that instead of a general shift, what develops is refinement in understanding how different features change in importance from moment to moment. Studying how attention relates to meaning in scenes is a difficult methodological problem. Consequently, I discuss how the use of a particular measure, attentional synchrony, which is the convergence of eye gaze in time and space across multiple individuals, can reveal developmental changes in attention that are sensitive to the dynamic nature of meaning. I conclude by proposing that hands and hand-object actions are meaningful features that potentially influence visual attention.

## 1.1 Visual attention in adults

The visual world is complex and dynamic. Yet, only a small fraction of the visual world can receive our attention at any moment. In order to allocate overt *visual attention*—that is, where our eyes are pointed—we must choose where to direct eye gaze from moment to moment to center the high-acuity region of vision (fovea) to information of interest. A central issue in theories of visual attention and this dissertation concerns how we guide our attention when viewing complex, dynamic scenes—scenes where the contents are continually changing—and the developmental progression of attention allocation within dynamic scenes. When there are multiple places to look at any moment, observers must choose to look at one location at the expense of visually attending to other locations, potentially missing important information conveyed at unattended locations. Thus, it is important to understand which features in a scene (whether bottom-up or top-down) influence where observers choose to look, because it underlies their ability to gather information from within a scene. How these influences change over development may potentially explain differences in what infants, children, and adults glean when watching dynamic scenes based on differences in visual attention.

Theories of visual attention describe two main categories of features that influence human gaze. *Bottom-up features* influence attention by standing out from the scene based on low-level visual properties of the stimulus, like color or motion (Borji & Itti, 2013; Itti & Koch, 2000). *Top-down features* receive attention due their semantic meaning based on knowledge, experience, or goals (Yarbus, 1967; Tatler, Hayhoe, Land, & Ballard, 2011). I will first describe these factors as they pertain to adult attention before moving on to a developmental account of attention to both types of influence.

### Bottom-up features

Bottom-up features are typically defined as simple image properties contained within a visual stimulus. For instance, contrast (Parkhurst & Niebur, 2004), color (Jost, Ouerhani, Wartburg, Müri, & Hügli, 2005), and motion (Mital, Smith, Hill, & Henderson, 2011) are features that attract attention by "popping out" from the rest of the stimulus. Collectively, bottom-up features contribute to visual salience. Using biologically-inspired algorithms to compute these features from pixel-level image data, researchers can create predictive 'maps' that represent relative values of 'pop out' in the stimuli. Values of color, intensity (amount of luminance), and orientation (direction of lines and edges) allow for each pixel to be evaluated relative to the pixels in the rest of the image. Areas of the map with a greater salience value are predicted to receive more attention than areas with low salience based on bottom-up attention. Indeed, when comparing adult eye movements to salience maps of scene images, adults tend to look at locations that contain higher salience (Peters, Iyer, Itti, & Koch, 2005; Parkhurst & Niebur, 2004). Salience maps are most predictive of attention during adults' first few fixations when inspecting a photograph (Parkhurst & Niebur, 2003). However, dynamic scenes (videos) consist of ever-changing information, so observers must actively distribute eye gaze to attend to new information. By calculating frame-to-frame changes in the intensity and orientation within a video stimulus, salience maps can generate predictions that include values of flicker and motion, respectively. In studies using dynamic stimuli, attention is predicted by salience maps at better-than-chance levels (Franchak, Heeger, Hasson, & Adolph, 2016; T. J. Smith & Mital, 2013).

Bottom-up features are considered exogenous factors, which means they occur externally to the observer. In fact, one methodological benefit of using salience maps is that all features can be computed with no intervention from a human (Mital et al., 2011; Parkhurst & Niebur, 2004). Some researchers argue that bottom-up features are sufficient for modeling attention during free viewing (Itti & Koch, 2000; Zelinsky & Bisley, 2015). Free viewing is an experimental task in which participants are asked to simply view a stimulus with no specific instructions or objective. Modeling attention in free viewing is of particular interest as all three studies contained within this dissertation require participants to freely view stimulus videos.

However, bottom-up features may not sufficiently explain attention in all free viewing tasks because computational salience from pixel level data cannot account for semantic information that might guide attention (Võ & Henderson, 2009). For instance, comparing semantically compatible and semantically incompatible scenes shows how prior knowledge guides looking. A semantically compatible scene could show a fork on a dinner table, while a semantically incompatible scene could show a toothbrush on a dinner table. An observer's knowledge of the semantic mismatch of a toothbrush on a dinner table may violate expectations and influence more attention towards the incompatibility. In studies that record eye movements when viewing scenes that contain semantically compatible and incompatible information, salience is not able to account for increased attention to incompatible objects that violate observers' expectations (Helo, van Ommen, Pannasch, Danteny-Dordoigne, & Rämä, 2017; Võ & Henderson, 2009). Such failures of a purely bottom-up approach have led others to theorize that top-down features are more influential on visual attention, which I will review in the next section.

### Top-down features

Attention to top-down features is driven by semantic relevance based on the knowledge, goals, and experience of an observer. Because observers attribute meaning to these features, they are considered endogenous rather than exogenous factors. As described briefly in the prior section, semantic information plays a role in free viewing tasks (Võ & Henderson, 2009). Beyond semantic compatibility, attention during free viewing is often allocated to socially-relevant features like bodies, hands and facial regions (Birmingham, Bischof, & Kingstone, 2008, 2009; Franchak et al., 2016; Frank, Vul, & Johnson, 2009; Foulsham, Walker, & Kingstone, 2011; Võ, Smith, Mital, & Henderson, 2012). Social features are considered top-down because they are meaningful due to the information that they convey regardless of any explicit task demands. Socially-relevant features provide information about a social agent's intentions, actions, emotions, communication and may serve as a 'default' area of attention when free viewing (Birmingham et al., 2009).

An observer's tasks and goals are another top-down influence on attention that attribute meaning to areas in a scene that are relevant to the observer. In eye tracking studies where participants are asked to complete screen-based tasks involving memorization or searching, visual behavior is deployed to serve the given task, for instance, looking longer at objects during memorization (Castelhano, Mack, & Henderson, 2009). In real-world tasks that require participants to perform more complex actions, attention is allocated spatially and temporally to the relevant locations of the task being completed (Ballard & Hayhoe, 2009; Land, 2009). For instance, when the task involves making a sandwich, eye movements cluster around task-relevant locations such as current and future objects that are manipulated by hands (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). From this perspective, goals and task demands can influence visual attention towards areas that are meaningful for planning and execution of motor behaviors. But even when just watching others perform manual actions, observers tend to look at similar task-relevant areas, despite not engaging in the action, because observers recognize the goals of other agents (Flanagan & Johansson, 2003).

Collectively, features that signal semantic relevance within a scene are indicators of meaning. A relatively novel approach to predicting adult attention involves creating 'meaning maps' based on the semantic relevance of many smaller overlapping regions, or patches, of a static scene (Henderson & Hayes, 2018). Meaning maps are generated by crowd-sourced semantic ratings of image patches. When applied to the image, these ratings create a spatial distribution of what adults have rated as the most semantically meaningful. When those ratings are combined into a meaning map across the entire image, patches with higher meaning ratings are more likely to attract eye movements. More recently, Rehrig and colleagues (2020) have generated 'grasping maps' by crowd-sourcing ratings of 'graspability' for small patches of the scene instead of more generic "meaning" ratings. These 'grasping maps' can accurately predict adults' visual attention to scenes in which they have been asked to describe possible actions. Meaning and grasping maps support the notion that visual attention is guided by what observers deem to be meaningful. While meaning maps have not been directly applied to dynamic videos, it is likely that a driving force in visual attention is to select areas that convey the most meaning.

### Combining bottom-up and top-down features

The prior sections indicate that both bottom-up and top-down features can predict adult visual attention. However, it is important to compare these influences and to understand how they may interact. Firstly, that both bottom-up and top-down features can predict attention suggests that both feature types overlap. For instance, while faces may typically be termed a top-down feature because they convey meaning; salience calculations indicate that face regions often contain high levels of visual salience (Henderson, Brockmole, Castelhano, & Mack, 2007; Torralba, Oliva, Castelhano, & Henderson, 2006; Wass & Smith, 2015). The overlap between bottom-up and top-down is exacerbated with the inclusion of dynamic channels in salience algorithms. Top-down features like objects and people are likely to move within a scene. When salience algorithms are applied to dynamic scenes, there is often overlap in what is considered salient and semantically meaningful (Mital et al., 2011). The overlap makes it difficult to delineate the individual contribution of each if bottom-up and top-down features are not mutually-exclusive influences on attention.

Secondly, direct comparisons between salience maps and top-down models of adult attention typically indicate that top-down models outperform salience models. Adult attention is better predicted using models that are based on faces (Frank et al., 2009; Rider, Coutrot, Pellicano, Dakin, & Mareschal, 2018) and meaning (Henderson & Hayes, 2018). However, research from my collaborators and I beyond this dissertation indicate that salience is still important. For instance, when adults viewed a narratively incoherent vignette with scenes out of order, salience was a stronger predictor of attention compared to the attention of adults who viewed a narratively coherent video (Jing, Kadooka, Franchak, & Kirkorian, in press). Attention to salience may be a strategy to find meaning by reverting to low-level visual information when top-down features are less available. Moreover, redundancy of bottom-up and top-down factors may help viewers to allocate attention more effectively by allowing for cue combination. Recently, we found that bottom-up salience and top-down biases towards the center of a screen, in combination, cue adults to look at faces (Franchak & Kadooka, 2022).

## 1.2 Visual attention in infants and children

### 1.2.1 Developmental changes in attention to bottom-up and top-down features

In the developmental attention literature, there is prevalent support for an age-related shift in the influences of attention from bottom-up factors to top-down factors, starting during infancy (Frank, Amso, & Johnson, 2014; Kwon, Setoodehnia, Baek, Luck, & Oakes, 2016; Frank et al., 2009; Gluckman & Johnson, 2013; Amso, Haas, & Markant, 2014; Rider et al., 2018). Theories of visual attention development suggest the shift involves improvements in voluntary control of gaze, which is the ability to intentionally orient attention (Colombo, 2001; Oakes & Amso, 2018). More specifically, young infants' attention is viewed as obligatorily 'stimulus-driven', meaning that areas of high visual salience capture infants' attention (Colombo, 2001; Stechler & Latz, 1966). Age-related improvements in voluntary control allow infants to increasingly inhibit attention to salient features, and instead voluntarily select where to look. If their attention is not captured by bottom-up features, infants can then choose to direct visual attention towards top-down features that are defined by infants' own goals and prior knowledge (Oakes & Amso, 2018).

Importantly, if development changes whether attention is involuntarily captured by bottom-up features versus free to voluntarily attend to meaningful, top-down features, it suggests that differences in attention to features should be consistently observed regardless of the stimulus. I refer to this idea as the 'Global Shift Hypothesis' in Chapter 2, where 'global' means that infants should consistently attend to bottom-up features regardless of the stimulus, whereas older participants should consistently attend more to top-down features regardless of the stimulus. Yet, this hypothesis has not been exhaustively tested. Although the prediction states that over development observers should increasingly attend to top-down features generally, past work testing developmental changes in top-down attention have focused on measuring attention to faces, specifically (Amso et al., 2014; Franchak et al., 2016; Frank et al., 2009; Kwon et al., 2016; Rider et al., 2018). By only considering faces as a top-down feature, development of attention to other semantically-meaningful top-down features is ignored. As discussed previously, there are also methodological issues with considering attention to faces as purely a top-down feature since faces are also visually salient (Wass & Smith, 2015). Consequently, evidence for the Global Shift Hypothesis is mixed.

Supporting the Global Shift Hypothesis, general increases in attention to faces are reported in the first year of life and beyond. Kwon and colleagues (Kwon et al., 2016) found that the attention of 4-month-olds was biased towards salient images in an image array, however, by 8 months, attention was preferentially allocated to faces even when more salient targets were available. In dynamic videos, a similar pattern has been observed in which a model based on salience is better able to predict the attention of 3-month-olds, but the attention of 6-month-olds, 9-month-olds, and adults was better predicted by a face-looking model (Frank et al., 2009). Both findings support the idea that younger infants' attention is captured by salient areas, but older infants can inhibit looking at salient areas to voluntarily look towards more meaningful areas (e.g., faces). Improvements in face-processing abilities in the first year of life may aid in discrimination and selection of faces (Farzin, Hou, & Norcia, 2012; Pascalis, de Haan, & Nelson, 2002). As described by Colombo (2001), infants younger than one year old undergo improvements in orienting and voluntary control of attention. Developmental changes in attention abilities may bias attention to faces by helping to select semantically-relevant features.

However, not all published studies have found consistent findings in support of a shift from bottom-up to top-down attention with age. In two separate studies, infants ranging from 3-30 months old and children from 6-8 years old exhibited age-related *decreases* in looking at faces for particular stimuli (Frank et al., 2009; Stoesz & Jakobson, 2014). The introduction of multiple agents in the scene moderated the influence of face looking for younger observers. Franchak et al. (2016) reported a similar effect of scene content, in which scenes with multiple agents (compared with a single agent) changed face looking behavior for adults, but not young infants. Taken together, age-related biases in attention to features do not appear consistent and may be idiosyncratic to the content of a scene. Wider testing of diverse stimuli is needed to determine the presence of developmental changes that are invariant to stimuli effects (i.e., 'global'). I address the question of a global shift in Chapter 2.

The Global Shift Hypothesis is motivated by the theory that voluntary control of attention becomes increasingly endogenous (Colombo, 2001; Kwon et al., 2016; Oakes & Amso, 2018), but a global shift is not the only way that increasingly endogenous attention may manifest. However, predicting a global shift in looking to different *types* of features may be an oversimplification of how attention is allocated. The results of the study in Chapter 2 suggest this is the case. The studies in Chapters 3 and 4 explore an alternative hypothesis: Increasing voluntary control of attention over development may lead to increases in prioritizing meaningful areas. That is, meaningfulness in a scene is not solely defined by a specific type of feature at all times. Faces may convey meaning at one second, but a gesturing hand may convey more meaning in the next second. Thus, better voluntary control of attention should manifest as increases in attention to meaning, which may appear as many different types of features. An observer with mature visual attention should be able to flexibly attend to any of the features that convey meaning from moment to moment. This presents a challenge to measuring meaning, given that singular features are unable to serve as locations of expected meaning for the entire duration of a stimulus or across many different stimuli. In the following sections, I provide evidence in support of this alternative approach to developmental changes in attention and how meaning may be measured.

### 1.2.2 Comprehension influences attention

If meaningful features guide attention, then developmental differences in understanding semantic information should impact visual attention. For instance, an infant that does not comprehend the meaning or semantic relevance of a TV remote may view a cluttered living room differently than an adult. Here, I consider the role of comprehension as an influence on attention that might change over development.

Observers distribute their attention towards information that conveys semantic relevance or meaning. Attention to scenes that violate expectations of semantic relevance can reveal the semantic expectations of an observer. For instance, adults comprehend the patterns of regularities in our environment which help direct their gaze (Helo et al., 2017). When presented with scenes containing irregularities (e.g. a bar of soap on a kitchen table), adults but not 2-year-olds were sensitive to the irregularity and looked longer to the irregular objects. Preschool-aged children watching incomprehensible TV show content exhibit shorter durations of looking and lower levels of overall attention compared to comprehensible scenes (Anderson, Lorch, Field, & Sanders, 1981). In this case, incomprehensible content is less meaningful and should receive less attention. However, the ability to modulate attention based on meaning requires a bare minimum of noticing differences in the first place. Infants less than a year old spend the same amount of time looking to comprehensible and incomprehensible scenes (Pempek et al., 2010). Along with increases in voluntary control of attention, developmental changes in the ability to comprehend information about a visual scene may drive observers to better attend to meaningful information.

### 1.2.3 Adult synchrony

In characterizing how visual attention develops, measurements of adult visual attention allow for comparisons along the expected age-related changes in attention to features that are expected to be meaningful. As an example, if a researcher expects developmental changes in attention to meaningful objects, the researcher must create, code, and calculate a measurement to index attention to objects and determine whether there are developmental changes. The difficulty of a single-feature approach is that comparisons require a different measurement for each feature and potentially new coding schemes for each stimulus. Further, concerns of validity are warranted for measurements that could be biased by a researcher's judgements of when a feature is meaningful. Prior sections in this review describe effects of idiosyncrasies in stimuli content (Frank et al., 2009; Franchak et al., 2016) and comprehension (Helo et al., 2017; Pempek et al., 2010) on attention. Taken together, adult visual attention is characterized by many different dynamic features but all in service toward guiding attention to information that the observer determines is most meaningful from moment to moment.

In response to the methodological challenge of indexing meaning, I propose the use of *adult synchrony.* Adult synchrony is the spatiotemporal consistency between an observer compared to adults based on a correlation of their eye movements when watching the same stimulus. By comparing an observer to adults, adult synchrony scores provide a metric of attention that captures how adult-like an observer's eye gaze is. Correlations closer to one indicate more adult-like attention. Franchak et al. (2016) showed that age predicted adult synchrony. Younger infants (6 months) had lower adult synchrony scores, whereas toddlers

(24 months) showed greater synchrony with adults. This suggests that adult synchrony can measure the progression towards more mature gaze by capturing the extent to which observers prioritize where to look in a way that is similar to adults.

Why is there an age-related increase towards greater adult synchrony? Adults may share similar concepts of what is important and meaningful in a dynamic scene. As a consequence, adult synchrony is sensitive to the spatiotemporal changes in meaning that occurs in a video from moment to moment. A face may be meaningful to look at briefly, but the meaningfulness may shift toward a held object in the next moment. A model based purely on a single feature, like looking to faces, would not capture this change in meaning.

A consistent finding in adult eye behaviors studies is that adults tend to look towards similar locations in time when viewing dynamic stimuli (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Franchak et al., 2016; Hart et al., 2009; Mital et al., 2011; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010; Wang, Freeman, Merriam, Hasson, & Heeger, 2012). Highly correlated adult eye movements are consistent even across wide ranges of stimuli involving live-action, animated, and professionally produced films (Dorr et al., 2010; Gannon & Grubb, 2022; Goldstein, Woods, & Peli, 2007; Shepherd et al., 2010; Franchak et al., 2016; Rider et al., 2018). Age-related increases towards more consistent adult gaze (less spatial variability) is supported by a few studies, (Frank, Vul, & Saxe, 2012; Kirkorian, Anderson, & Keen, 2012) but we are among the first with Franchak et al. (2016) to specifically use adult synchrony as a general similarity index of attention and test factors that may predict developmental changes in attention.

Adult synchrony provides a measurement that is specific to a single stimuli and is agnostic to constructs that are imposed by the researchers. The role of multiple features can be tested against one developmental change (i.e. towards greater synchrony) to determine the individual contribution of each feature in accounting for age-related increases in adult-like attention. Adult synchrony, as a product of prioritization of meaning, also leads to interesting predictions about how adult synchrony may change as a result of improvements in comprehension. As infants and children gain perceptual and motor experiences that improve comprehension, adult synchrony may also improve as they learn to prioritize meaningful information in ways that are similar to adults. In the next section, we apply this concept within the context of goal-directed manual actions.

## 1.3 Development of attention towards meaning in hands

In Chapters 3 and 4, I test two features related to hands as predictors of adult synchrony: hands and hand-object actions. First, attention may be allocated to hands because hands can convey meaningful information related to social relevance (Fausey, Jayaraman, & Smith, 2016; Frank et al., 2012), non verbal communication (Bertenthal & Boyer, 2012; Tomasello, Carpenter, & Liszkowski, 2007), joint attention, (Yu & Smith, 2013; Deak, Krasno, Triesch, Lewis, & Sepeta, 2014), and manual action (Flanagan & Johansson, 2003). Second, there are developmental changes in children's comprehension of hands over the first few years of life as they learn to comprehend pointing (Bertenthal & Boyer, 2012), hand cues in early word learning (de Villiers Rader & Zukow-Goldring, 2010) and more specifically about information conveyed during hand-object action: object properties (P. Tseng, Bridgeman, & Juan, 2012), goals (Woodward, 2009), and affordances (Klatzky, Pellegrino, McCloskey, & Doherty, 1989). Improvements in comprehension could be reflected in more adult-like prioritization of hands and hand-object actions. However, sensitivity to the information conveyed by hands, does not necessarily mean that infants and children will select hands as meaningful locations of attention, especially dynamic scenes with various features competing for attention.

The selection of hands as a predictor is driven by prior studies indicating that increased comprehension of action should lead to more attention to hand locations (Bertenthal & Boyer, 2012; P. Tseng et al., 2012; Woodward, 2009). Although other aspects of action understanding might also drive observers' attention, a practical benefit of using hands is that they are easy to spatially define in an image. For example, the attention of actors within a scene may cue areas for joint attention, but determining where actors were looking would be challenging to reliably identify in a video for every frame. Hands are physical entities and hand-object actions are defined by the time when hands are in physical contact with movable objects. We predicted that observers who synchronize their attention to hands and hand-object actions, spatially and temporally, would correspond with observers who synchronized attention in an adult-like way.

I propose that infants' and children's developing action understanding should produce greater attentional synchrony with adults when viewing manual actions. Prior work with adults shows that when observers comprehend the action they are watching, their eye gaze moves to the same locations at the same time as the actor performing the action (Flanagan & Johansson, 2003; Kochukhova & Gredebäck, 2010). The relationship between where a person looks and what a person understands is reciprocal. In Chapter 4, I will test whether providing perceptual opportunities to improve comprehension leads to increases in children's adult synchrony. Perceptual opportunities may increase knowledge about an action and lead to greater attention at the locations that are meaningful. In daily life, perceptual opportunities occur in the ways we explore and interact with our environment. In training studies, the role of perceptual opportunities is understood to improve infants (and adults) ability to discern regularities and goal information contained with actions (Sommerville, Woodward, & Needham, 2005; Monroy, Gerson, & Hunnius, 2017). However, it is not known if improvements in action comprehension are also changing eye movements to look at specific locations at specific times.

# Chapter 2

# Developmental changes in infants' and children's attention to faces and salient regions vary across and within video stimuli

## 2.1 Abstract

Visual attention in complex, dynamic scenes is attracted to locations that contain socially-relevant features, such as faces, and to areas that are visually salient. Previous work suggests that there is a *global shift* over development such that observers increasingly attend to faces with age. However, no prior work has tested whether this shift is truly global, that is, consistent across and within stimuli despite variations in content. To test the global shift hypothesis, we recorded eye movements of 89 children (6 months to 10 years) and adults while they viewed seven video clips. We measured the extent to which each participant attended to faces and to salient areas for each video. There was no evidence of global age-related changes in attention: Neither feature showed consistent increases or decreases with age. Moreover, windowed analyses within each stimulus video revealed significant moment-to-moment variations in the relation between age and each visual feature (via a bootstrapping analysis). For some time windows, adults looked more often at both feature types compared to infants and children. However, for other time windows the pattern was reversed—younger participants looked more at faces and salient locations. Lack of consistent directional effects provides strong evidence against the global shift hypothesis. We suggest an alternative explanation: Over development, observers increasingly *prioritize* when and where to look by learning to track which features are relevant within a scene. Implications for the development of visual attention and children's understanding of screen-based media are discussed.

## 2.2 Developmental changes in infants' and children's attention to faces and salient regions vary across and within video stimuli

The visual world is dynamic. Since we cannot "rewind" or "pause" events in real life, we must look at the right place at the right time to glean the most important information. Poor visual acuity in the peripheral areas of the visual field means that humans must make eye movements to direct the high-acuity fovea towards the informative areas in a scene from moment to moment (Land & Fernald, 1992; Westheimer, 1982). What influences where observers look, and how do those influences change over development? Two influences that have been widely studied in the developmental literature are *socially-relevant features* (e.g., faces) and *visually-salient features.* Faces influence visual attention by drawing gaze towards socially-meaningful locations that convey information such as affect, attention, and speech (for a review, see (Bruce, 1993)). Visually-salient features attract gaze to locations whose appearance (e.g., color, motion) stands out from the surrounding scene (Borji & Itti, 2013; Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Itti & Baldi, 2005). Developmental changes in attention from visually salient features to meaningful areas in a scene, such as faces, could be indicative of a shift in attentional biases. We will refer to this as the *global shift hypothesis*, and review the evidence in greater detail below.

However, little scrutiny has been given to whether developmental changes in attention to faces and salient areas are truly **global**. A global developmental change in attention should be found consistently across and within stimuli that vary in content. Yet, prior developmental studies of free viewing often present only a small number of stimuli of short duration and/or aggregate looking measurements over an entire stimulus rather than test the consistency of attention patterns across and within stimuli. To address these limitations and test the consistency of age-related changes in attention, we measured infants' and children's (6 months to 10 years) and adults' eye movements across and within a wide set of stimuli with diverse content. Using a sufficiently large data set, we measured attention to faces and visually-salient locations to examine whether developmental changes in attention to each type of feature were global—that is consistent across and within stimuli.

Understanding whether and how visual features' influence on attention changes over development has broad significance. For instance, atypical patterns of looking to faces has been implicated in identifying infants and children who are at risk for Autism Spectrum Disorder (Klin & Jones, 2008; Klin, Jones, Schultz, Volkmar, & Cohen, 2002). Using a diverse stimulus set can inform on whether there are developmentally-normative changes in face looking that are independent of stimulus variations. It is also important to understand attention development in the context of viewing screen-based media. Watching TV shows and videos/DVDs is pervasive: 35% of children aged 0-2 are exposed to screen media on a daily basis, and those that are exposed average 42 minutes of viewing per day (Rideout, 2017). Viewing becomes more common and more extensive with age: 67% of children aged 2-4 exposed to screen media each day with an average duration of 159 minutes. Despite the purported educational benefits of media intended for infants and children, there are well-documented limits on what children actually learn (Wartella, Richert, & Robb, 2010). Studying how visual features influence looking behavior has potential implications for understanding how media should be designed to improve children's learning of educational content.

### 2.2.1 Faces and salient locations attract adults' attention

Adults distribute their attention to socially-relevant locations, such as people's bodies (Foulsham et al., 2011), eyes (Birmingham et al., 2009), and locations relevant to the goals of others' motor actions (Ballard & Hayhoe, 2009; Land, 2009). Faces are a particularly strong feature that captures adults' attention (Birmingham et al., 2009; Franchak et al., 2016; Frank et al., 2009; Shepherd et al., 2010). Heightened attention to faces and facial features is found for both static (photo) and dynamic (video) stimuli (Birmingham et al., 2008, 2009; Võ et al., 2012; Yarbus, 1967), even when there is no specific viewing task (Birmingham et al., 2009), suggesting that socially-relevant features serve as a "default" location of interest.

Adults also look towards visually-salient features, which capture attention based on their appearance. Locations that are colorful (Jost et al., 2005), high contrast (Parkhurst & Niebur, 2004; Reinagel & Zador, 1999), and contain motion (Mital et al., 2011) attract eye gaze regardless of whether the location is meaningful because those areas "pop out" by having a different visual appearance from the surrounding scene. For example, a lone painting hung askew will catch the eye when placed on a wall of properly-leveled artwork because of its unique orientation (rather than the content depicted in the painting). To quantify the degree to which a location in a scene differs in appearance from its surroundings, biologically-inspired computational saliency models have been devised to calculate relative saliency of locations based on different feature channels (Borji & Itti, 2013; Itti & Koch, 2000; Itti et al., 1998; Itti & Baldi, 2005). Comparing model predictions to adult gaze patterns confirms that fixated locations tend to have higher visual saliency compared with non-fixated locations when viewing both static images and dynamic scenes (Parkhurst & Niebur, 2004; Peters et al., 2005; T. J. Smith & Mital, 2013).

### 2.2.2 Evidence for and against a global developmental change in visual attention

Consistent with the global shift hypothesis, several developmental studies have found age-related increases in looking at faces in static images (Amso et al., 2014; Gluckman & Johnson, 2013; Kwon et al., 2016). Infants younger than 6 months spent more time looking towards salient images of objects, but older infants attended to faces despite the presence of non-face images with greater visual salience (Kwon et al., 2016). A similar trend has been found in studies using dynamic stimuli (Frank et al., 2014, 2009; Rider et al., 2018). Frank and colleagues (2009) showed that 3-month-old infants' eye movements when watching an animated clip were better predicted by a low-level salience model, but eye movements of 6-month-olds, 9-month-olds, and adults were better predicted by a face-looking model. A study comparing children (6-14 years) and adults watching videos found that face models were better or equal to salience models (depending on the stimulus) at all ages, however, face models were more predictive of adults' attention compared with children's attention (Rider et al., 2018). Increases in face-looking rates with age—particularly in the first year of life—parallel developmental improvements in infants' visual search skill (Frank et al., 2014) and infants' ability to discriminate and process faces (Farzin et al., 2012; Pascalis et al., 2002).

However, an *increase* in face looking does not necessarily entail a corresponding *decrease* in attention to areas with high saliency. Although several studies indicate decreasing influences of saliency on attention with age (Helo, Pannasch, Sirri, & Rämä, 2014; Kwon et al., 2016; Açik, Sarwary, Schultze-Kraft, Onat, & König, 2010), others find that saliency models are *more* predictive of adults' gaze compared with infants and children (Rider et al., 2018; Franchak et al., 2016; Frank et al., 2009). One explanation is that faces tend to have higher saliency than irrelevant locations in a scene (Torralba et al., 2006; Wass & Smith, 2015). Thus, a global developmental increase in looking at faces may or may not be accompanied by a change in looking to salient regions depending on the correspondence between salience and faces in a given stimulus. Variations in the salience of faces in different videos—such as when comparing videos intended for child and adult audiences (Wass & Smith, 2015)—further motivates the need to test the consistency of age-related changes in attention across a wider set of stimuli.

Despite evidence in support of the global shift hypothesis, there are a few conflicting results. Frank, Vul, and Saxe (2012) found different age-related changes in 3- to 30-month-olds' attention to faces depending on scene content: Age predicted an *increase* in looking at faces for scenes that contained close-ups of children but predicted a *decrease* in face looking for scenes that included wide shots with multiple agents. Similarly, Franchak and colleagues (2016) found variations between infant and adult eye movements depending on scene content. For scenes with one agent, adults' gaze was predicted by both saliency and looking to the actor's face. However, for scenes with multiple agents, adults suppressed looking to salient areas, looked at the main actor's face, but rarely looked at the other actors.

In contrast, young infants looked at moderately-salient locations and infrequently looked at the main actor's face regardless of how many agents were in view. By 24 months, toddlers' viewing patterns were adapted to scene content in a similar way to adults. Lastly, Stoesz and Jakobson (2014), found that the addition of more actors in a scene led to decreases in face looking that were more pronounced for children than for adults.

Given the moderating role of scene content, support for the global shift hypothesis requires testing whether age differences in looking to faces and salient locations are invariant to differences in stimuli. However, prior studies have primarily tested only a single stimulus video or a small set of stimuli, which limits the ability to detect a global pattern across diverse content (Franchak et al., 2016; Kirkorian et al., 2012; Frank et al., 2009, 2014). Furthermore, the studies above suggest that different priority may be given to different features from scene to scene within a stimulus video. However, most studies of visual attention report average measures over the entire duration of a stimulus (Frank et al., 2009; Kirkorian et al., 2012; Rider et al., 2018) or compare a few select scenes or scene types (Franchak et al., 2016; Frank et al., 2012, 2014; Stoesz & Jakobson, 2014). Finer temporal granularity—that is, determining whether there are age differences on shorter time windows—could reveal whether age differences in attention are robust to variation in scene content within a stimulus.

## 2.3 Current study

The goal of the current study was to test the global shift hypothesis by assessing age differences in attention to faces and visual saliency across and within scenes with varying content. Because visual attention changes throughout infancy and childhood (Colombo, 2001; Oakes & Amso, 2018), we tested participants across a wide age range (6 months to 10 years and college-age adults). To our knowledge, no prior work has examined changes in looking to faces and salient locations in video stimuli that spans from infancy to adulthood, making this dataset unique. Participants watched seven 2-minute video clips from various child-friendly media while eye movements were recorded. We chose videos with diverse content to determine whether changes in visual attention to specific feature types are global, that is, invariant across stimuli. Exemplar screenshots and descriptions of the seven stimuli used in this study are available on Databrary (https://nyu.databrary.org/volume/1007). We calculated the proportion of time spent looking at faces in each scene (*face looking*) and the visual saliency of areas attended by each participant (*gaze saliency*) based on calculations from a saliency model (Harel, Koch, & Perona, 2006; Itti & Baldi, 2005).

First, we tested whether there were consistent age-related changes in attention to faces and salient areas across the stimulus set. If there is a global shift towards looking more often at faces, we predict a consistent age-related increase in face looking for each of the seven videos. We made no specific prediction concerning global changes in gaze saliency given that saliency of faces may vary. An additional consideration is how best to represent the trajectory of age-related differences in visual attention. Previous work found rapid increases in face looking during infancy followed by a modest rate of change for older children and adults (Amso et al., 2014). For this reason, we calculated logarithmic in addition to linear functions to model age differences.

A second set of analyses tested how consistently age-related changes in attention to faces and salient locations exist over changes *within* each video. Rather than defining ad hoc scenes of interest as in past work (Franchak et al., 2016), we objectively and exhaustively tested temporal changes in eye movements by using a sliding window analysis. For each stimulus video, we defined 10-s windows every 5 s, resulting in 22 windows. Face looking and gaze saliency were calculated within each window for each participant to capture differences in attention as the scene changes. Evidence of truly global age-related changes in attention would entail greater face looking with age that is invariant over time within a video. Alternatively, if changes in scene content alter the importance of different features over time, age differences in gaze saliency and face looking may vary across windows. For example, in one window (or a few successive windows) adults may attend towards faces more so than children (a positive correlation between age and face looking). At a different time window faces might be less important to the scene, resulting in adults looking less often at faces compared to younger participants (a negative correlation between age and face looking). Such variation in the direction of age differences in face looking from moment to moment would provide evidence against the global shift hypothesis.

It is important to note that saliency models are developed and evaluated by using databases of adult eye behaviors, raising a potential concern that saliency models may not be equally valid when applied to infants' and children's data. However, a recent comparison of saliency model performance found that the model used in the current study is one of the best models for predicting both adult and infant gaze to static images across seven evaluative metrics (Mahdi, Su, Schlesinger, & Qin, 2017). Since no past work has evaluated different dynamic saliency models for modeling infant eye movements, we cannot rule out that a model tuned to infants and/or children would perform better. We also note that different models capture different visual features. The model we chose uses flicker and motion on a pixel level to measure dynamic change in a scene, but other models use visual entropy or a Bayesian representation of surprise (Itti & Baldi, 2009) and could provide different insights of developing attention.

## 2.4 Method

### 2.4.1 Participants

Our goal was to analyze continuous effects of age from 6 months to 10 years. To ensure that sufficient data were collected across that entire range, participants were recruited from 7 narrower age ranges: 6-to 11-month-olds; 12-to 17-month-olds; 18-to 23-month-olds; 2-to 4-year-olds; 4-to 6-year-olds; 8-to 10-year-olds; and college-aged adults. Infant age ranges were spaced closer together compared with child age ranges because past work indicated rapid developmental change in infancy followed by more gradual changes during childhood (Amso et al., 2014). We defined an a priori stopping rule based on data quality: Run participants until each **stimulus** has data of sufficient quality from 10 participants within each age range. Eye movement data were considered insufficient and excluded on a stimulus-by-stimulus basis if any of the two criteria were met: 1) the participant's eye gaze data were missing (e.g., eye occlusion, looking away) for $> 50\%$ of the frames of a video, or 2) eye gaze data were missing for any single continuous period of $> 30$ s. Due to these exclusion criteria, participants in the final sample provided data for between 2 and 7 stimulus videos. For example, if an infant attended the first three videos but then refused to watch the remainder, the infant contributed data to 3/7 stimuli. At minimum, 10 adults and 60 children (6 months to 10 years) would be required. However, it was necessary to run additional younger participants because they were less likely to stay engaged through the entire session and consequently failed to contribute data to all 7 stimuli. Beyond the minimum of 10 per age range, we were required to run an additional ten 6-to 11-month-olds, three 12-to 17-month-olds, three 18-to 23-month-olds, one 2-to 4-year-old, one 4-to 6-year-old, and one 8-to 10-year-old to ensure that each stimulus had sufficient data. Each video had data ranging from 76 to 87 participants. Table 2.1 displays the final sample size for each of the seven videos and shows the smallest age effect ($r$) that could be detected at 80% power. Based on these effect size calculations, the study was adequately powered to detect medium effects of age.

Table 2.1: Sample size ($n$) and smallest effect size ($r$) that could be detected with 80% power for each video

| | $n$ | $r$ |
|---|---|---|
| Video 1 | 85 | .29 |
| Video 2 | 87 | .29 |
| Video 3 | 83 | .30 |
| Video 4 | 75 | .31 |
| Video 5 | 83 | .30 |
| Video 6 | 87 | .29 |
| Video 7 | 82 | .30 |

The final sample consisted of 79 children ranging in age from 6-months-old to 10-years-old (42 female) and 10 college-aged adults (5 female). All participants in the final sample had normal or corrected-to-normal vision with no color blindness or history of familial color blindness. Families were recruited from the Riverside County area. Participating children were identified by their caregivers as Black/African American ($n = 1$), American Indian/Alaskan Native ($n = 5$), non-Hispanic White ($n = 21$), Hispanic or Latino(a)/White ($n = 31$), and more than one race ($n = 21$). Adults were college undergraduates recruited from the departmental participant pool and received course credit for participation. Adult participants identified as American Indian/Alaskan Native ($n = 1$), non-Hispanic White ($n = 2$), Asian ($n = 3$), and Hispanic or Latino(a)/White ($n = 4$). Families received $10 and a small gift or book for participating. The study procedure conforms to the US Federal Policy for the Protection of Human Subjects and was approved by the Institutional Review Board of the University of California Riverside under protocol HS-16-126: "Development of visual exploration while watching videos". Participants (or their caregivers) signed an informed consent document after hearing the details of the study. Children aged 2-4 years gave verbal assent and children aged 5-10 years provided written assent.

Eleven additional participants were tested (9 infants/children and 2 adults) but their data were excluded completely due to issues affecting the entire experimental session: failed to complete the experiment due to fussiness/inattention to all seven videos (9 children), falling asleep (1 adult), and distraction (checking a mobile phone instead of looking at the stimuli, 1 adult).

### 2.4.2 Stimuli

Seven child-friendly videos were selected to present stimuli with diverse content: three Sesame Street videos, three music videos, and one children's science demonstration video. Each video was 2 min in duration with limited graphical elements and no cuts (i.e., each stimulus was presented as a continuous shot). Beyond these criteria, the selected videos varied in a number of ways: the number of agents on screen, the types of actions performed, the presence of non-human agents, and the presence of non-agentive movement. Each stimulus video, overlaid with data from infants and adults, is available to view on Databrary (https://nyu.databrary.org/volume/1007). To isolate the role of visual information on visual attention, audio cues that would inform gaze location (Coutrot & Guyader, 2014) were removed by replacing the original audio tracks with children's instrumental music. Every participant received the same pairing of music and video.

### 2.4.3 Apparatus

Each stimulus video was presented at 30 Hz on a 43.2 cm (diagonal) wide-screen monitor at a viewing distance of  60 cm. Stimulus videos subtended a visual angle of 31°×19°. The monitor was affixed to an adjustable arm and equipped with an Eyelink 1000 Plus remote eye tracker (SR Research Ltd.). Eye movements (right eye only) were recorded with a temporal resolution of 500 Hz.

### 2.4.4 Procedure

Participants sat in a viewing room that was separated from the experimenter by a hanging curtain. A target sticker was placed on the forehead to facilitate the eye tracker detecting the observers' eyes. Infants sat in a high chair with a 5-point harness to reduce body movement. Infants' caregivers sat behind infants and were instructed to refrain from interacting with infants, pointing at the screen, or speaking. Children and adults sat in a chair (with a booster seat for younger children).

At the beginning of the study, the experimenter adjusted the monitor and calibrated the eye tracker. For infants, an attention-getting video played while the experimenter adjusted the monitor. A 5-point calibration routine was used for participants of every age and was followed by a 5-point validation check. Validation data were used to calculate the average error in degrees of visual angle between the target location and estimated point of gaze location. The calibration process was repeated if validation indicated an average error of less than 1.5° of visual angle. As described by Wass, Smith, and Johnson (2013), infant eye tracking data is often lower quality compared with older participants, which impacts both accuracy (disparity between reported and actual point of gaze) and precision (disparity between successive samples of reported point of gaze). Accuracy averaged $M$ = 0.54° ($SD$ = 0.26) across age. The correlation between age and average spatial errors was marginally significant ($r$ = -.204, $p$ = .055) with older participants having higher accuracy. However, when comparing average visual error by age groups, there was only a difference of 0.2° of visual angle between 6-to 11-month-olds ($M$ = 0.66°, $SD$ = .29) and adults ($M$ = 0.43°, $SD$ = .15), suggesting that differences in accuracy would have a minimal effect on analyses. Precision for each participant was calculated following a published method (Wass, Forssman, & Leppänen, 2014) using data from each video for which the participant contributed data. Precision averaged $M$ = 1.68° ($SD$ = .32) and was not significantly correlated with age ($r$ = -.044, $p$ = .680).

After calibration and validation, participants were shown the 7 stimulus videos in a randomized order. Adults and children were instructed to simply watch the videos. Each stimulus video was preceded by a gaze-contingent target in the middle of the screen that required a fixation for $>$ 250 ms to trigger the video to start.

### 2.4.5 Data processing

Because of concerns about the validity of fixation detection algorithms when applied to younger participants with less robust data (Wass et al., 2014), raw eye tracking data were used to measure gaze behaviors. Data were extracted as a time series of horizontal and vertical gaze coordinates for each observer for each of the 7 stimuli. Time points were excluded if gaze locations exceeded the screen boundaries or were otherwise missing (eyes closed, turned away from screen, and eye occlusions).

#### Face looking

The proportion of time spent looking at the faces of agents was obtained using dynamic area of interest (AOI) analyses. For each video frame, Dataviewer software (SR Research Ltd.) was used to draw elliptical AOIs around the heads of each humanoid agent (i.e., human actors and Muppet characters) as they moved in the scene. *Face looking* was defined when the gaze location fell within the boundary of a face AOI. To compare across stimuli which had varying amounts of times with faces present on screen—and between participants who had different amounts of missing data—face looking rates were calculated for each participant by dividing the number of samples looking at faces by the number of samples with faces present during which the participant had valid (non-missing) data.

### Gaze saliency

*Gaze saliency* was calculated to determine the relative saliency of visually attended locations in comparison to the rest of the scene as in past work (Franchak et al., 2016; Tatler et al., 2011; T. J. Smith & Mital, 2013). Video frames were converted to images at the rate of presentation (30 Hz). Using the algorithm of Itti and Baldi (2005) as implemented in the GBVS toolbox (Harel et al., 2006), the relative salience of each pixel was calculated for each frame based on a combination of five feature maps (contrast, orientation, color, flicker, and motion). Dynamic features—flicker and motion—were calculated by comparing differences in successive video frames. For example, pixel changes that occur as a character moved from off-screen to on-screen would indicate greater flicker and motion relative to an otherwise still background. Image feature maps were weighted equally to create a composite saliency map, integrating static and dynamic features. Each pixel within the map was assigned a rank between 1 to 100 which reflected its saliency relative to the other pixels in the video frame; the most salient pixel ranked 100. An example of an heatmap showing the saliency of different regions on a video frame is available on Databrary (https://nyu.databrary.org/volume/1007). For every frame of each video, the average saliency rank of pixels was calculated within a 1.2° diameter circle around the point of gaze. Larger gaze saliency scores indicate that the participant looked at a relatively more salient location within the frame.

## 2.5 Results

The first set of analyses assessed the global shift hypothesis by measuring age-related changes in gaze towards faces and salient features across the seven stimulus videos. The second set of analyses tested the global shift hypothesis by measuring age-related differences in attention from moment to moment within each stimulus video.

### 2.5.1 No consistent age differences in face looking or gaze saliency across stimuli

To assess the global shift hypothesis, we tested across videos for consistent age-related **increases** in face looking and consistent changes (either increases or decreases) in gaze saliency. We used generalized estimating equations (GEE) to model the age-related changes. Like a regression model, GEEs estimate change in criterion variables (i.e., gaze saliency and face looking) from predictors (i.e., age as a continuous variable and video as a categorical factor). The main advantage of using GEEs is that they can handle participants contributing varying amounts of data in a repeated measure, whereas an ANCOVA would require excluding participants who do not contribute to every level of the repeated measure (i.e., participants who did not watch all 7 videos—likely infant observers—would be excluded). We tested two GEE models for each visual feature: A model with age as a linear (continuous) predictor and a model with log-transformed age as a continuous predictor; both included video as a categorical predictor. Follow-up analyses examined each video separately with regression models testing for changes in looking according to age.

### Face looking

The proportion of time spent looking at faces varied widely between videos (Figure 2.1), ranging from $M$ = .017 ($SE$ = .001) for Video 5 to $M$ = .831 ($SE$ = .009) for Video 1. The overall range in face-looking rates speaks to the diversity of the content. Face looking was high in Video 1, which depicted multiple agents playing a song together, but was low in Video 5, which focuses on a series of mechanical events (human agents play a peripheral role). However, contrary to the global shift hypothesis there was no uniform age-related increase in face looking across these diverse videos (regardless of age model, linear or logarithmic). A GEE model with linear age showed a significant effect of video (Wald's $\chi^2$ = 5534.56, $p$ ¡ .001) and a significant linear age×video interaction (Wald's $\chi^2$ = 21.66, $p$ = .001), but no main effect of linear age. In contrast, the logarithmic model did find a significant main effect of log age (Wald's $\chi^2$ = 6.16, $p$ = .013) in addition to the significant effect of video (Wald's $\chi^2$ = 245.31, $p$ ¡ .001). However, the significant logarithmic age parameter was *negative*, opposite to the global shift hypothesis prediction. Further, a significant age×video interaction (Wald's $\chi^2$ = 34.11, $p$ ¡ .001) moderated the main effect of age, casting doubt on a *global* age effect in face looking across videos.

To further explore the moderating effect of stimulus video on age-related changes for individual videos, regressions were fit using linear and logarithmic age to predict face looking separately for each video (Table 2.2). Only Video 4 demonstrated significant age-related change; linear ($R^2$ = .091, $p$ = .009) and logarithmic ($R^2$ = .240, $p$ ¡ .001) age significantly predicted face looking. Surprisingly—and contrary to the global shift hypothesis—negative changes in Video 4 indicate that face looking *decreased* with age. Moreover, the

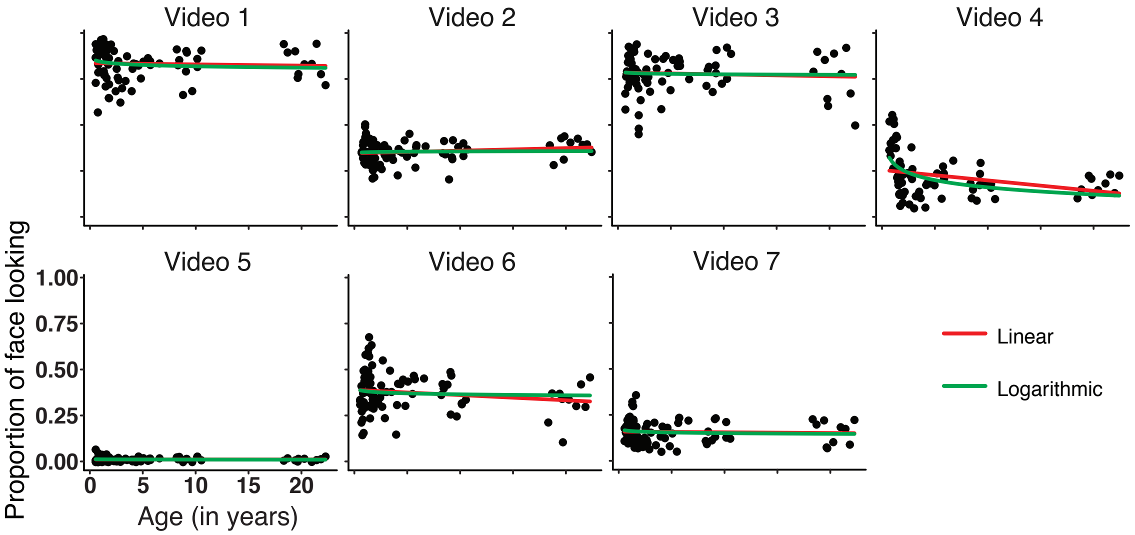


Figure 2.1: Changes in face looking as a function of age for all seven videos. Linear and logarithmic functions are plotted for each stimulus video.

lack of significant age effects across the remaining videos runs contrary to the prediction of a global increase in attention to faces.

Lastly, past research identifying the global shift hypothesis has examined changes in face looking over infancy. An additional analysis ruled out that a global shift would be found when testing only participants younger than 18 months. The linear age GEE with the restricted age range indicated no significant effect of linear age, but did indicate a significant effect of video (Wald's $\chi^2$ = 441.52, $p$ ¡ .001) and a significant linear age×video interaction (Wald's $\chi^2$ = 43.41, $p$ ¡ .001). The analysis using logarithmic age revealed a similar result. The model found no significant effect of log age but there was a significant effect of video (Wald's $\chi^2$ = 39.58, $p$ ¡ .001) and a significant log age×video interaction (Wald's $\chi^2$ = 37.17, $p$ ¡ .001).

### Gaze saliency

Similar analyses were performed for orienting to salient locations. For each participant, a composite gaze saliency score for each video was calculated by: 1) averaging the saliency ranks of pixels within a 1.2 diameter of the participant's point of gaze on every frame, and then 2) averaging across all frames in each video. Across age and stimuli, observers on average looked towards relatively more salient areas of the scene with a grand mean gaze saliency rank (out of 100) of $M = 81.53$ ($SD = 7.93$). Although consistently high, gaze saliency differed between the seven videos, with mean ranks (collapsing across age) ranging from 66.01 to 85.32. However, as is evident from inspecting the graphs in Figure 2.2, there were no consistent linear or logarithmic age-related changes in gaze saliency across videos. The linear age GEE model confirmed a significant effect of video (Wald's $\chi^2 = 628.28$, $p$ ¡ .001), but did not find significant age or age×video effects. Similarly, the logarithmic age model showed a significant effect of video (Wald's $\chi^2 = 50.51$, $p$ ¡ .001) and failed to find a main effect of age. However, there was a significant age×video interaction in the logarithmic model (Wald's $\chi^2 = 12.63$, $p = .049$), suggesting that age differences in looking at salient regions depended on the stimulus.

To further explore the age-related changes in gaze saliency for individual videos, regressions were fit using linear and logarithmic age to predict gaze saliency separately for each video (Table 2.2). For five of the stimuli, neither linear nor logarithmic changes in gaze saliency were found as a function of age. Two stimuli indicated significant fit with a logarithmic function, Video 2 ($R^2 = .055$, $p = .029$) and Video 5 ($R^2 = .126$, $p = .001$). Both videos revealed age-related increases in looking to salient areas, but effect sizes were

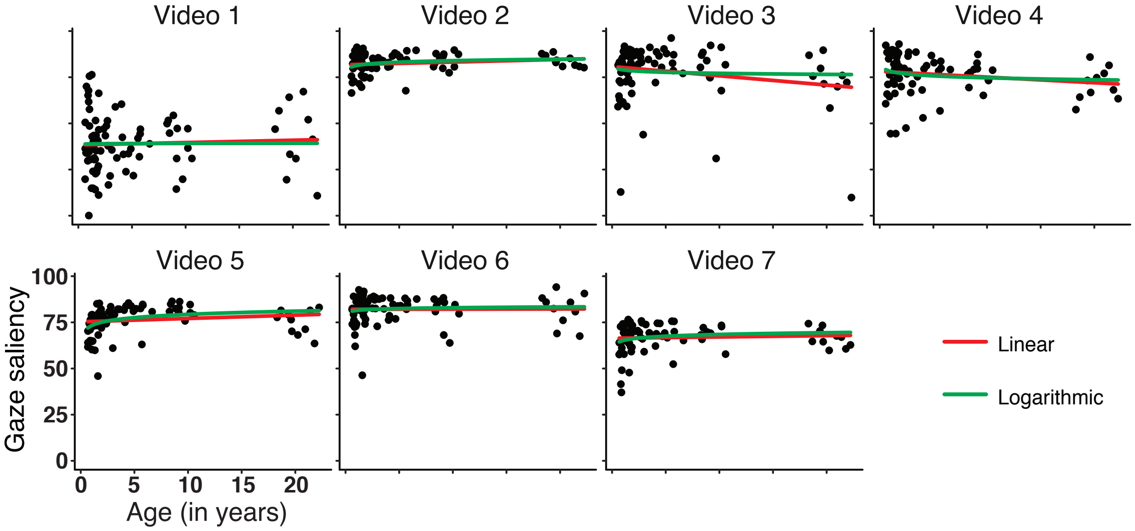


Figure 2.2: Changes in gaze saliency as a function of age for all seven videos. Linear and logarithmic functions are plotted for each stimulus video.

modest. In summary, no main effect of either linear or log-transformed age was found, which reflects a lack of a global age-related change across videos.

As with face looking, we ruled out that global changes would be found when restricting the analyses to participants ¡ 18 months. The linear age GEE found no significant effect of age, but did find a significant effect of video (Wald's $\chi^2$ = 46.35, $p$ ¡ .001) and a significant linear age×video interaction (Wald's $\chi^2$ = 16.42, $p$ ¡ .012). The GEE model using logarithmic age showed no significant effects of log age, but there were significant effects of video (Wald's $\chi^2$ = 13.03, $p$ = .042) and a age×video interaction (Wald's $\chi^2$ = 14.50, $p$ = .024).

Table 2.2: Regression parameters for linear and logarithmic age-related changes in face looking and gaze saliency for each stimulus video.

| | Face looking | | | | Gaze saliency | | | |
|---|---|---|---|---|---|---|---|---|
| | Linear | | Log | | Linear | | Log | |
| Video | $b$ | $R^2$ | $b$ | $R^2$ | $b$ | $R^2$ | $b$ | $R^2$ |
| 1 | -0.046 | .002 | -0.131 | .017 | 0.049 | .002 | 0.008 | ¡ .001 |
| 2 | 0.136 | .019 | 0.033 | .001 | 0.166 | .027 | 0.234 | .055* |
| 3 | -0.061 | .004 | -0.035 | .001 | -0.204 | .042 | -0.058 | .003 |
| 4 | -0.301 | .091* | -0.490 | .240* | -0.177 | .031 | -0.147 | .022 |
| 5 | -0.016 | ¡.001 | -0.059 | .003 | 0.146 | .021 | 0.354 | .126* |
| 6 | -0.155 | .024 | -0.073 | .005 | 0.006 | ¡ .001 | 0.086 | .007 |
| 7 | -0.038 | .001 | -0.088 | .008 | 0.059 | .003 | 0.198 | .039 |

*$p < .05$

### 2.5.2 Within-stimulus variability moderates age differences in visual attention

Next, we tested for consistency in age differences in looking towards faces and salient areas over time within each stimulus. One possibility is that age-related increases in looking to faces occur during particular moments within the videos (and null effects of age at other moments), consistent with the global shift hypothesis. A second possibility is that the direction of age differences to each feature changes as a function of time (e.g., adults looking more at faces/salient areas compared to infants/children at one time and less at faces/salient areas at another time). Such inconsistent age differences would provide strong evidence against the global shift hypothesis. To test these possibilities, we used a sliding window analysis to measure the differences in attention to visual features as a function of age at different points of time within each video. Each 2-min video was segmented into 10-s windows that were distributed evenly throughout the video. The first window started at the beginning of the video, and each subsequent window was placed 5 s after the start of the previous window resulting in 22 overlapping windows. Face looking and gaze saliency were recalculated for each participant within every 10-s window.

We calculated separate GEEs for each video to predict attention to each visual feature (face looking, gaze saliency) based on window (as a factor) and age (as a continuous predictor). Significant age×window interactions would suggest that attention to visual features differed by age over the course of the video. Such interactions would indicate features had a differential, age-dependent influence on visual attention at different points in a video as the scene content changes. Significant age×window interactions were followed up with separate correlations between age and visual features to determine the **direction** and **strength** of the age difference within each window. The consistency of the direction of age correlations was of key interest in differentiating between the possibilities above. Because there was greater evidence in the prior section for logarithmic effects of age, we used log-transformed age in the models testing these possibilities and in follow-up correlation tests. We tested parallel models using linear age; however, we omitted those results for brevity because there were no substantive differences that would affect the interpretation of the findings.

### Changes in scene content within a video moderated the direction of age effects on face looking

For all videos, the relationship between age and face looking varied significantly from window to window. Figure 2.3 shows the fluctuating relationship between age and face looking across windows of each video. To better illustrate the differential effects of age, separate lines are plotted to show face looking values for infants (6- to 24-months), children

(2- to 10-years), and adults. However, in GEE models and correlations, logarithmic age was analyzed as a continuous variable: In Figure 2.3, the inset figures illustrate the continuous functions underlying three exemplar windows. The direction and strength of correlations between age and face looking are represented in Figure 2.3 based on color shading over each time window. As evidence of the changes from moment to moment, windows within each stimulus show age effects that vary both in strength and direction (red bars indicate less looking to faces with age whereas blue bars indicate more looking to faces with age).

Seven GEE models were calculated (one for each video) to test for the effects of age, window and interactions between age and window on face looking. Table 2.3 shows that for all seven videos, there were significant age×window interactions, indicating varying age-related differences in face looking from moment to moment. We also found a significant effect of window for all 7 of the GEE models indicating significant mean level fluctuations in face looking as the scene content changed within these videos, irrespective of age. Lastly, only one main effect of log-transformed age was found (Video 4), as was observed in the stimulus-level analysis of face looking in the previous section.

Follow-up analyses explored age×window interactions by measuring the correlation between logarithmic age and face looking for each time window (e.g., the $r$ values depicted in Figure 2.3). Figure 2.4A shows a frequency distribution of every correlation between age and face looking for each window aggregated across the seven videos. Two findings emerge from examining the distribution of correlations across videos. First, the presence of both positive and negative correlations indicate that there are both times in which adults look more to faces than infants but also other windows when infants attend more towards

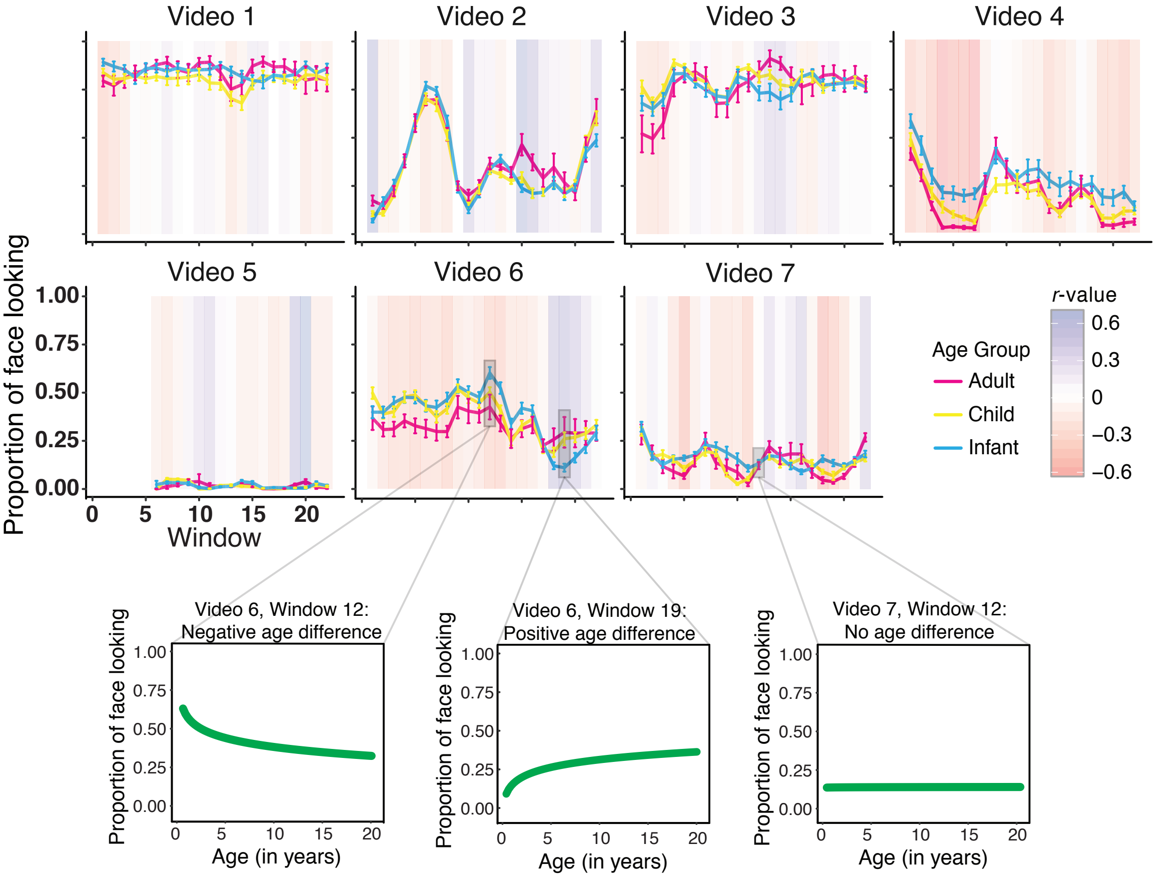


Figure 2.3: Windowed analyses of age differences in face looking over the duration of all 7 videos. Age was analyzed as a continuous variable, but for illustration purposes age was averaged into three groups (infants: 6-24 months; children: 2-10 years; adults: 18-22 years). Colored vertical bars represent strength and direction of correlation between age and face looking for every window. Darker colors indicate stronger correlations. No data are plotted for the first 5 windows of Video 5 because no faces were present during that portion of the video. Insets depict examples of 3 individual windows to show a negative correlation, positive correlation, or no correlation between age and face looking with age represented as a continuous predictor.

faces than adults. The second finding from these correlations is that a relative minority of windows show statistically-significant correlations. This indicates that infants, children, and adults more often prioritized faces in a similar rather than a different way.

However, given the total number of windows in which correlations were calculated, the probability of spurious correlations is high. To estimate the expected range of correlation values due to chance, we created a bootstrapped null distribution by randomly re-assigning the age labels to each eye movement time series (Figure 2.4B). Ages were randomly shuffled within videos but not for each window in order to preserve the temporal ordering between windows for any given participant. Age correlations with face looking were recalculated for the randomly shuffled data. This was repeated 1000 times to produce a null distribution of correlations between age and face looking. Figure 2.4B shows this distribution of randomized correlations with vertical lines indicating the range in which 95% of the correlations occurred. Next, we determined how many correlations in the observed data were more extreme compared to the 95% range from the null distribution (arrows on the x-axis of Figure 2.4A). Only 5% of the correlations should fall outside of this range by chance, however, in the observed data 26.8% fell outside of the range. Using the *’multicon’* package in R (Sherman & Serfass, 2015), we conducted a randomization test that confirmed the number of significant correlations found was greater than chance. The test indicated that there was significant difference between the expected number of significant correlations ($M$ = 7.43, $SE$ = 4.04) and the 40 observed statistically significant correlations, $p$ ¡ .001. This indicates that the prevalence of significant correlations is not spurious and points to real age-related differences in orienting to faces. Moreover, of the 40 significant correlations, 28 were negative and 12 were positive, providing further evidence against the notion of a global increase in face looking with age.

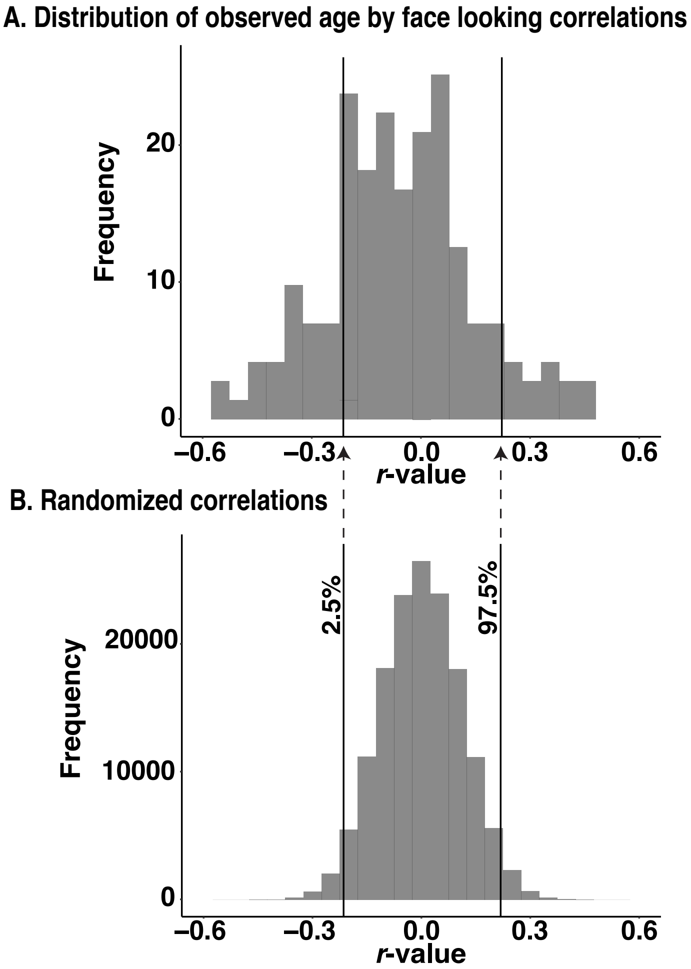


Figure 2.4: (A) Observed distribution and (B) randomized null distribution of correlations between age and face looking for each window aggregated across videos. Vertical black lines mark the 95% range of expected correlations in the null distribution.

### Age-related differences in gaze saliency from moment to moment

Similar to the patterns observed in face looking, for most stimulus videos, the relationship between age and gaze saliency varied significantly from window to window. Figure 2.5 shows the changing relation between age and gaze saliency across windows in each of the videos. Again, for illustrative purposes, separate lines are plotted to show gaze saliency means for infants (6-24 months), children (2-10 years), and adults; however, logarithmic age was analyzed as a continuous variable in GEE models. As with face looking, variation in the relation between age and face looking over windows provides evidence of age-related changes from moment to moment.

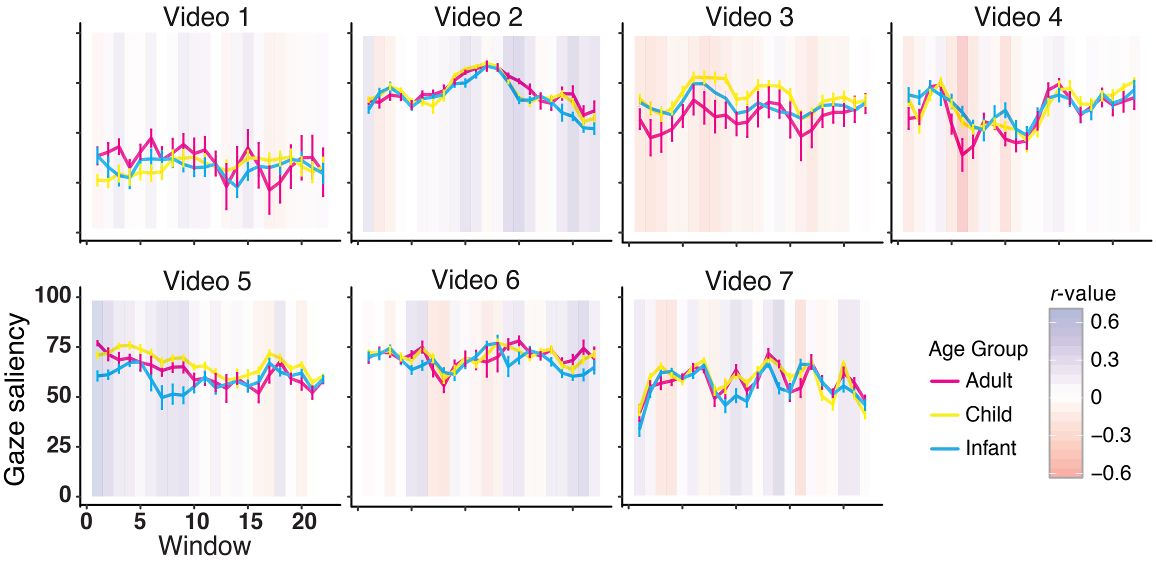


Figure 2.5: Windowed analyses of age differences in gaze saliency by video stimulus. Age was analyzed as a continuous variable, but for illustration purposes age was averaged into three groups (infants: 6-24 months; children: 2-10 years; adults: 18-22 years). Colored vertical bars represent strength and direction of correlation between age and gaze saliency for every window. Darker colors indicate stronger correlations.

Similar to face looking, seven GEE models were calculated to test for effects of age, window, and age×window interactions on gaze saliency. As shown in Table 2.3, significant age×window interactions were found for all seven videos, indicating varying age-related differences in gaze saliency from moment to moment. In addition to the significant interactions, there were significant main effects of window for 6/7 videos indicating mean level differences in gaze saliency over time irrespective of age. Finally, as seen in the previous analyses, Videos 2 and 5 showed main effects of log age.

Table 2.3: Generalized Estimating Equation Wald's $\chi^2$ for effects of window, age, and age×window for each video stimulus.

| | Face looking | | | Gaze saliency | | |
|---|---|---|---|---|---|---|
| Video | Window | Age | Window×Age | Window | Age | Window×Age |
| 1 | 57.88* | 1.20 | 56.47* | 45.20* | .001 | 44.31* |
| 2 | 310.42* | .20 | 179.47* | 116.96* | 7.88* | 75.16* |
| 3 | 35.48* | .22 | 36.68* | 29.21 | .00 | 41.74* |
| 4 | 121.95* | 21.82* | 103.71* | 75.95* | .15 | 77.49* |
| 5 | 38.39* | .12 | 39.71* | 116.04* | 11.53* | 120.89* |
| 6 | 188.12* | .38 | 113.27* | 123.38* | 1.79 | 115.81* |
| 7 | 122.98* | 1.71 | 142.09* | 188.89* | 3.72 | 114.90* |

*$p < .05$

We explored the age×window interactions by examining the distribution of all correlations between age and gaze saliency (Figure 2.6A). As with face looking, the distribution of correlations clustered around $r = 0$, indicating that participants across ages more often prioritized salient locations in a similar way. We created a null distribution of correlations based on 1000 iterations of reshuffling age labels and eye movement data. Figure 2.6B depicts this distribution, with vertical lines delineating the middle 95% of the data. The original observed data was compared to the 95% range in the null distribution to determine whether the number of significant windows could be due to chance. As seen in Figure 2.6A, 22.07% of the observed correlations exceeded the 95% range of the null distribution. Using the *'multicon'* package in R (Sherman & Serfass, 2015), we conducted a randomization test that confirmed a significant difference between the average expected number of significant correlations due to chance ($M$= 7.4, *SE* = 4.09) and the 34 statistically significant correlations observed in the study, $p$¡.001. Unlike the age-face looking correlations, significant age-saliency correlations tended to be positive (30/34) rather than negative (4/34).

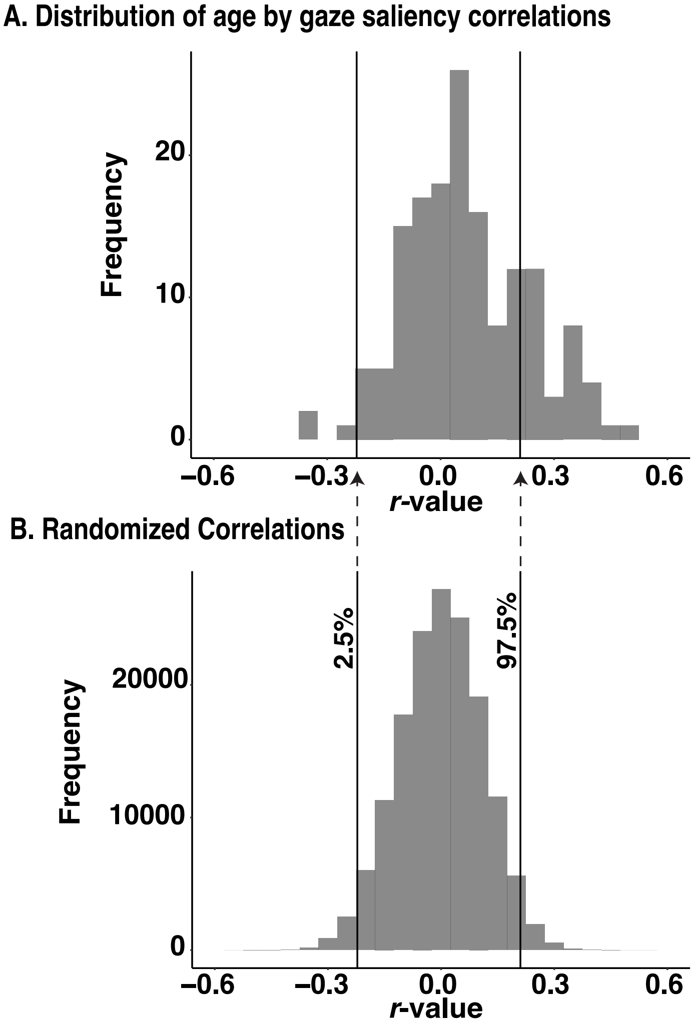


Figure 2.6: (A) Observed distribution and (B) randomized null distribution of correlations between age and gaze saliency for each window aggregated across videos. Vertical black lines mark the 95% range of correlations in the null distribution.

## 2.6 Discussion

The current study measured the eye movements of infants, children, and adults across and within seven videos to test the global shift hypothesis. No global shift in attention was discerned at any level of analysis. We found no consistent age-related changes in looking to faces or to visually-salient locations across videos. Of the seven videos, only two videos showed modest age-related increases in gaze saliency and only one video showed an age-related *decrease* in face looking. No video showed an age-related increase in face looking. However, there were moment-to-moment age differences in looking at both faces and salient locations within videos.

These findings suggest that the global shift hypothesis does not appropriately capture the nuances of developmental change in visual attention. Age differences in looking at both types of visual features only emerged at shorter time scales. Sliding window analyses revealed that the relation between age and each visual feature was in constant flux: For some time windows, age was correlated with face looking and gaze saliency, but for other windows participants of all ages attended to features in a similar way. When age did predict differences in face looking, we found both positive and negative correlations, suggesting that age differences were not global but rather depended on different prioritization of faces according to age. Sometimes adults looked more often at faces, but other times infants looked more often at faces.

### 2.6.1 Lack of global changes in attention

The lack of overall age-related changes in visual attention to salient areas and faces differs from many prior studies that found such effects: increases in face looking (Franchak et al., 2016; Frank et al., 2009; Kwon et al., 2016; Amso et al., 2014), decreases in looking to salient areas (Helo et al., 2014; Kwon et al., 2016; Açik et al., 2010), or increases in looking to salient areas (Rider et al., 2018; Franchak et al., 2016; Frank et al., 2009). Many of these studies focused on developmental changes that occur during infancy, which could explain a discrepancy in findings. However, when restricting analyses to participants ¡ 18 months, we still did not find global age-related change for either feature. There are several other differences between the current study and past work that may explain conflicting findings. One potential explanation is the duration of the selected stimuli. Each video clip used in the current study was 2 min, but most past studies used either static images or video stimuli that were shorter in duration: one 60-sec video (Franchak et al., 2016); twelve 20-sec videos (Frank et al., 2012); twelve 4-sec videos from one television program (Stoesz & Jakobson, 2014); 24 4-sec clips from a single video (Frank et al., 2009)). Averaging looking behavior over short stimuli likely misses the heterogeneity present in longer videos that would yield evidence of moment to moment changes. Indeed, our windowed analyses indicate that there is over a 20% chance of randomly picking a 10-s window from our stimuli that would show an age-related change in face looking. Thus, studies that use only one stimulus or a few stimuli with short durations may be at risk for selection effects that could lead to incorrect generalizations.

Using images and short videos may also capture unique age-related differences in early scene inspection that are not characteristic of visual attention more broadly. Within the first few seconds of examining a new scene (e.g., following a cut), adults move from frequent, quick fixations to longer fixations associated with inspecting objects while infants persist slightly longer with rapid fixations (Helo, Rämä, Pannasch, & Meary, 2016). Other studies have found a bias in adults, but not young infants, towards looking at the center of the scene immediately following a cut or at the onset of a stimulus (Mital et al., 2011; Kirkorian et al., 2012; Wang et al., 2012). This has been attributed to an adult viewing strategy that expects screen-based media to center relevant information in the image frame

(P.-H. Tseng, Carmi, Cameron, Munoz, & Itti, 2009). Therefore, studies that use short stimuli or stimuli with frequent scene cuts may see biases that result from age differences in early scene viewing.

However, other studies used longer stimuli, such as three 5-minute videos (Rider et al., 2018) and two 2-minute videos (Frank et al., 2014), and did find consistent age-related changes in the influences of saliency and/or faces on eye movements. However, since neither study systematically tested for changes in attention to saliency/faces on shorter timescales within each video, it is unclear whether the age differences at the video level are due to consistent effects over time or from local effects confined to particular times. Indeed, in the few videos that showed overall age differences in the current study, it was clear on closer inspection that those overall effects were in fact driven by differences in how adults or infants selected faces or salient locations for a few time windows as opposed to consistent effects across the entire video.

Could the lack of a global age-related difference in gaze saliency be the result of the saliency algorithm being tuned to adults? We would argue that the opposite is true. If the saliency algorithm was a better measure across the board for adults compared to younger participants, we would expect to see higher gaze saliency values in adults for every video. Instead, we found that for many videos there was no substantive difference in saliency across ages. This suggests that even though the saliency model is trained on adult data, it performed similarly when applied to infant and child data for the majority of the time.

The lack of global age-related changes in the current study may be a consequence of the particular video content we selected. Wass and Smith (2015) found that television programs designed for toddlers more often contain a speaking character whose face is salient compared with programs designed for adults. It is possible that past studies used more infant or child focused video content, which could bias looking towards both faces and salient areas. In the current study, we selected the seven videos to provide diverse content that would be engaging to participants across the ages we tested, which included media designed both for children and for adults. Yet, delineations in our stimuli between videos designed for adults or children provide no insight to why particular videos showed age-related changes. For instance, age-related changes in saliency were observed for both child-directed (Video 2) and adult-directed (Video 5) stimuli. Diversity in video content provides the opportunity to investigate how other properties may explain the results we present. However, the challenge with this type of post hoc approach is identifying which of the countless properties that vary between videos or scenes can explain the findings. Possibly, diversity in the content we chose accounts for why there were no consistent age-related changes across videos. Past studies might have found more consistent effects because stimuli were homogeneous in content.

Finally, the use of dynamic versus static stimuli in the current study versus past investigations may account for differences in face looking. Many (but not all) of the studies that found consistent age-related trends in face looking used static images (Helo et al., 2014; Kwon et al., 2016; Açik et al., 2010; Amso et al., 2014), whereas studies that found inconsistent effects of face looking (Frank et al., 2009; Franchak et al., 2016) used dynamic videos. Recent work demonstrated that face-looking preferences are greater in static as opposed to dynamic stimuli (Libertus, Landa, & Haworth, 2017; Stoesz & Jakobson, 2014). Faces may be the most relevant place to look in a static image, but in videos that display complex actions involving hands and objects, faces may less often be the most important location. Beyond screens, real-life visual attention is not only used for passively viewing events but for actively controlling movements. Accordingly, infants infrequently look at caregivers' faces and spend more time looking at objects (Yu & Smith, 2013; Franchak, Kretch, & Adolph, 2018)—presumably to support object-related manual actions. Thus, previously-measured global changes in looking to faces may be a byproduct of using less ecologically-relevant stimuli, such as static images, that do not convey as much information about action.

### 2.6.2 Development of visual attention involves changes in prioritizing features

How, then, does visual attention to faces and salient features develop? We argue that children become better able to *prioritize* which features to attend to—whether faces or salient locations—depending on the particular content in a scene. Prior work has shown that adult observers prioritize which visual features to attend to based on their importance within a scene or their relevance to a task (Franchak et al., 2016; Henderson & Hayes, 2018; Henderson, 2017; Ballard & Hayhoe, 2009; Rothkopf, Ballard, & Hayhoe, 2007; T. J. Smith & Mital, 2013). The current study provides evidence that infants and children often, but not always, prioritize visual features in a similar way as adults. At the overall video level, age differences in gaze saliency and face looking were marginal. At the window level, most time periods within videos showed no age differences. Since gaze saliency and face looking changed greatly from moment to moment, this suggests that even observers as young as 6 months are responding to changes in feature relevance in a similar way as older children and adults. The most striking example is the change in face prioritization over windows 2-10 of Video 2; Figure 2.3 shows that face looking jumps from 13% to 72% and then back to 13% in a short time for participants of every age. Related work found that increasing homogeneity in infants' eye movements patterns within age groups could be explained by increasing similarity to adults' eye movement patterns, suggesting a quantitative rather than qualitative change in how visual features attracted attention over development (Franchak et al., 2016). Similarities between infants' and adults' prioritization is also consistent with prior work showing that many other aspects of visual attention are mature by 6 months of age (Oakes & Amso, 2018), with some visual processing abilities reaching adult-like levels: scanning and fixations to simple shapes (Bronson, 1994), configural face processing (Cashon & Cohen, 2004), and perception and discrimination of object features (Colombo, Mitchell, Coldren, & Atwater, 1990).

Despite similarities in how infants and adults prioritized faces and salient locations, age differences in attention to each feature could be detected for some time windows. Indeed, all seven videos showed age×window interactions for both visual features. Developmental differences in prioritization are consistent with past work showing that age moderates the degree to which infants' face looking and gaze saliency varied across different types of scenes (Frank et al., 2012; Franchak et al., 2016). The current study extends these findings by showing that changes in prioritization are evident from infancy through childhood. Moreover, these differences emerged when scenes were defined in an objective way—evenly-spaced time windows that are agnostic to video content—rather than an ad hoc way—defining scenes based on particular content features. Furthermore, the current study is unique in showing that infants' prioritization differs from adults' both in looking *less often and more often* at visual features depending on the time window. Thus, the developmental difference in prioritization cannot be explained by a global deficit in selecting (or inhibiting) a particular feature type.

What might account for developmental differences in prioritization? First, temporal and spatial changes in attention may account for age differences in prioritizing features. Although some aspects of attention are nearly adult-like in the youngest participants we tested, other aspects are not. Infants' temporal processing, or the rate at which infants are able to isolate individual changes in a stimulus, is much coarser than adults' (Farzin, Rivera, & Whitney, 2011). In a dynamically changing scene, infants may be slower to change their prioritization of visual features to reflect what is important from moment to moment. Additionally, the development of endogenous attention—that is, the ability to exert voluntary control to select and inhibit where to attend—shows protracted improvements throughout infancy and early childhood (Colombo, 2001; Oakes & Amso, 2018). For example, children's ability to sustain attention to a particular target while inhibiting distraction from other targets improves from 2.5 to 4.5 years (Ruff, Capozzoli, & Weissberg, 1998). Note that these same changes in attention motivate the global shift hypothesis—that is, increasing endogenous control allows infants to inhibit looking to irrelevant, salient areas while actively selecting faces. However, the current results suggest something more subtle: Increasing endogenous control allows infants to better prioritize information by inhibiting competition from faces and/or salient regions while sustaining attention towards locations they deem informative, whatever those might be.

The second possibility is that differences in prioritization reflect developmental changes in how infants and children comprehend scene content and determine which locations are most informative. Deficits in infants' understanding of media are especially notable, as children under 24 months fail to even notice when scenes in a video narrative are presented in a scrambled order (Pempek et al., 2010). Such deficits in scene comprehension are likely a key factor that accounts for differences in how infants and children distribute eye movements while watching videos (Franchak et al., 2016; Helo et al., 2017; Kirkorian et al., 2012; Kirkorian & Anderson, 2018). It is important to note that in the current study we analyzed overall rates of face looking irrespective of *which* face observers fixated. Many scenes had multiple faces in view, so it is possible for observers of different ages to have similar face-looking rates while attending to different targets. Moreover, face-looking rates could be similar for two observers who looked at the same face for the same duration but at different times (even in the short, 10-s windows). Thus, it would be incorrect to interpret similar face-looking rates (and gaze saliency scores) between observers or between age groups to indicate similar comprehension of the scene. A more nuanced analysis of synchrony in looking at specific faces at specific times might bear on this issue; however, this was beyond the scope of the current investigation.

Finally, attention and comprehension likely interact in several ways which would lead to age-related differences in viewing behavior. First, prior research shows that children's gross attention to media depends on their understanding (Anderson et al., 1981; Lorch & Castle, 1997): Children are more prone to distraction and visually attend less while watching content that is beyond their comprehension. Although we excluded participants who had large missing sections of gaze data, it is still possible that lower engagement in younger participants who did not understand what they were watching could have impacted their overall attention. Looking away from the video would prevent observers from monitoring key visual targets in the scene and disrupt following the narrative. Second, prior work shows age-related differences in how salient visual features interact with understanding of scene content in determining where observers look. For example, when viewing static images altered to include inconsistent objects (i.e., a bar of soap on a kitchen table), adults spend long periods fixating inconsistent objects regardless of their saliency but 24-month-olds only do so when those objects are visually salient (Helo et al., 2017).

### 2.6.3 Implications for attention development and media viewing

In sum, the current study demonstrates that the developmental changes in eye movements while watching complex, dynamic stimuli reflect age differences in how observers prioritize different features as opposed to a global age-related shift in the selection of specific features. The results from this study add to a growing literature showing that singular feature based-approaches are insufficient to capture the complexity in gaze allocation (Tatler et al., 2011; Sailer, Flanagan, & Johansson, 2005; Henderson & Hayes, 2018; Land & McLeod, 2000; Pereira, Birmingham, & Ristic, 2019). What is meaningful in a scene changes dynamically and may not predictably map on to distinct visual features, making it challenging to determine why observers prioritize locations in a particular way. More work is needed to map out the degree to which changes in attention and/or comprehension account for developmental changes in prioritizing where to look. The current study makes an informative methodological contribution in showing that variability is the rule, not the exception. Improving our understanding of how visual exploration changes with development will depend on studying a wider array of complex stimuli (and real-world situations) and analyzing gaze behavior across different timescales.

Furthermore, as the first study to compare eye movements across a large sample, wide age range, and large, diverse set of video stimuli, our results have broad implications for understanding infant and child viewing of screen-based media. Since media viewing is a common and frequent childhood occurrence, it is important to understand how changes in visual attention might contribute to children's understanding of screen-based media. One implication is that the challenge children face in learning the 'right' features is more complex than previously thought—there is no 'one size fits all' solution because the relevance of different features is in constant flux. Still, our work raises potential avenues for designing media to improve comprehension. First, designers of children's media could restrict how often particular features change in relevance over time to improve children's comprehension. Second, children should benefit from scenes in which different types of features converge rather than compete (Wass & Smith, 2015; Amso et al., 2014) to reduce the pressure on prioritization. Future work should seek to test children's learning from video clips that systematically vary the need to change prioritization of visual features over time to track key educational content.

## 2.7 Acknowledgements

I acknowledge that Chapter 2 of this dissertation was published in Developmental Psychology in November of 2020

# Chapter 3

# Attention to hands during manual actions account for developmental increases in attentional synchrony

## 3.1 Abstract

Consistency of attention, both spatially and temporally, increases with age from infancy into childhood, when viewing dynamic stimuli (Franchak et al., 2016; Kirkorian et al., 2012). We propose that developing increasingly synchronous visual attention involves improvements in prioritizing meaningful information at each moment. The current study is a secondary analysis of Kadooka and Franchak (2020) in which eye movements of 79 children (6 months to 10 years) and 20 adults were recorded as they viewed five dynamic videos. We tested whether infants' and children's synchronous attention to two meaningful locations, hands and hand-object actions, could account for age-related increases in adult-like attention. Improvements in looking to these hand features may indicate a convergence towards adult-like prioritization of meaning. Findings show that the degree to which infants and children look to hands and hand-object actions can account for increases in similarities to adult attention, beyond age alone. In considering the spatiotemporal variability of meaning in a scene, we suggest that attention development involves changes in looking to semantically-relevant information, broadly.

## 3.2 Attention to hands during manual actions account for developmental increases in attentional synchrony

The visual attention system involves a complex process of prioritizing different information from moment to moment. When passively watching a scene unfold in the real world, we are constantly moving our eyes to direct our attention, roughly three times per second (Schiller, 1998). Looking out a window, an observer may allocate their attention towards a busy walkway, leaves fluttering in the wind, the face of a familiar acquaintance, the iridescent flash of a bluebird or any number of nearly limitless possibilities. Yet, with all these potential targets of attention, adults tend to show high eye movement synchrony in where they look when viewing dynamic scenes (Dorr et al., 2010; Franchak et al., 2016; Hart et al., 2009; Mital et al., 2011; Shepherd et al., 2010; Wang et al., 2012). This level of synchrony in attention is not present in infants. But over the first few years of life, infants increasingly develop greater synchrony with adults (Franchak et al., 2016; Kirkorian et al., 2012).

How do infants and children achieve more adult-like gaze behaviors? As we will discuss in subsequent sections, there are inconsistent findings about whether attention to certain types of visual features contributes to the development of adult-like gaze. In the current study, we test whether two understudied features – hands and hand-object actions – can account for age-related increases towards greater adult synchrony. Hands are a rich source of social, communicative, and semantic information. Furthermore, prior studies indicate that infants and children increasingly extract meaning from hands: social communication and word learning via gestures and pointing (Tomasello et al., 2007; de Villiers Rader &

Zukow-Goldring, 2010), joint attention by following hands (Yu & Smith, 2013; Deak et al., 2014), the affordances of tools from grip positions (Barrett, Davis, & Needham, 2007), and the goals and intentions of actions (Woodward, 2009, 1998). It is unknown whether these abilities to understand the meaningfulness of hands translates to greater attentional synchrony to these features when watching dynamic scenes. If visual attention development involves better prioritization of meaningful information, then increased synchrony to these meaningful features could partially account for the age-related changes towards greater adult synchrony.

### 3.2.1 Developmental changes in synchronization of attention

Adults exhibit highly correlated gaze behaviors in looking to similar places at similar times (Dorr et al., 2010; Franchak et al., 2016; Hart et al., 2009; Mital et al., 2011; Shepherd et al., 2010; Wang et al., 2012). By recording the eye movements of adults when watching the same stimulus, researchers can calculate similarity in gaze location compared to other observers across the duration of the stimulus, which is known as the inter-subject correlation or ISC. Higher ISCs indicate greater spatiotemporal similarity in gaze. Adults, when compared to other adults, show high ISCs across a wide range of stimuli. In commercially produced films, adult attention is similar likely because directors and editors have designed their films with the explicit goal of guiding attention to specific information (Dorr et al., 2010; Gannon & Grubb, 2022; Goldstein et al., 2007). But even in more naturalistic live-action stimuli (Shepherd et al., 2010; Franchak et al., 2016; Rider et al., 2018), animated videos (Rider et al., 2018), and virtual reality experiences (Farmer et al., 2021), adults show synchronous eye movements.

Past research reveals age-related increases towards more consistent eye movements when watching dynamic videos (Franchak et al., 2016; Frank et al., 2009; Kirkorian et al., 2012). One way to examine this is to measure developmental differences in the spatial distribution of gaze. Kirkorian and colleagues (2012) found that several metrics of spatial variability decrease with age (i.e., became more similar) when comparing the gaze of 1-year-olds, 4-year-olds, and adults while watching a video from Sesame Street, a live-action children's program. In a different study, Frank, Vul, and Johnson (2009), found a similar trend for 3-month-olds, 6-month-olds, 9-month-olds, and adults when watching Charlie Brown, an animated television series. Eye movements were defined by less spatial variability as age increased. However, less spatial variability does not necessarily mean greater synchrony with adults as it is possible that children may increasingly look to a feature that adults do not look at. Franchak et al. (2016) directly addressed this by comparing ISCs between infants and adults as they watched a Sesame Street music video. By calculating the ISCs between an infant observer and a comparison group of adults, these correlations measured how adult-like their eye movements were. Synchrony with adults increased with age for infants ranging from 6-months to 24-months old.

Age significantly predicts increases in adult synchrony (Franchak et al., 2016), but age alone holds limited explanatory value in determining what is changing. Age-related changes to several features have been proposed to partially account for developmental changes in attention. In studies with dynamic stimuli, the influence of visually salient locations (e.g. color, motion, contrast) increased with age (Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018). Other studies have identified age-related changes in attention to socially relevant features and faces (Franchak et al., 2016; Frank et al., 2009, 2012, 2014; Stoesz & Jakobson, 2014). However, the content of the scene cannot be ignored. Scene factors like the number of agents (Franchak et al., 2016; Frank et al., 2012; Stoesz & Jakobson, 2014), centering of faces (Franchak et al., 2018), and dialogue (Frank et al., 2014) all influence the relationship between attention to faces and age. This suggests that scene content influences what is most meaningful. Neglecting to account for this shifting relevance leads to an incomplete characterization of how attention is developing.

Therefore, we posit that synchrony with adults is not a global developmental shift in attention to features, but rather a change in prioritization of features depending on their relevance. For instance, an attentional bias for looking at objects was observed for 4-6 year old children, yet adults only looked to objects if it was relevant for the task (Darby, Deng, Walther, & Sloutsky, 2021; Spelke, 1990). Furthermore, adult gaze behavior when viewing a static image is reliably predicted by judgements of semantic meaning (Henderson et al., 2007; Henderson, 2017; Henderson & Hayes, 2018). For dynamic scenes, the meaningfulness of any given feature may change from moment to moment. Indeed, examining smaller 10 second windows of time within video stimuli revealed that the direction of age-related differences in attention to both salient regions and faces were in constant flux (Kadooka & Franchak, 2020). In any given stimulus, sometimes adults may look more towards a feature, yet at other times, infants look more towards that feature. Developing adult synchrony occurs when infants and children become more adult-like in their concept of what features are relevant and when they are relevant. Importantly, age-related changes in attention to particular features are still helpful in accounting for developing adult synchrony, however, particular consideration must be taken towards measuring attention when those features are meaningful.

### 3.2.2 Age-related changes in attention to hands and hand-object actions

In the current study we selected two features that we predicted would account for age-related changes in synchrony with adults: hands and hand-object actions. Hand-object actions are times when hands actively interact with objects in moving, manipulating, or other goal-directed actions. These features were chosen for three reasons: a) hands and hand-object actions convey meaning, b) infants and children improve in their ability to detect meaning from these features, c) hands and hand-object actions are spatiotemporally identifiable. These characteristics make hands and hand-object actions good candidates for investigating adult synchrony of attention. If attentional synchrony involves becoming more adult-like in prioritizing meaning, then changes in ability to extract meaning from hands should lead to greater attentional synchrony with adults.

Hands are meaningful as a primary way that our bodies act on the world. The wide range of prehensile dexterity in the human hands is a defining feature of our species that sets us apart from other primates (Napier, 1956). Hands serve social, communicative, and semantic functions. When hands and objects interact, additional information is communicated about the objects (P. Tseng et al., 2012), affordances (Klatzky et al., 1989), goals and intentions (Zacks & Tversky, 2001). An adult neuroimaging study has identified a hands-specific region in the extrastriate body area of the visual cortex that is sensitive to hands but not other non-hand body parts (Bracci, Ietswaart, Peelen, & Cavina-Pratesi, 2010), highlighting the importance of hands.

What evidence is there for age-related changes in attention to hands? A head camera study from Fausey et al. (2016) found that during everyday activities of infants between 1 and 24 months old, the distribution of hands in view shifts towards an increasing amount of hands, especially the hands of other people. One possibility is that this shift reflects normative motor development that occurs in the first two years of life, like reaching and independent sitting, which may structure daily experiences towards seeing more hands. It is also possible that this may be driven by an infant's own active selection of hand stimuli. It is likely that both the ability to control attention and exposure to hand stimuli help to regulate attention to hands. Frank et al. (2012) found increases in looking to hands in dynamic stimuli for 3- to 30-month-olds especially when the stimulus contained manual actions with objects, indicating an active selection for certain hand stimuli.

Hands may provide information and structure to visual scenes by guiding infants' attention to locations that are meaningful especially within a social context. For instance, pointing and gesturing facilitate early word learning by offering a synchronous cue that connects language to objects in the world (de Villiers Rader & Zukow-Goldring, 2010; Tomasello et al., 2007). Evidence also suggests that looking to hands is important in the development of joint attention between caregivers and 1-year-olds, particularly during goal-directed actions with objects (Yu & Smith, 2013). When engaged in goal-directed actions, visual attention of the actor is tightly coupled with the action being performed. Therefore, when watching caregivers complete actions, infants can coordinate their attention spatially and temporally with the attention of their caregiver by looking to the hands. In fact, the gaze of the actor, hands, and objects may provide redundant spatiotemporal information about attention that allows scaffolding for more sophisticated social abilities in joint attention and gaze following (Shepherd, 2010). During the first two years of life, infants develop the ability to extract information about tool affordances based on hand grip positions (Barrett et al., 2007) and are sensitive to goals and intentions based on their own and the observation of other's actions (Woodward, 1998, 2009).

It is clear that attention to hands increases with age and children are sensitive to the wide range of meaningful information that is conveyed by hands. However, it is not clear if these changes in the ability to extract information to hands leads to prioritizing hands as a meaningful region both spatially and temporally. Synchronizing attention to hands or hand-object actions may provide a developmental scaffold for prioritizing meaningful information in ways that are more adult-like. Certainly, hands and hand-object actions are not the only features that influences adult synchrony. But the spatial discernibility of physical hands and hand-object actions over time makes these features easily definable compared to other spatiotemporally sensitive features that convey meaning like goal intentions (e.g. the location that an agent plans to place a object), distal referents (e.g. the referent in a point or gesture), or locations of an agent's attention. For these reasons, we investigated whether synchrony in looking to hands (hand synchrony) and hand-object actions (hand-object synchrony) can account for age-related increases in adult synchrony beyond age.

### 3.2.3 Current Study

The primary aim was to test whether synchronous attention to hands and hand-object actions can predict changes in attentional synchrony with adults. We performed secondary analysis of data from Kadooka and Franchak (2020) in which eye movements were recorded during children's (6 months to 10 years) and adult's free-viewing of complex, dynamic video stimuli. Past work characterizing visual attention highlights the need to capture the moment-to-moment prioritization of what is meaningful. Averaging attention to a feature across an entire stimulus may artificially suppress effects if those features were not meaningful over the entire duration of the stimulus. To address this, we selected two features (hands and hand-object actions) and determined the times that each feature occurred on screen. Past studies have also recognized that different scene content may lead to different attention patterns (Franchak et al., 2016; Frank et al., 2009). Explanations of visual attention development that are specific to the stimulus are limited in utility. Therefore, our use of the previously collected dataset takes advantage of the long duration (two-minute videos) and variability (5 different live-action child friendly videos) of stimuli to ensure that detected effects are invariant to stimuli.

First, we replicate past work on the age-related changes in *adult synchrony*. As a robust trend (Frank et al., 2009; Kirkorian et al., 2012) and direct methodological replication (Franchak et al., 2016), we expect to find attentional synchrony with adults will increase with age. In other words, infants and children will increasingly look to the same locations at the same time as adults. Past work in attentional development has found logarithmic patterns associated with age in which rapid changes in infancy are followed by slower changes into late childhood and beyond (Amso et al., 2014; Kadooka & Franchak, 2020). Therefore, a logarithmic model of age will be tested. If our expectations are met, adult synchrony will be used as a baseline to determine the unique contribution of other age-related factors above and beyond age.

Second, we will assess whether attention to hands can predict adult synchrony. Although past work shows hands attract attention as a meaningful and socially-relevant location for communication, gestures, joint attention, and manual actions (Bertenthal & Boyer, 2012; Frank et al., 2012; Fausey et al., 2016; de Villiers Rader & Zukow-Goldring, 2010; Yu & Smith, 2013), we calculate a novel measure of attention to hands—*hand synchrony*—that describes the degree to which infants and children follow hands in the scene. We predict that hand synchrony will predict adult synchrony above and beyond age alone. In other words, attention to hands will increase with age, regardless of the stimulus, because hands are a meaningful feature of scenes.

Lastly, we will test whether attention to hands interacting with objects can predict adult synchrony. *Hand-object synchrony* will provide a measure of attention that is sensitive to the information cued when agents engage in manual actions with objects. Past research has identified that infants and children become increasingly capable of extracting information from hand-object actions like joint attention, affordances, and goals (Barrett et al., 2007; Shepherd, 2010; Woodward, 1998; Yu & Smith, 2013). However, it is unexplored whether the ability to extract information means more attention to hand-object actions in dynamic scenes. We predict that changes in hand-object synchrony will account for changes in adult synchrony above and beyond age alone.

## 3.3 Methods

### 3.3.1 Participants

The current study involves secondary analysis of prior research (Kadooka & Franchak, 2020) on the developmental changes in visual attention of children (6 months to 11 years) and adults when watching videos. The original sample consisted of 79 children and 10 college-aged adults. Data from an additional 10 adults were collected as part of the current study to serve as a comparison group. In total, 79 children and 20 adults were used as participants for this study. For detailed information about the distribution of participant ages and exclusion criteria, see the original paper (Kadooka & Franchak, 2020).

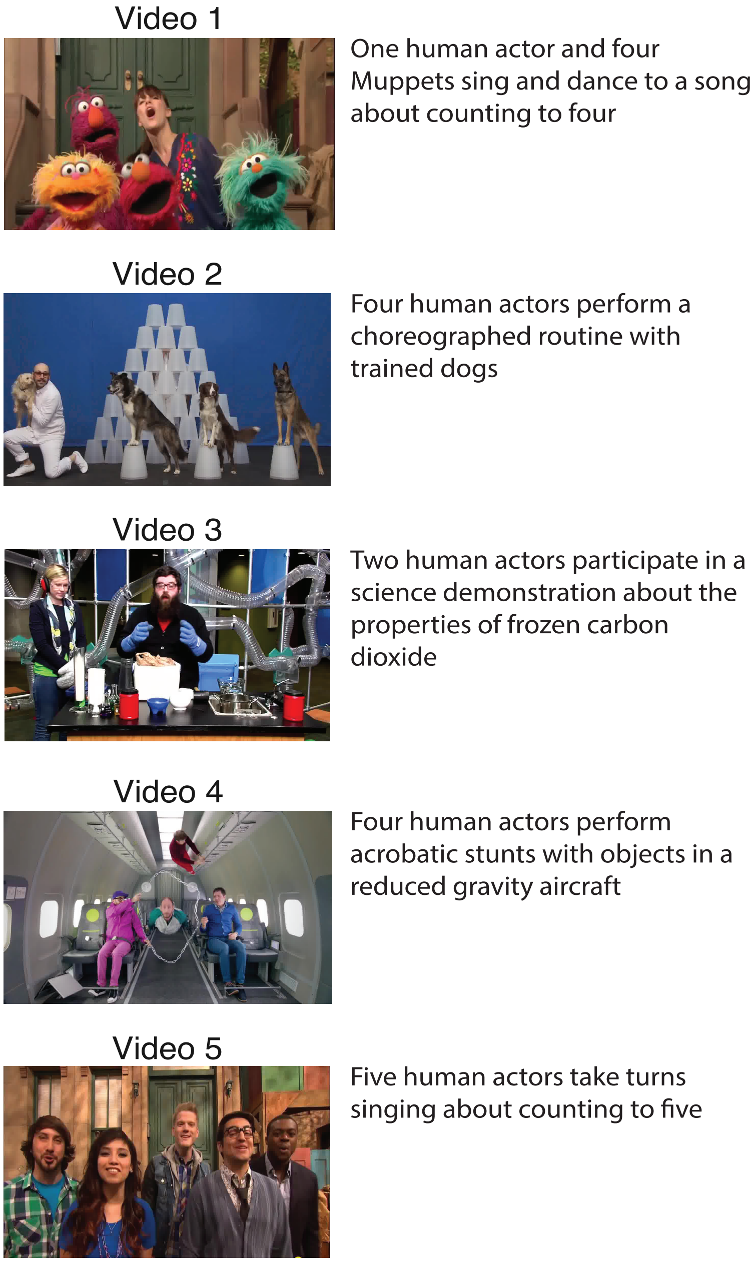
Participating families were recruited from the Riverside County area and received a book or small toy and $10. Adult participants were undergraduates from the University of California, Riverside who received credit towards completing a course requirement. Participating children were identified by their caregiver as Black/African American ($n$ = 1), American Indian/Alaskan Native ($n$ = 4), non-Hispanic White ($n$ = 23), Hispanic or Latino(a)/White ($n$ = 30), and more than one race ($n$ = 21). Adult participants identified as Black/African American ($n$ = 2), Asian ($n$ = 8), non-Hispanic White ($n$ = 1), Hispanic or Latino(a)/White ($n$ = 8), and more than one race ($n$ = 1). All participants had normal or corrected-to-normal vision. Written informed consent was provided by adult participants and parents or legal guardians of child participants. Approval for the study was given by the Institutional Review Board of the University of California, Riverside.

### 3.3.2 Stimuli

Eye tracking data from five of the seven original video stimuli were selected for secondary analysis. Each of the seven videos were chosen to present 2 minutes of diverse, child-friendly scene content with no cuts and limited graphical animations. One video was excluded because it contained too few instances of hands in the scene and the other was excluded because it contained over 30 hands from 18 agents which were too widely distributed on screen to differentiate attention to hands vs non-hand areas. The five remaining videos included two Sesame Street videos, two music videos, and a children's science demonstration. Figure 3.1 provides an exemplar screen shot and brief description of the video stimuli. The original audio tracks were replaced with children's instrumental music to isolate the role of visual information. Videos of the stimuli with overlaid gaze from participants is available on Databrary (https://nyu.databrary.org/volume/1007). Participants in the final sample provided data for a minimum of one video but could provide data for up to five videos. The average number of videos of sufficient data per participant was $M$ = 4.65 ($SD$ = 0.82).

### 3.3.3 Apparatus

Participant eye movements (right eye only) were recorded using an Eyelink 1000 Plus remote eye tracker (SR Research Ltd.) at a temporal resolution of 500 Hz. Videos were presented at 30 Hz. on a 43.2 cm (diagonal) monitor which subtended a visual angle of 31° x 19°. The monitor and eye tracker were mounted on an adjustable arm to facilitate different viewing heights and positions.



Video 1

One human actor and four Muppets sing and dance to a song about counting to four

Video 2

Four human actors perform a choreographed routine with trained dogs

Video 3

Two human actors participate in a science demonstration about the properties of frozen carbon dioxide

Video 4

Four human actors perform acrobatic stunts with objects in a reduced gravity aircraft

Video 5

Five human actors take turns singing about counting to five

Figure 3.1: Exemplar screenshots and brief description of the five stimuli videos

### 3.3.4 Procedure

Participants sat in a room separated from the experimenter. Infants sat in a highchair and were secured with a harness to reduce body movement. Caregivers sat behind the highchair and were instructed to not interact with the infants. Adults and children sat in a chair facing the monitor. A target sticker was placed on the forehead to facilitate eye position tracking and distance estimation by the eye tracker. Angle and distance of the eye tracker was adjusted to be  60 cm from the participant's eyes. All participants completed a 5-point calibration and subsequent validation procedure to ensure that calibration accuracy was 1.5° of error or less. If necessary, the calibration and validation was repeated until sufficient calibration was met. The video stimuli were then presented in a randomized order. Prior to each video starting, participants were required to fixate a gaze-contingent trigger target located at the center of the screen. This ensured that participant gaze always started in the center of the stimulus.

The quality of infant eye tracking data is often lower than adults' (Wass et al., 2013) which can impact accuracy (error between the true point of gaze and reported gaze) and precision (variation in reported gaze over successive samples). In our sample, average spatial errors across all ages were less than 1° ($M$ = 0.53°, $SD$ = 0.25) but were negatively correlated with age ($r$ = -.27, $p$ = .006). Yet, this age disparity only resulted in a difference of 0.22° between adults and the youngest infants which suggests minimal impacts to data quality. Following the method from Wass, Forssman, and Leppänen (2014), average precision across age was calculated ($M$ = 1.65°, $SD$ = 0.33) and was also negatively correlated with age ($r$ = -0.20, $p$ = .048). Similar to accuracy, the difference of 0.18° was minimal between adults and infants. However, concerns about fixation-detection algorithms for lower quality data in infants (Wass et al., 2014) led us to avoid fixation detection out of caution.

### 3.3.5 Data processing

For each participant, raw eye tracking data was extracted as a time series of horizontal and vertical gaze coordinates for the duration of each stimulus. Periods spent with the eye occluded, eye closed, or looking off screen were excluded from analyses.

#### Adult synchrony

Attentional synchrony is the degree to which there is spatiotemporal consistency in gaze behaviors. Synchrony with adults' gaze provides a baseline for assessing predictive factors. To calculate attentional synchrony with adults, participant eye movements were compared to a group of ten comparison adults. By using a comparison group, this ensures that all participants were compared to the same independent group of adults. Following the metric used by Franchak and colleagues (2016), inter-subject correlations (ISCs) were calculated for each participant paired with every adult in the comparison group to determine adult synchrony for each video. For each pair of observers, ISCs were calculated by: 1) calculating the correlation coefficient for the vertical time series, only for samples at which both observers have valid data, 2) calculating the correlation coefficient for horizontal time series, only for samples during which both observers have valid data, and 3) averaging the horizontal and vertical correlation coefficients. An individual participant's adult synchrony is the average ISC between the participant and each of the ten comparison adult observers. As a correlation, ISCs are bounded between -1 and +1, therefore adult synchrony scores closer to 1 indicates greater spatiotemporal similarity with adults.

### Hand synchrony

To capture hand synchrony, we used dynamic area of interest (AOI) analyses. For each video, elliptical AOIs were drawn around the hands of each agent using Dataviewer Software (SR Research Ltd.). AOIs changed in size and location to accommodate character and camera movement across all frames in which the hands were in view. Across the five videos, 44 hands belonging to 22 agents were coded in this way. On average, hands were visible 84% of the time. However, this varied from video to video, with visible hands ranging from 62% to 100% of the video. Table 3.1 contains a breakdown of time that hands were visible on screen for each stimulus. It is important to note that while hands were prevalent, the average percent of time that an individual hand was visible in a stimulus was 35% (min = 1%, max = 98%) and was often discontinuous. In Figure 3.2, the blue color shows when hands of two agents were visible for the 2 minute duration of a Video 3. While hands were visible for the entire duration, this does not mean that every hand was visible at all times.

Table 3.1: Percentage of time hands and hand-object actions were visible for each video

| | Hands visible | Hand-object actions visible |
|---|---|---|
| Video 1 | 68.94% | 37.11% |
| Video 2 | 92.92% | 61.67% |
| Video 3 | 100% | 72.81% |
| Video 4 | 98.07% | 42.88% |
| Video 5 | 61.65% | No hand-object actions |

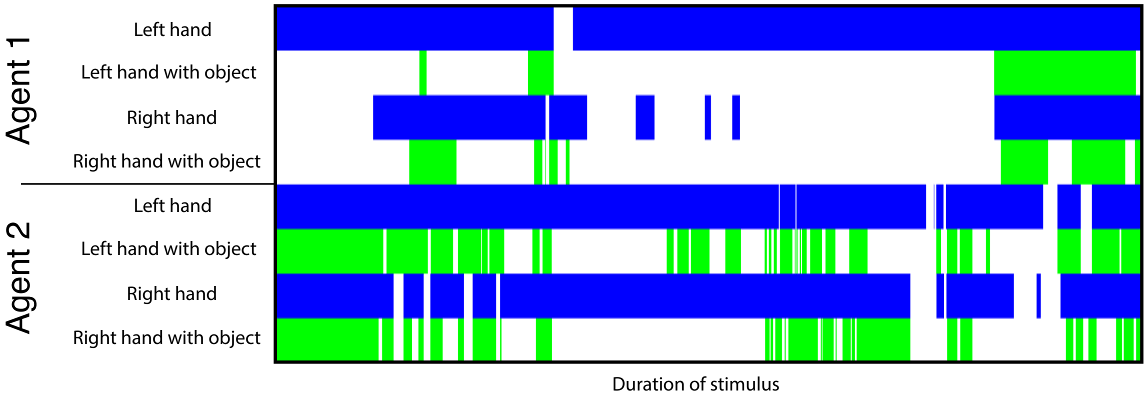


Figure 3.2: Exemplar visualization of times when hands and hand-object actions are visible for two agents in Video 3. Blue horizontal bars indicate when hands are visible for each hand. Green horizontal bars indicate when hand-object actions are visible for each hand.

For each hand AOI, a time series of Cartesian coordinates (X,Y) defined the ellipse of the AOI in units of pixels. To calculate hand synchrony, each participant's gaze data was then correlated with the center of the nearest hand AOI. This was determined by the shortest Euclidean distance in 2-dimensional space between gaze and the centers of any present AOIs. Thus, this measure could account for simultaneous hand AOIs and capture dynamic changes in attention to different hands from moment to moment. Similar to calculating synchrony with adults, correlation coefficients in the horizontal and vertical dimension were calculated for all times in which hand AOIs existed and participants had valid gaze on screen. Correlation coefficients in both dimensions were averaged to produce a single measure of hand synchrony for each participant in each video. Higher correlation values indicate greater hand synchrony.

### Hand-object synchrony

Hand synchrony when interacting with objects provides a spatiotemporal cue about the manual actions of agents. We coded the times when each hand was interacting with an object. Interacting with an object involved being physically in contact with an item that could be carried or moved. Hands interacting with the agent's own body/clothing, other agents, immovable furniture, or surfaces like walls and floors were not counted as manual actions with objects. One out of the five videos did not contain manual actions, therefore analyses involving hand-object synchrony only included four videos. Across the four videos, manual actions occur on average 54% of the time but range from 37% to 73% depending on the video (see Figure 3.1). Of the periods that hands were in view, 59% involve manual actions. Figure 3.2 provides a visual representation of periods of time when hand-object actions were occurring in green.

The process of calculating synchrony with hand AOIs was repeated but only for portions of the videos in which a participant provided valid gaze and hands were interacting with objects. As before, if there were simultaneous hand AOIs, gaze was correlated to the center of the nearest hand AOI. Correlation coefficients in the horizontal and vertical direction were averaged for each video to produce scores of attentional synchrony to hands when interacting with objects. Hand-object synchrony scores closer to 1 indicate greater attentional consistency towards hand-object actions.

## 3.4 Results

Three sets of analyses were performed in order to assess what predicts attentional synchrony with adults (adult synchrony) across diverse video stimuli. First, to replicate past work (Franchak et al., 2016), we tested whether adult synchrony increased with age. Second, hand synchrony was added as a predictor to determine its explanatory power of adult synchrony above and beyond age alone. Third, hand-object synchrony was tested as a predictor of adult synchrony. Due to only four videos containing hand-object actions, the first and second analyses were repeated with a subset of data including only the four videos so that contribution of hand-object synchrony in variance reduction could be identified relative to age and hand synchrony.

Statistical analyses were conducted in R (R Core Team, 2017). Linear mixed-effects models (LMMs) were constructed using the *'lme4'* package (Bates, Mächler, Bolker, & Walker, 2015) to predict adult synchrony. The current data is structured such that participants viewed multiple videos yet they may have only contributed data to a subset of the videos. LMMs are well suited for modelling this type of repeated measure structure to account for similarities within participants and within videos.

### 3.4.1 Attentional synchrony with adults increases with age

Prior work has demonstrated logarithmic changes in attention development (Amso et al., 2014; Kadooka & Franchak, 2020) which rapidly progress during infancy but slow as individuals enter late childhood. Therefore, log-transformed age was used to predict adult synchrony. Preliminary analyses confirm non-transformed age led to poorer fit compared to log-transformed age. A LMM with log-transformed age and random intercepts for participant and video confirmed prior findings (Franchak et al., 2016; Frank et al., 2009; Kirkorian et al., 2012) that age significantly predicts adult synchrony ($beta = 0.03$, $p < .001$). Table 3.2 shows fixed and random effects for this model under Model Age - 5 Video. Eye movements become more adult-like with age following a logarithmic curve for all videos as seen in Figure 3.3.

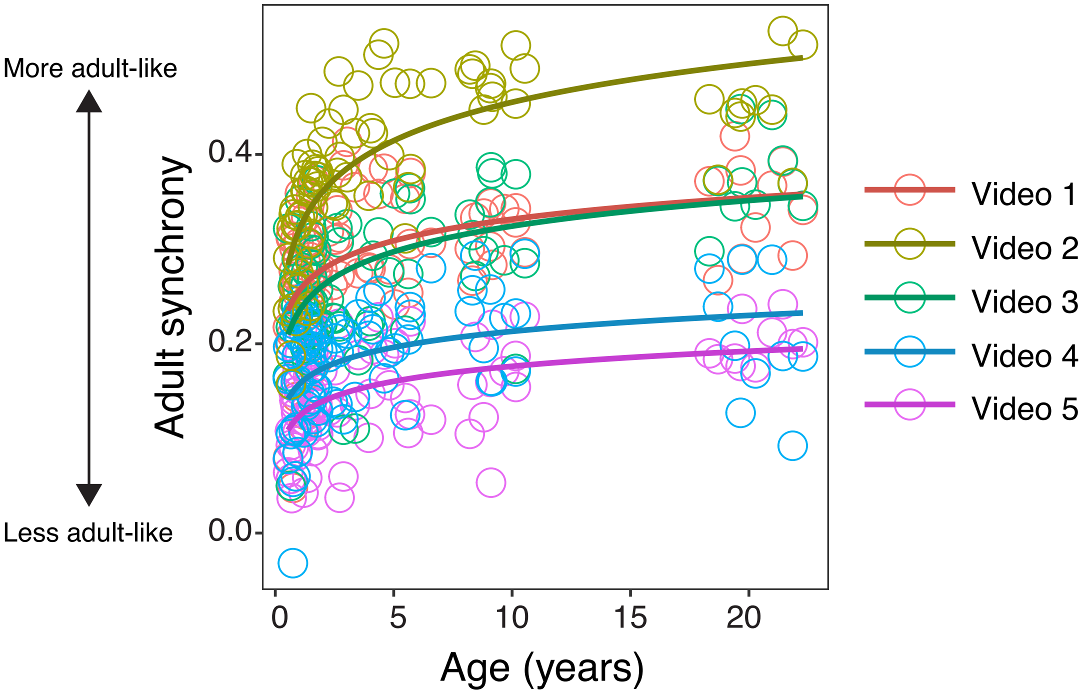


Figure 3.3: Changes in adult synchrony as a function of age. For comprehensibility, logarithmic functions are plotted for each stimulus video

### 3.4.2 Hand synchrony predicts adult synchrony

Overall, attention was fairly synchronous with hands ($M = .49$) but varied widely ($SD = .21$, min = -.16, max = .80). Hand synchrony was positively correlated with adult synchrony ($r = .40$, $p < .001$). However, this relationship was stronger for some video than others (Figure 3.4). To determine whether attention to hands could account for the age-related increases in adult synchrony, hand synchrony was added as a fixed effect to the LMM constructed in the prior section. The fixed effect of log-transformed age and random intercepts for participant and video remained in the model. A model with random slopes for video failed to converge and was not included. Hand synchrony significantly predicted adult synchrony ($beta = 0.17$, $p < .001$) indicating a positive correlation with adult synchrony. Log-transformed age remained significant ($beta = 0.03$, $p < .001$). Table 3.2 shows fixed and random effects estimates for this model under Model Hands - 5 Video. Importantly, hand synchrony significantly improved model fit beyond age alone $X^2(1) = 22.69$, $p < .001$).

Table 3.2: Comparison of linear mixed-effect model predicting attentional synchrony with adults (adult synchrony) from log-transformed age and hand synchrony. Random effects of subject and video estimate standard deviation (SD) of parameters.

| | Model Age - 5 Video | | | Model Hands - 5 Video | | |
|---|---|---|---|---|---|---|
| | Fixed Effects | | | | | |
| Predictor | *B* | *SE* | *p* | *B* | *SE* | *p* |
| Intercept | 0.009 | 0.043 | 0.83 | -0.056 | 0.041 | 0.196 |
| log(Age) | 0.035 | 0.003 | < 0.001 | 0.032 | 0.003 | < 0.001 |
| Hand Synchrony | - | - | - | 0.176 | 0.036 | < 0.001 |
| | Random Effects (*SD*) | | | | | |
| Subject | 0.016 | - | - | 0.019 | - | - |
| Video | 0.073 | - | - | 0.083 | - | - |
| AIC | -1165.5 | - | - | -1186.2 | - | - |

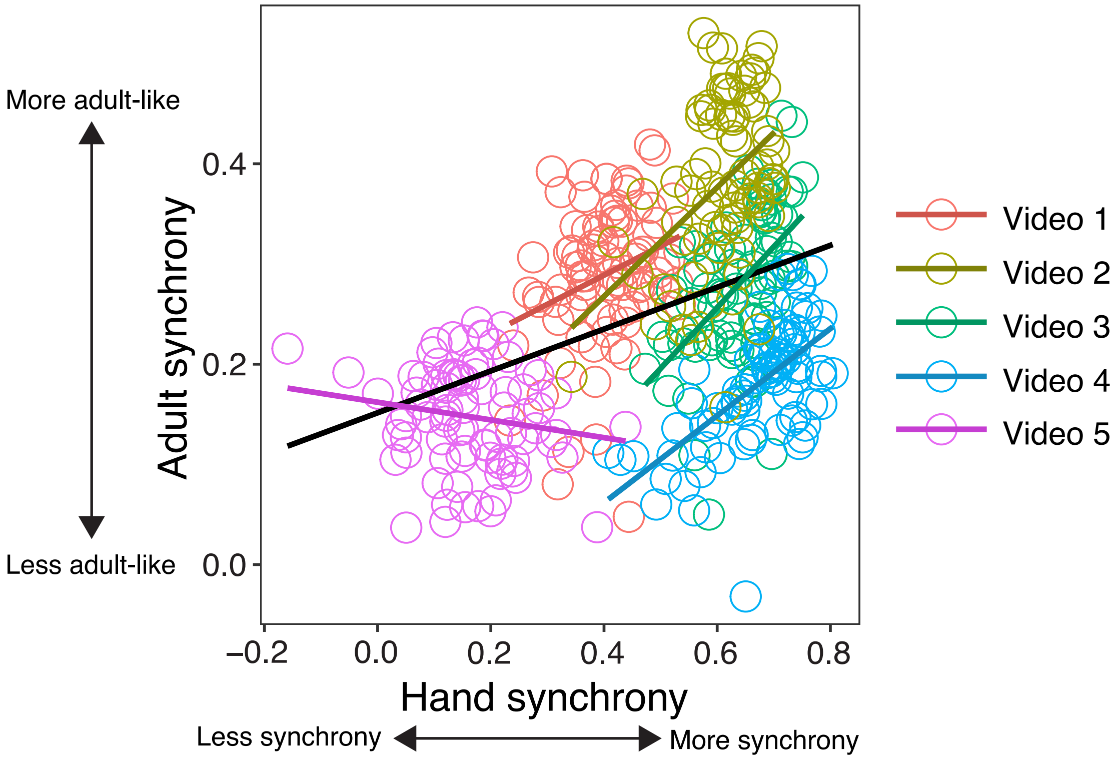


Figure 3.4: Relationship between hand synchrony and adult synchrony. Each circle represents a single participants observation for one video stimulus. The overall effect of hand synchrony on adult synchrony is plotted in black. For additional comprehensibility, correlations are plotted for each stimulus video

### 3.4.3 Attentional hand-object synchrony predicts adult synchrony

Only 4 videos contained hand-object actions. Among those videos, hand-object synchrony averaged $M = .41$, but ranged from a minimum of -.25 to a maximum of .72 ($SD = .18$). Similar to hand synchrony, hand-object synchrony was correlated with adult synchrony ($r = .61$, $p < .001$). Although a positive correlation was consistent across all videos, Figure 3.5 shows variation in this relationship by video. To determine the contribution of hand-object synchrony as a fixed predictor relative to log-transformed age and attention to hands, three nested models were compared. In all models, age-related increases in adult synchrony were predicted by the fixed effect of log-transformed age and the random intercepts for participant and video. No additional parameters were added to Model Age. Model Hands added hand synchrony as a fixed effect to Model Age. Model Hands-Object added hand-object synchrony as a fixed effect to Model Age. Table 3.3 provides model estimates and model fit (AIC). The effect of hand-object synchrony was significant ($beta = 0.153$, $p < .001$) and indicated a similar positive correlation with adult synchrony. Both hand synchrony and hand-object synchrony improved model fit when added to Model Age ($X^2(1) = 43.46$, $p < .001$, $X^2(1) = 20.53$, $p < .001$, respectively). While model fit was higher for Model Hands over Model Hands-Object, neither were able to explain more variance than the other.

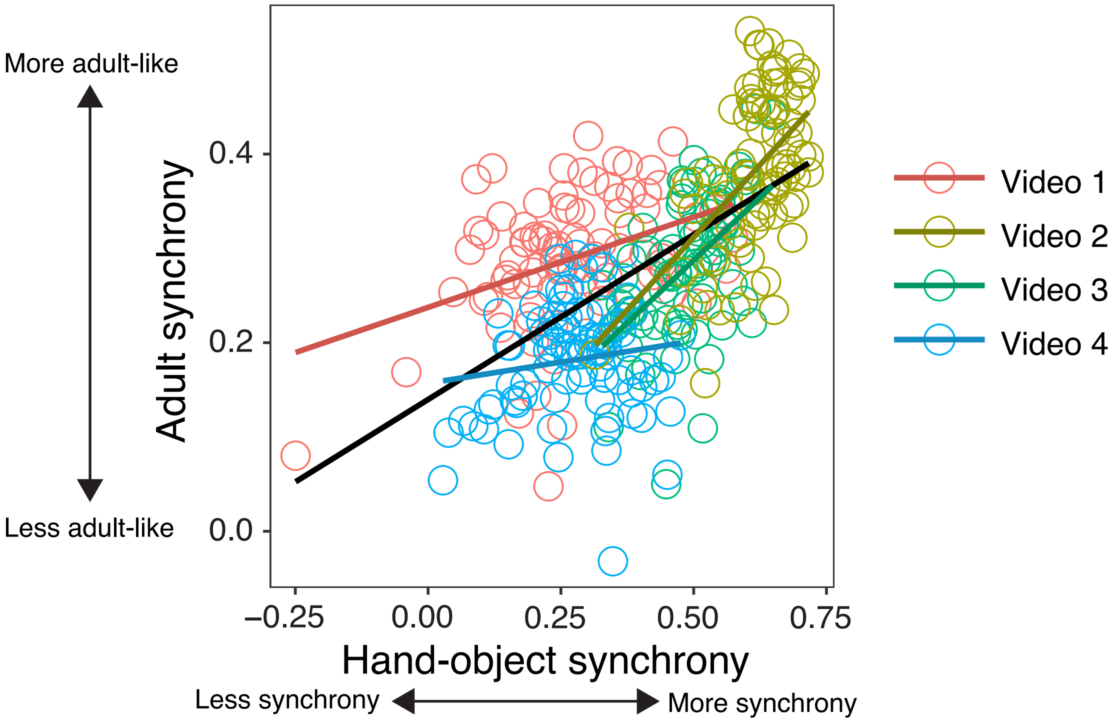


Figure 3.5: Relationship between hand-object synchrony and adult synchrony. Each circle represents a single participants observation of one video stimulus. The overall correlation is plotted in black.

Table 3.3: Linear mixed-effect model comparison predicting attentional synchrony with adults (adult synchrony) from three models. Log-transformed age, hand synchrony, and hand-object synchrony as fixed effects. Random effects of subject and video estimate standard deviation (SD) of parameters.

| | Model Age | | | Model Hands | | | Model Hands-Object | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Fixed Effects | | | | | |
| Predictor | *B* | *SE* | *p* | *B* | *SE* | *p* | *B* | *SE* | *p* |
| Intercept | 0.015 | 0.042 | 0.729 | -0.121 | 0.051 | *0.0455* | -0.015 | 0.036 | *0.656* |
| log(Age) | 0.038 | 0.003 | < 0.001 | 0.032 | 0.003 | < 0.001 | 0.034 | 0.003 | < 0.001 |
| Hand Synch | - | - | - | 0.306 | 0.044 | < 0.001 | - | - | - |
| Hands-Obj Synch | - | - | - | - | - | - | 0.153 | 0.033 | < 0.001 |
| | | | | Random Effects (*SD* | | | | | ) |
| Subject | 0.016 | - | - | 0.008 | - | - | 0.014 | - | - |
| Video | 0.070 | - | - | 0.085 | - | - | 0.054 | - | - |
| AIC | -902.30 | - | - | -943.75 | - | - | -920.84 | - | - |

## 3.5 Discussion

In summary, we measured age-related changes in attentional synchrony with adults and investigated whether attentional synchrony with two features, hands and hand-object actions, can account for age-related differences in synchrony with adults. Our results confirm past findings (Franchak et al., 2016) of age-related changes towards increasingly synchronous eye movements that were robust across five varied and dynamic stimuli. Furthermore, both hand synchrony and hand-object synchrony predicted adult synchrony above and beyond age alone. The degree to which infants and children look to hands and hand-object actions reliably accounts for similarities in how those observers prioritize meaningful information. Here we will discuss the implications of these findings and a possible mechanism for the changes observed.

Our findings support the perspective that development of visual attention involves learning to better prioritize meaningful features in ways that are increasingly more adult-like. Hands and hand-object actions were not always present in the scenes, yet there is a strong relationship between attention to these features and adult synchrony. This suggests that looking to hands and hand-object actions may be one of many attentional behaviors that develop together towards more adult-like prioritization of meaningful features. In other words, it is likely that the degree to which an observer synchronizes attention to hand-object actions is also related to more general attentional biases that occurs when hand-object actions are not occurring.

Becoming more adult-like in looking to hands may be helpful in developing key experiences and abilities that are developmentally appropriate. Extracting information from manual actions and grip positions during infancy coincides with infants being motorically ready to perform grasping actions of their own (Libertus & Needham, 2010; Needham, Barrett, & Peterman, 2002). Similarly, developing joint attention to objects by looking to hands occurs around when infants are learning names for objects (Yu & Smith, 2011). The development of looking to hands and hand-object actions may be developmentally gated to aid in the timing of these experiences. In other words, attention, comprehension, and the motor system may work together to structure the visual environment (L. B. Smith, Jayaraman, Clerkin, & Yu, 2018; Yoshida & Fausey, 2019) by increasing attention when the infant is ready to comprehend information, in this case manual actions. Too soon, and the infant is provided visual experiences that add unnecessary complexity, too late, and the infant may miss out on critical information.

We identify two plausible and non-mutually exclusive mechanisms that would both be supported by our findings. The first mechanism involves maturation of general attention skills. From this account, development in prioritizing hands, hand-object actions, and adult-like prioritization of meaningful regions is the result of improvements in general attentional control. Attentional control refers to the ability to orient and select visual information while also inhibiting attention to other features (Colombo, 2001). Infant attention to static images are impacted by their attentional selection and orientation abilites (van Renswoude, Visser, Raijmakers, Tsang, & Johnson, 2019) and attentional control may be even more relevant in dynamic stimuli when orienting and selecting information is time sensitive. Looking to hands and hand-object actions could be the result of actively selecting these features that convey important information in the moment. Likewise, developing attentional control would involve suppressing attention to other features in favor of hands. Improvements in attentional control are observed throughout infancy and childhood into at least the early teenage years (Oakes & Amso, 2018; Ruff et al., 1998; Paus, Babenko, & Radil, 1990; Aring, Grönlund, Hellström, & Ygge, 2007). Further, this general ability develops rapidly during early infancy which could account for the logarithmic pattern in developing adult synchrony.

The other plausible mechanism involves improvements in comprehending a scene's meaning. Through social input, exploration, and experience, infants may learn to better comprehend the content of a scene. As previously described, interactions with adults lead to experiences in which attention is guided towards meaningful locations like hands and hand-object actions (Yu & Smith, 2013). With repetitive and redundant encounters with cues that signal meaningful information, infants and children may modulate their prioritization as they increasingly comprehend the importance of social partners, hands, faces, eyes, affordances, and information that is relevant to their past experiences. This account is also plausible in describing the results of this study.

Most likely, both mechanisms are at play in the development of visual attention. Comprehension and attentional skills interact and recursively improve the ability to prioritize meaningful information. Both these mechanisms lead to broad cascading changes across development, which makes it difficult to establish temporal precedence. Individual differences in attentional control are associated with several achievements including word learning (Yu & Smith, 2011), walking ability (Mulder, Oudgenoeg-Paz, Verhagen, van der Ham, & Van der Stigchel, 2022), and executive function (Veer, Luyten, Mulder, van Tuijl, & Sleegers, 2017). It is entirely possible that any of these achievements could be a driving force behind developing adult-synchrony by offering greater comprehension of the visual world, rather than attentional control alone. Similarly, improvements in comprehension comes from gaining experience across a wide variety of knowledge; narrative structure (Kirkorian et al., 2012; Kirkorian & Anderson, 2018), goal-directed actions (Flanagan & Johansson, 2003), film techniques (Kirkorian & Anderson, 2017), language (Anderson et al., 1981), and regularities in scenes (Helo et al., 2017) all influence attention. But better attentional control may lead to voluntary detection of this information.

### 3.5.1 Limitations and future direction

Our findings indicate a strong influence of attention to hands that changes with age. While hands were selected as a semantically relevant feature, it is possible that hands were visually salient. Other features that have been selected as ‘top-down’, like faces, tend to contain regions of high salience (Henderson et al., 2007; Torralba et al., 2006; Wass

& Smith, 2015). However, this is unlikely in our case given that, in preliminary analyses, salience was unable to consistently account for changes in adult synchrony. Further, no age-related effect was detected in the the original data analysis (Kadooka & Franchak, 2020). Incorporating low-level information, in addition to other features, would help to deepen understanding of visual attention and explain changes in adult synchrony. Additionally, knowing how attention changes for particular features means that future studies could manipulate infants' prioritization of meaningful information with targeted manipulations of comprehension.

In this study we used a measurement, adult synchrony, that indexes the development of visual attention irrespective of features. Although this metric is spatiotemporally sensitive to changes in meaning, there are some limitations. Calculations of ISCs are likely confined to screen-based stimuli because it requires that observers view the same stimulus. Unfortunately this would limit the ability to test other potentially important changes in visual attention beyond the screen like motor ability or navigation. However, newer technology like virtual reality and head mounted displays (Farmer et al., 2021) with built-in eye tracking function could present identical stimuli. Head-mounted eye tracking would require constructing reproducible visual scenes in real life which is plausible but difficult.

# Chapter 4

# Attentional synchrony when viewing manual actions

## 4.1 Abstract

Developing adult-like visual attention involves changes in synchronizing gaze to meaningful information at the right time. How do children develop the ability to prioritize meaningful information from moment-to-moment? We propose that perceptual experiences allow for increased comprehension of the information that conveys the most meaning. For manual actions, the meaningful information occurs when hands interact with objects. In the current study, we tested whether prior viewing of a novel manual action could change attention of 4-years-olds towards more adult-like prioritization of hands-object actions. Participants either viewed a live demonstration of a manual action or talked about the objects involved with the same action. Eye movements were recorded during a subsequent screen-based viewing of the action. Results show that prior opportunity to view the action did not change attention towards greater synchrony with adults or greater attention to hands-object actions. Comprehension and attention was similar between groups and was largely unaffected by a prior visual experience. However, synchrony with adults was significantly correlated with attention to hand-object actions. More effective manipulations of comprehension are discussed.

## 4.2 Attentional synchrony when viewing manual actions

The development of visual attention in infants and children involves a progression towards more synchronous adult-like eye movements when watching dynamic stimuli (Franchak et al., 2016; Kadooka & Franchak, 2020, in prep). In other words, where children choose to allocate their gaze becomes increasingly similar to adults as they get older. Our past work suggests that this attentional synchrony is the result of improvements in prioritizing the meaningful information that is conveyed in a scene from moment-to-moment (Kadooka & Franchak, 2020). For instance, looking towards an actor's hands interacting with objects can account for changes in attentional synchrony with adults beyond age alone (Kadooka & Franchak, in prep).

We propose that one possible route to developing more adult-like visual attention involves gaining experiences that improve scene comprehension leading to changes in how observers prioritize meaningful information. In the context of manual actions with objects, greater understanding of an action has been shown to change visual attention to hand-object interactions (Filippi & Woodward, 2016; Hard, Meyer, & Baldwin, 2019; Rotman, Troje, Johansson, & Flanagan, 2006). Experiences that allow for better understanding of an action result in a feedback loop in which greater comprehension supports subsequent attention to meaningful information, which in turn provides a deeper understanding of the action. However, it has not been tested whether this reciprocal relationship specifically leads children's eye movements to better synchronize with adults' eye movements. That is, can providing an opportunity to extract meaning from an action lead to increased adult synchrony? If so, are there changes in looking towards the locations where hands and objects interact?

## 4.3 Developmental increases in adult synchrony

Adult synchrony is the degree to which an observer's eye movements correlate with the gaze behavior of adults (Franchak et al., 2016). This comparison is measured as an inter-subject correlation (ISC) which is the correlation of eye gaze location coordinates over time when viewing the same stimulus. The utility of measuring adult synchrony comes from the reliable overlap in where adults look, both spatially and temporally. A high level of synchrony persists across a wide range of dynamic stimuli including animated cartoons, naturalistic live-action videos, and Hollywood-produced films (Dorr et al., 2010; Franchak et al., 2016; Gannon & Grubb, 2022; Hart et al., 2009; Mital et al., 2011; Shepherd et al., 2010; Rider et al., 2018; Wang et al., 2012). This is to say, adult synchrony is a metric of attention that indexes how adult-like an observer's eye movements are. Importantly, adult synchrony is agnostic to expectations of where researchers think attention should be allocated. Of course, this does not mean that adults look to exactly the same place at every moment, but rather, there is considerable convergence between different adults' attention.

How does age relate to adult synchrony? Eye movements of infants and children become increasingly adult-like with age (Franchak et al., 2016; Kadooka & Franchak, 2020). A large body of developmental literature has identified age-related changes in visual attention to scene features including saliency, faces, bodies and hands (see Kadooka and Franchak (2020) for a more thorough review). However, in our prior work (Kadooka &

Franchak, 2020), we provide evidence that the development of adult synchrony is not based on global changes in attention to features but rather driven by improvements in prioritizing meaningful information from moment to moment. At one moment, adults may prioritize looking to a face of a central character, but within a few seconds adults' attention may shift towards a gesture of an entirely different person. Measurements of global attention that are based on single features, like faces, may fail to capture the spatiotemporal changes in relevance to many different features. On the other hand, adult synchrony is an index of similarity between a given observer and adults' actual spatiotemporal allocation of attention. Attentional synchrony is achieved when there is agreement with adults on how to prioritize information within dynamic scenes.

The dynamic nature of the visual world means that different features may be prioritized depending on what is meaningful in that moment. As a scene unfolds, the importance of a person's face may wane after that person points towards a toy and grasps it. This emphasis on understanding the influence of 'meaning' on attention is supported by Henderson et al. (2017) in which 'meaning maps' were computed to measure the semantic relevance of static images. Meaning maps were created by asking crowd-sourced raters to judge the meaningfulness of many overlapping image regions which were then used to create a 'map' that evaluates the spatial distribution of meaningfulness. In several variants of this methodology, they have shown that regions rated as more meaningful receive greater attention (Henderson, 2017; Henderson & Hayes, 2018; Rehrig et al., 2020).

## 4.4 Comprehension drives prioritization of meaning

How does comprehension relate to attention? There is likely a reciprocal relationship between what an observer understands and how they allocate their attention to meaningful information. When comprehension is high, attention should be allocated to semantically-relevant information that is important for understanding the scene. Because the relationship between comprehension and attention should progress together, one approach to disentangling these constructs involves violating semantic expectations. Helo and colleagues (2017) compared adults and 2-year-olds as they viewed scenes with irregularities (e.g. a bar of soap on a kitchen table) and scenes without irregularities. Adults, but not 2-year-olds, were sensitive to regularities in the environment and looked longer to the irregular objects. Similarly, when adults view static scenes that contain 'impossible' floating objects (e.g. a saucepan hovering in a kitchen), greater attention is allocated to these impossible objects (Võ & Henderson, 2009). Observers use their comprehension to determine meaningful information that should receive more attention. However, selection of meaningful information also implies inhibiting attention to information that is not meaningful. When preschool-aged children were presented with TV show content that was edited to be incomprehensible based linguistic or narrative manipulations, look duration was shorter and gross attention decreased compared to comprehensible scenes (Anderson et al., 1981). Information that is less comprehensible and meaningful should not receive as much attention. However, modulating attention based on meaning requires the observer to be sensitive to the difference. Infants younger than one year old look to comprehensible and incomprehensible scenes for the same amount of time (Pempek et al., 2010). Therefore, developmental improvements in comprehension helps children to determine and attend to meaningful information. Likewise, attention is allocated to the information that is most meaningful for comprehending the scene.

The relatedness of comprehension and attention through the selection of meaningful information provides context for age-related differences in attention to hand-object actions previously observed (Kadooka & Franchak, in prep). When infants, children, and adults viewed live-action videos, greater spatiotemporal agreement between their gaze and the locations of hand-object actions accounted for attentional synchrony with adults. Attentional synchrony with adults occurs as part of a feedback loop in which comprehension refines our attention towards meaningful information and this refinement towards attending to meaningful information allows us to notice or grasp new comprehension of the visual world. In the following section, we will describe this feedback loop within the context of manual actions.

## 4.5 Attending to and comprehending manual actions

To understand the relationship between attention and comprehension in manual actions, we need to know the meaningful information that exists when hands and objects are interacting. The action literature provides several featural and structural avenues that convey meaning. For instance, goal construal is signaled by the path of the hands (Sommerville et al., 2005), intentionality is signaled by hand trajectories (Cannon & Woodward, 2012), and object affordance is signaled by hand shapes (Barrett et al., 2007; Ambrosini et al., 2013).

A more global indicator of proficient attention and comprehension is the ability to make predictive eye movements to the goals of actions. For instance, when observing simple actions, like moving an object from one location to another, adults' eye movements move along the path of the hands ending in a predictive look to where the object will be placed ~150 ms before the hand arrives at that goal location (Flanagan & Johansson, 2003; Rotman et al., 2006). Differences in predictive looks between knowledgeable and naive observers provide insight on how changes in comprehension can change attention. Comparisons of adults' eye movements between predictable and unpredictable actions show that looks to the goal are delayed when the observer is uncertain of the goal (Rotman et al., 2006). However, adults quickly recover by using other features like trajectory of the hand to reestablish a prediction. When faced with unpredictability, adults' recruitment of other meaningful information likely reflects an accumulation of experience.

Prediction of action goals helps identify the developmental changes in action comprehension and attention. In a study by Kochukhova and Gredebäck (2010), past familiarity with actions led to differences in predictive looks for 6-month-olds, 10-month-olds, and adults. When observing an actor use a spoon, an action that all age groups were familiar with, predictive looks to the mouth were observed for all ages. However, when the actor used a comb, only adults made predictive looks to the head, reflecting adults' familiarity with this action. Infants made reactive looks to the head in which their gaze followed the hand. In another study involving eating-utensil actions, Swedish and Chinese 8-month-olds made predictive looks to the mouth only for utensils that were familiar to them (Green, Li, Lockman, & Gredebäck, 2016). Chinese infants predicted the action only when chopsticks were used, but Swedish infants made predictive looks to the mouth only when a spoon was used. Importantly, infants in this study were at an age in which they had been fed with these utensils but were not yet able to use them on their own. While motor abilities may not be required to develop comprehension of actions, the importance of motor abilities is well established in the literature for adults and infants. Expert adult athletes, compared to novices, are better able to predict trajectories of thrown projectiles based on human movements (Moore & Müller, 2014; Williams, Ward, Knowles, & Smeeton, 2002). The ability to detect statistical regularities in the goals of actions were correlated with 8- to 11-month-olds ability to perform those actions (Monroy et al., 2017), and prediction of goals to grasping actions correlates with 4- to 10-month-olds ability to perform grasping actions (Kanakogi & Itakura, 2011).

## 4.6 Manipulating comprehension and attention in manual actions through action experience

The feedback loop involved in comprehension and attention is directly related to the experiences of the observer. Repeated action experiences (either viewing or doing an action) provide opportunities to attend to meaningful information and build on prior comprehension. In previously discussed work, differences in age or culture stood in for these experiences (Green et al., 2016; Kanakogi & Itakura, 2011; Kochukhova & Gredebäck, 2010; Monroy et al., 2017). However, direct manipulations of exposure to experiences are effective at influencing comprehension and attention. These training studies measure the impact of providing opportunities to experience actions. In the "Sticky Mittens" paradigm, 3-month-old pre-grasping infants are trained to 'grasp' objects with the use of Velcro-lined mittens, prior to when infants typically learn to intentionally grasps objects. Studies using this paradigm have shown that training leads to several cascading effects in infants' subsequent attention, comprehension and action, including greater visual and object exploration, better understanding of reaches as goal-directed, and higher sensitivity to changes in the goals of an actor's reach (Libertus & Needham, 2010; Needham et al., 2002; Sommerville et al., 2005). Another study using a similar design has shown improvements in 13-month-olds ability to predict actions based on observed hand shapes and hand kinematics (Filippi & Woodward, 2016). By gaining specific experiences, infants improve comprehension of meaningful aspects of actions and increase sensitivity of meaningful information when attending to actions.

Combining past work, manual actions provide an avenue for understanding how experiences change comprehension and attention. The reviewed literature on manual actions has focused on changes in infants' and children's sensitivity to features of action based on coarse measurements of attention like looking time or dwell time. Therefore it is unknown if action experience changes attention towards greater adult synchrony, specifically. Providing opportunities for observers to learn about an action may improve their sensitivity and prioritization of meaning in ways that are more similar to adults.

## 4.7 Current Study

In the current study, we tested whether action observation increased children's attentional synchrony with adults when viewing that action. The eye movements of 4-year-olds and a group of comparison adults were recorded while viewing a novel action after they had observed either a live demonstration of the action or an irrelevant experience. Because attention and comprehension of action is tied to prior experiences and abilities (Kanakogi & Itakura, 2011; Needham et al., 2002; Rotman et al., 2006), we selected an action that was novel and unfamiliar to 4-year-olds. By providing an opportunity to observe the action, we expected that this may allow 4-year-olds to identify meaningful information involved in the action, particularly the ways in which hands interact with objects. As a result, we predicted that children who had a relevant action experience would have higher adult synchrony than those with an irrelevant experience. Additionally, we predicted synchronous attention to hand-object actions would mediate this effect, since attention to hand-object actions has been closely related to adult synchrony in our prior work (Kadooka & Franchak, in prep).

## 4.8 Method

### 4.8.1 Participants

The final sample consisted of 10 college-aged adults (4 female, $M = 19.69$, $SD = 1.27$ ) and 26 4-year-olds (15 female, $M = 3.94$, $SD = .13$). Due to the COVID-19 pandemic, child data collection was stopped prematurely and fell short of the 40 children originally planned for this study. Six additional 4-year-olds participated in the study, but their data were excluded due to issues with cooperating and following directions. All participants in the final sample had normal or corrected-to-normal vision with no color blindness or history of familial color blindness.

Families were recruited from the Riverside County area. Adults were college undergraduates recruited from the departmental participant pool and received course credit for participation. Participating children were identified by their caregivers as Black/African American ($n$ = 1), American Indian/Alaskan Native ($n$ = 1), Asian ($n$ = 1), non-Hispanic White ($n$ = 6), Hispanic or Latino(a)/White ($n$ = 10), and more than one race ($n$ = 6). One parent did not report race/ethnicity. Families received \$10 and a small gift or book for participating. Adult participants identified as American Indian/Alaskan Native ($n$ = 1), Asian ($n$ = 4), non-Hispanic White ($n$ = 1), Hispanic or Latino(a)/White ($n$ = 2), and more than one race (n = 1). One adult participant did not report race/ethnicity. Participants or their caregivers provided informed consent after hearing the details of the study. Children gave verbal assent. The study procedure conforms to the US Federal Policy for the Protection of Human Subjects and was approved by the Institutional Review Board of the University of California Riverside.

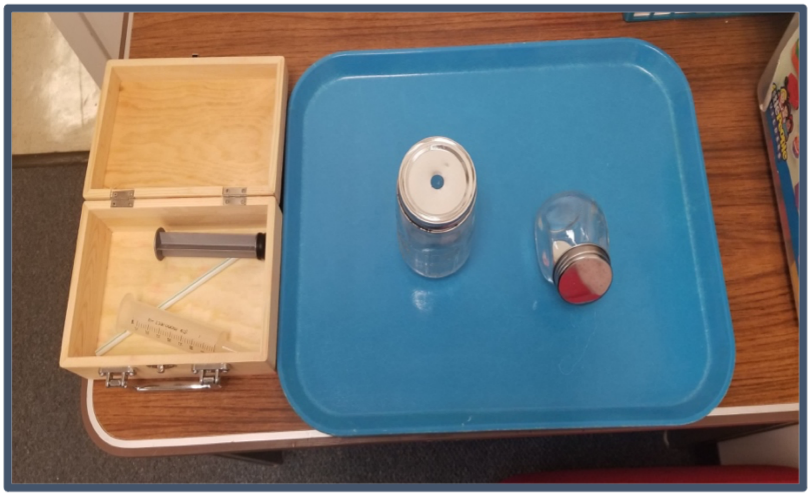
### 4.8.2 Design

All participants engaged in a two-part task: a *live demonstration* followed by a *video observation* task. Participants were randomly assigned to one of two conditions that varied the live demonstration to provide either a relevant or irrelevant perceptual experience. In the *relevant action condition*, participants watched the experimenter perform the target action. In the *irrelevant action condition*, participants talked about the items involved in the target action but did not receive a live demonstration of the target action. In the video observation task, all participants watched the same video recording of the target action while an eye tracker recorded their eye movements.

### Stimuli

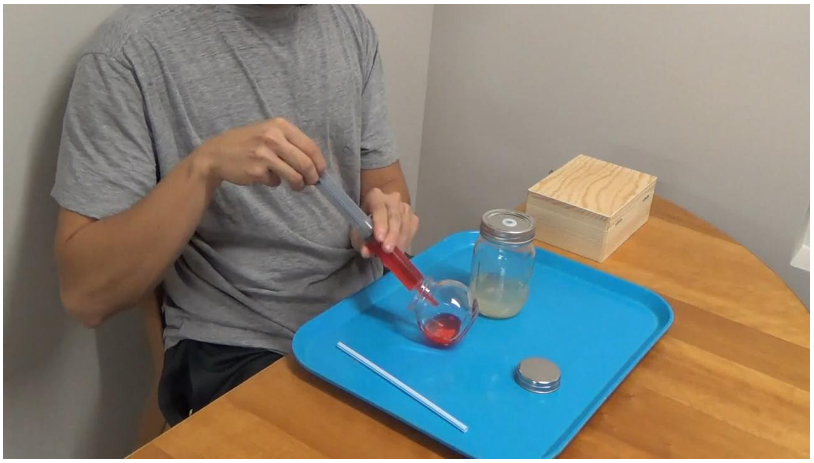
The target action consisted of moving colored juice from one container to another using a plastic syringe. This specific goal-directed action was chosen because pilot testing revealed that it was an unfamiliar action sequence for 4-year-olds but not adults. Materials for the live demonstration involved a plastic 35.56 cm x 45.72 cm tray, a 35 ml plastic syringe (plunger and barrel), a 440 ml wide-mouth mason glass jar with a straw hole lid, a 147.8 ml tilted glass jar with a lid, a disposable plastic straw, a 15.24 cm x 20 cm x 7.93 cm hinged wooden box, and powdered juice mix (lemonade and cherry Kool-Aid). Figure 4.1 presents the materials used during the target action. A video was filmed at 30 fps that depicted an actor performing the target action with same materials used in the live demonstration. The total time of this video was 45 seconds. This video was viewed during the video observation task. Figure 4.2 is a frame from the video depicting the target action.

Figure 4.1: Items Used in Target Action



Note. Materials used during the target action: a plastic tray, a 35 ml plastic syringe (plunger and barrel), a wide-mouth mason glass jar with a straw hole lid , a tilted glass jar with a lid, a disposable plastic straw, and a hinged wooden box.

Figure 4.2: Example frame from the Stimulus Video



Note. A frame from the video presented during the video observation task. This video depicted an actor using a large plastic syringe to move colored juice from one container to another.

### Eye Tracking Apparatus

The video was presented on a 43.2 cm (diagonal) widescreen monitor at 30Hz and subtended a visual angle of 31°x19°. An Eyelink 1000 Plus remote eye tracker (SR Research Ltd.) was mounted below the monitor on an adjustable arm. Eye movements (right eye only) were recorded with a temporal resolution of 500 Hz.

## 4.8.3 Procedure

Prior to the arrival of participants, materials for the live demonstration were prepared. The straw, plunger, and barrel were placed in the wooden box. 35 ml of lemonade was placed in the small, titled jar and 70 ml of Kool-Aid was placed in the large jar. Both jars were placed adjacent to each other, approximately 5 cm apart (see Figure 4.1). Once participants arrived and completed consent procedures, they were seated at a table across from the experimenter. Child participants were seated on a booster seat. A research assistant would enter the room carrying the prepared tray and wooden box. The experimenter would proceed with the *live demonstration* and *video observation*.

### Live Demonstration

For the *relevant action* condition, the tray was placed in front of the experimenter. The experimenter explained their goal was to mix the two flavors of juice using the items in the box in order to taste the combined flavors. Participants were instructed to watch the experimenter as they mixed the juice. Adults were given the following prompt: “I want to know how these two liquids taste when they are mixed together. Please observe the sequence of actions I make to move the liquid from the small jar to the large jar and then take a sip of it.” Children were given a similar but more child-friendly prompt: “I want to know what these two flavors of juice taste like when I mix them together. Watch me as I use the items in the box to move the red juice to the jar with the yellow juice and then take a sip of it.” Experimenters were trained to use a sequence of timed actions that matched the timing of the actions in the stimulus video. Participants in this condition viewed the entirety of the target action sequence.

For the *irrelevant action* condition, the tray was placed in front of the participant. The researcher would ask the participant to talk about each item in front of them including the items in the wooden box. This allowed participants the opportunity to familiarize themselves with the materials but did not allow for observation of the target action sequence. Adults were given the following instructions: “We want to know what the items on the table are. Please identify each item in front of you, as well as the ones in the box, and say the material the item is made out of." Children were given the following instructions: "We want to know what all these things on the table are. Can you point to each of these objects, including the objects inside the box, and tell me what you think they are made out of?".

### Video Observation

Once participants completed the live demonstration, all participants were moved to a separate area containing the eye tracker in order to complete the video observation task. Participants sat in a viewing area that was separated from the experimenter by a hanging curtain. A target sticker was placed on the forehead to facilitate the eye tracker detecting the observers' eyes. Children and adults sat in a chair (with a booster seat for children).

The experimenter adjusted the monitor such that participants were at a viewing distance of  60 cm. Calibration involved a 5-point calibration routine followed by a 5-point validation check. Validation calculated the average error as the disparity between the target location and estimated point of gaze location. This calibration process was repeated until validation indicated 1.5° of average error or less.

After calibration and validation, participants were shown the pre-recorded video of the actor performing the target action of moving the juice using the syringe with no audio. Adults and children were instructed to watch the video as their eye movements were recorded. Importantly, materials and placement in the video were identical to the live demonstration. Timing of actions in the video were comparable to the live demonstration for the relevant action group.

### 4.8.4 Data Processing and Measures

Raw eye gaze locations during the video observation were extracted as a time series of horizontal and vertical gaze coordinates for each observer. Time periods when gaze was off screen, eyes were closed, or eyes were otherwise occluded were excluded from analyses.

#### Adult synchrony

Adult synchrony measures the degree of spatiotemporal similarity between an observer and a comparison group of adults. For this study, adult data was collected to serve as the comparison group. Half of the adults were assigned to each condition and participated in the study to ensure minimal differences between children and adult attention as a result of the study procedure. As a familiar action, adult attention to the target action was synchronized. Using a previously applied process (Franchak et al., 2016; Kadooka & Franchak, in prep), attentional synchrony was calculated by comparing each child participant to every adult in the adult comparison group. For each comparison made, the inter-subject correlation (ISC) was calculated as the correlation coefficient for the time series of eye coordinates. This was calculated separately for the horizontal and vertical direction, then averaged. An individual child's adult synchrony score was the average ISC between that child and each of the adult observers. Adult synchrony scores closer to 1 indicate more adult-like gaze.

#### Hand-object Synchrony

Hand-object synchrony measured the spatiotemporal agreement between an observer's gaze and the location of hands during hand-object actions in our video stimulus. Following the process described in Kadooka and Franchak (in prep), elliptical areas of interest (AOIs) were drawn around hands at all times that hands were visible using Dataviewer software (SR research Ltd.). Next, hand-object actions were coded as the times a hand was physically interacting with an object that could be moved. For both hand AOIs, a time series of coordinates was created that described the center of the ellipse for the entire stimulus duration. To calculate hand-object synchrony, each participant's eye movements were then correlated with the coordinates of the nearest hand AOI center. For times when both hands were visible, the nearest AOI was determined by the shortest Euclidean distance. Similar to ISCs, correlations were calculated in the horizontal and vertical dimensions separately then averaged. Hand-object synchrony scores closer to 1 indicated greater spatiotemporal attention to the locations of hands when interacting with objects.

## 4.9 Results

We predicted that the relationship between condition and adult synchrony would be mediated by hand-object synchrony. Therefore, we performed a mediation analysis. To establish a mediation, the relationship between condition and adult synchrony was examined first. A linear regression using condition to predict adult synchrony indicated no total effect of condition, ($beta = 0.043$, $p = .419$). Synchrony did not significantly differ between children who received relevant action experience ($M = .609$, $SD = .086$) and those who received irrelevant action experience ($M = .566$, $SD = .161$).

Past guidance of mediation models would suggest that a significant total effect is necessary to move forward with the mediation model, however, recent recommendations suggest that continuing without a significant effect can still provide useful information

(Hayes, 2018). Therefore, we proceeded to model the effect of condition on our mediating variable, hand-object synchrony. A linear regression predicting hand-object synchrony from condition found no effect of condition ($beta = 0.05$, $p = .25$). Children in the relevant condition ($M = .65$, $SD = .08$) looked to hand-object actions similarly to those who were in the irrelevant action ($M = .60$, $SD = .14$). Without a significant effect of condition on hand-object synchrony, there is no mediation effect. Contrary to our expectations, children viewed the action in similar ways, regardless of whether they saw a live demonstration of the target action prior.

In past work, adult synchrony was predicted by attention to hands (Kadooka & Franchak, in prep). Therefore, we examined our data to determine if this effect replicated. Since condition did not impact either variable, we combined data from both conditions. There was a strong linear correlation between hand-object synchrony and adult synchrony, ($r = .96$, $p < .001$). This is evident in Figure 4.3 showing that hand-object synchrony is highly predictive of adult synchrony.

## 4.10 Discussion

In this study we attempted to influence adult synchrony and hand-object synchrony by providing 4-year-olds an opportunity to observe a live demonstration of a novel action prior to viewing a video of the action. Our results indicate that regardless of whether 4-year-olds observed the live demonstration, their comprehension of and attention to the novel action was largely unaffected. However, we did confirm past results indicating that

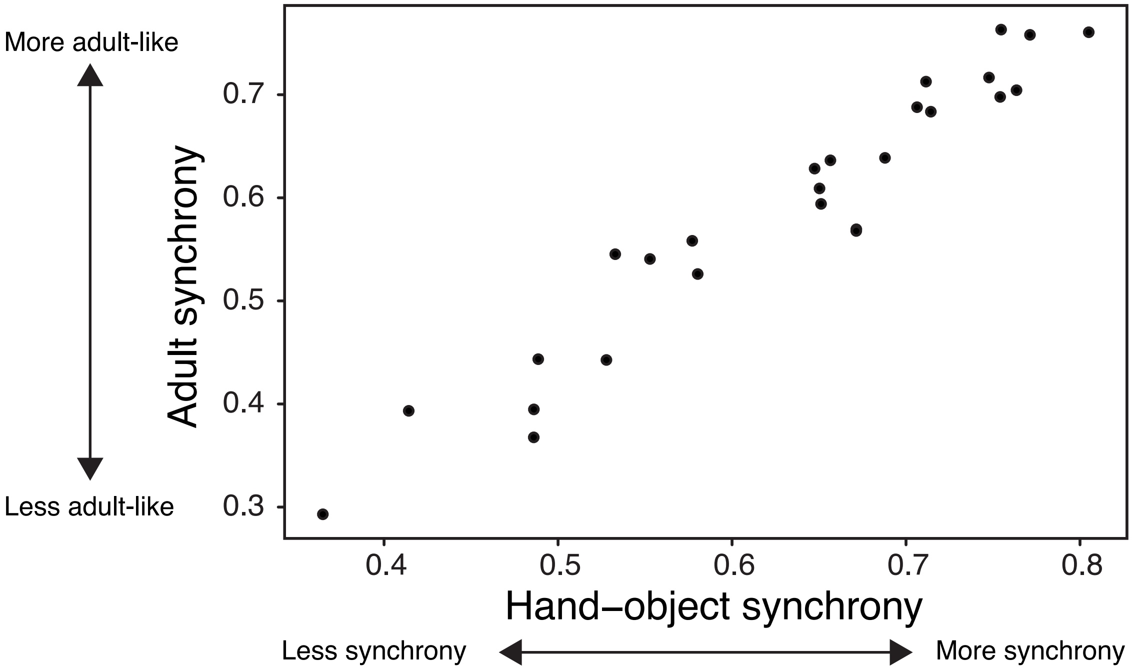


Figure 4.3: Relationship between hand-object synchrony and adult synchrony when viewing target action video

attention to hand-object actions provides considerable explanatory value in how similar observers are to adults.

Contrary to our expectation, opportunities to view a novel action did not influence attention. Perceptual experiences can provide opportunities to learn a wide range of information about what is meaningful during an action but we did not observe any effect of condition. One possibility is a ceiling effect in attentional synchrony with adults and hand-object actions for this particular action. This might occur if the action is too simplistic for this age and opportunities to learn about the action provide no additional knowledge. While the average correlation values in adult synchrony and hand-object synchrony were high, Figure 4.3 indicates that both adult synchrony and hand-object synchrony varied quite widely. Adult synchrony ranged from a minimum = .29 to a maximum = .76 (*SD*= .13) and hand-object synchrony had a similar range with a minimum = .36 and a maximum = .80 (*SD*= .11). Therefore it is unlikely to be a ceiling effect, as there were improvements to be gained by at least some of the 4-year-olds.

As it is applied in this study, adult synchrony and hand-object synchrony are likely to be related to action comprehension in ways that limit our conclusions. In this action, greater adult synchrony was related to looking at the hands interacting with objects. This is exactly what would be expected for knowledgeable adults: attention should being synchronized to meaningful hand-object interactions. As a closely related metric, high comprehension would also lead attention to hand-object actions. Therefore, comprehension is expected to be related to greater hand-object synchrony and greater adult synchrony. But this is not obligatory of these measures. Rather, this particular stimulus likely presents an action that leads to the collinearity that is observed in Figure 4.3. In a different hypothetical action stimulus, we can imagine that greater comprehension might lead attention to an area that is not the hands. For example, if the action involved manipulating a puzzle box to navigate out a rolling marble, then attention would likely be allocated to the marble as the most meaningful information. Simply looking like an adult (or looking to hand-object interactions) does not equate to understanding the meaningful information involved in the action.

Our attempt to influence comprehension of and attention to meaningful information was not effective. In past work that changed attention or comprehension, the most effective manipulations have involved providing experiences that allow for self-directed exploration of the action. For example, in studies involving “sticky mittens” (Needham et al., 2002; Sommerville et al., 2005), infants are given repeated opportunities to engage in the action themselves, with longer training involving 10 minutes a day for two weeks and shorter changes occurring only after 200 seconds of training. Opportunities to discover the relationship between their own actions and ‘grasping’ the Velcro objects may provide a stronger or longer-lasting change to the infants’ comprehension of meaning in grasps and attention to meaningful aspects of grasp-like actions they subsequently observe. Some researchers point to activation of the infants’ own motor system as the reason why these manipulations are effective (Libertus & Hauf, 2017). While the role of the motor system is crucial, simply observing actions can still influence an observer’s comprehension of action. Repeated, self-paced viewing of actions lead to reorganization of observers understanding of an action’s hierarchical structure (Hard et al., 2019). Self-paced viewing is an opportunity for observers to discover meaningful information about action structure by explore the stream of action at their own pace. Another common aspect of successful manipulations is the repeated nature of the opportunities provided. In line with theories of statistical learning (Saffran & Kirkham, 2018), repeated exposures help to reveal regularities in the type of information that is important to look at or that change comprehension. Therefore, we suggest that a sufficient manipulation in the current study may involve allowing children to explore the action by repeatedly attempting the action on their own.

### 4.10.1 Limitations and future direction

In this study, the stimulus was selected to provide an action that was just within the motor abilities of 4-year-olds but unfamiliar. However, there is a striking degree of similarity between adult synchrony and hand-object synchrony. How much of this similarity can be attributed to this specific action and how much can be attributed to all goal-directed manual actions? In this study, the action stimulus was designed to be fully in-view, with no face visible, take up a large area of a screen, and be goal-directed with no interleaved breaks or hesitations. These factors likely contribute towards there being few, if any, times that adults look to a location that did not involve hands and objects. However, coordination between eye movements and hands is a know feature of manual action (Flanagan & Johansson, 2003; Rotman et al., 2006) with only a couple hundred milliseconds of difference between eyes and hands. In future studies, consideration towards tasks and the metrics used to measure them should be central, especially when exploring new measurements.

Within this stimulus, there is a variety of actions involving simple grasps to more complex syringe use. While attentional synchrony with adults and hand-object synchrony is sensitive to spatiotemporal variability, investigating specific portions of the action may reveal differences in attention. For portions that show simple grasps, children's synchrony with adults could be high, but this may not be the case for the more novel actions involving the syringe. A closer analysis of that separates out different actions may show that our study conditions were effective at changing attention but only for actions that were entirely novel to 4-year-olds.

The action literature provides detailed descriptions of attention to simple actions. One consistent finding is that eye movements of experienced observers are predictive in looking to goal locations. Novice observers tend to be reactive and follow behind the hand. However, we did not examine this aspect of attention. Identifying times at which attention is most synchronous, predictive, and reactive, compared to adults may show that children are predictive of familiar actions but reactive for novel actions. Time lagged cross correlations would be helpful in this endeavor.

# Chapter 5

# Conclusions

In this dissertation, I described the results of three studies that add to our understanding of visual attention development in dynamic scenes. In Chapter 2, I reported the role of two visual features, faces and salience, to test whether there is a developmental shift from bottom-up to top-down attention. The results revealed that visual attention is a dynamic process, in which people prioritize which features are important from moment to moment. Building on the idea of prioritization, Chapter 3 measured developmental changes in attention to hands and hand-object actions. Using newly-developed measurements of synchrony, I found further evidence in support of a developmental account of visual attention that emphasizes prioritization. In Chapter 4, I applied this developmental account to visual attention when viewing goal-directed manual actions. To test whether action comprehension changes visual attention to actions, I experimentally manipulated 4-year-olds' prior experience in viewing an action sequence. However, I found no difference in subsequent action attention between children who had previously viewed the action sequence and those who had not. In the investigation of developmental changes in visual attention, I revealed complex variability and applied nuanced perspectives to develop better conceptual approaches that takes this complexity into consideration.

The main theoretical contribution from the present work is showing that feature accounts of development, whether bottom-up or top-down, fail to capture age-related changes in visual attention when meaning is not considered. Whereas prior work (Franchak et al., 2016; Frank et al., 2009; Kwon et al., 2016; Rider et al., 2018) suggested that attention develops via changes in looking at faces and salient locations, we provided evidence that development is unlikely to involve a global shift from bottom-up to top-down features. Change occurs quantitatively, rather than qualitatively, in the ability to prioritize meaningful information. In Chapter 3, I directly tested whether a general measure (adult synchrony) that captures the spatiotemporal variability of attention is predicted by age. Indeed, unlike faces or salience in Chapter 2, there was a consistent effect of age on adult synchrony. This suggests that changes in how more mature participants prioritize what is meaningful to look at cannot be reduced to simply, "look less at salient areas and more at faces". Although the increase in adult synchrony with age suggests that observers become increasingly better at prioritizing attention towards meaningful areas, an alternative explanation is that what infants think is meaningful is different compared to what adults think is meaningful. This cannot be ruled out from the present work, however, past work suggests this is not the case (Franchak et al., 2016): 6-month-olds and 9-month-olds showed highly idiosyncratic gaze patterns rather than settling on a single, "meaningful" area, whereas 24-month-olds showed greater consistency within their age group.

A secondary theoretical contribution from this work is identifying attention to hands and hand-object actions as meaningful features that show developmental increases in the organization of attention. The designs of both measures take into account the spatiotemporal variability of features within a stimulus and were directly informed by insights from Chapter 2. For naturalistic scenes on a screen, hands and manual actions are a developmentally informative feature for influencing how infants and children become more adult-like observers. Building on work from Chapter 4, future research could determine if hands and hand-object actions are important in developing adult-like attention beyond the screen. Prior work already points to the ability of 1-year old infants to use the hands of caretakers during object interactions to engage in joint attention without looking to the faces of their caretaker (Yu & Smith, 2013). Parents who leverage infants' attention to hand-object actions to scaffold joint attention can also provide opportunities for infants to discover other types of information about gaze following, coordinated social attention, object properties, or motor abilities. This illustrates the complexity of developing systems and how development of visual attention is inherently connected to development beyond simply looking to hands. Understanding antecedents that promote visual attention development and developmental cascades that arise from visual attention development will require rigorous interdisciplinary work.

My studies indicated that measuring meaning is vital for characterizing the development of visual attention beyond measuring simple features. Faces are not always meaningful and can change in meaning within a scene. What does the emphasis towards meaning suggest about visual attention theory as it relates to both bottom-up and top-down features? Firstly, given the overlap between salient locations and faces (Torralba et al., 2006; Wass & Smith, 2015), it is likely that locations in a scene that convey meaning are also visually salient. Conversely, bottom-up features can also be meaningful. Both bottom-up and top-down features are capable of cueing observers to areas of importance. Secondly, from a developmental perspective, the overlap of both feature types may provide redundant cues for infants and children to learn associations between features and areas of meaning. Redundancy may present statistical regularities that are detectable by infants (Saffran & Kirkham, 2018) and may structure the visual environment to promote scaffolding towards better comprehension of meaning or development of social abilities (Shepherd, 2010). Lastly, I suggest the separation of bottom-up and top-down features in a visual scene presents a false dichotomy. While bottom-up and top-down features are categorically distinct, the human visual attention system may organize these features hierarchically in service of determining meaning in a scene. For instance, low-level features in a scene like contours, colors, or motion may support better conceptualization of meaningful features like objects and social agents which in turn supports even more complex meaning like features involved in the intentions of people or information relevant to tasks. Observers are expected to attend to features that convey meaning but where would a person look if they were faced with a scene that they did not understand like watching an unfamiliar sport? In a hierarchically organized system of meaning, an observer may rely on a lower-level of features to find meaning like looking to the ball or perhaps attending to salient features like where the most movement is occurring. In recent work, my collaborators and I found support for this notion by showing salience was a better predictor of attention for adults who view videos with shots edited to occur out of order compared to adults who viewed intact videos (Jing et al., in press). When there is not enough information to select a meaningful location, observers may rely on features that are lower in the hierarchy of meaning. This may also explain why salience better predicts attention during the first fixations of static images (Parkhurst & Niebur, 2003).

A hierarchical organization of meaning for visual attention would also apply to the development of attention. If visual attention is allocated to the highest level of meaning available for an observer, this could still account for the attention of young infants as 'stimulus-driven' (Oakes & Amso, 2018; Stechler & Latz, 1966). In this case, infants may have an undeveloped hierarchical system of what is meaningful and therefore use less informative features, yet features that are most meaningful to them, which would be low-level salient features. Infants may choose to look to meaningful information that is appropriate for their level of comprehension complexity. Past studies have found a similar self-selection in infants for visual information that is not too complex nor too simple (Kidd, Piantadosi, & Aslin, 2012) and has been used to explain novelty/familiarity preferences in infants (Kidd et al., 2012; Kidd, Piantadosi, & Aslin, 2014). Kidd et al. (2012) suggest several explanations of this phenomenon including difficulty of encoding, efficiency of computational resources, and selection of the optimal level of complexity for learning. Despite uncertainty of the mechanism, infants' attention based on a hierarchical organization of meaning seems compatible with past work indicating infants choose to attend to information that exists within a cognitive 'Goldilocks' zone.

If meaning is organized hierarchically based on the observers individual experiences, there are interesting implications to how this applies to cultural differences. For instance, cultural-specific practices or tools may lead to different comprehension of meaning in a scene. One such difference is observed in attention to the use of eating utensils between Swedish and Chinese infants in predicting looks to the mouth based on whether spoons or chopsticks are used (Green et al., 2016). It could be argued that differences in infants' attention to eating utensils comes from a more general deficit in cognitive reasoning and inexperience with human actions rather than a cultural difference. However, adults also display cultural differences in understanding meaning when comparing naive and experienced observes. Comprehension of video clips was poorer for a group of adults from Turkey with no prior media experience compared to adults with either low or high media experience (Ildirar & Schwan, 2015). Without a developed hierarchy of meaning for media conventions, comprehension of scenes that involved cuts led to failures in interpreting temporal and spatial relations between scenes (e.g. cuts to a different angle of an animal misinterpreted as the animal rotating or cuts from establishing shots outside to inside not being connected as the same location or time). These findings suggest that acquired experience from everyday visual experiences beyond formal learning environments support a more developed hierarchy of meaning. Day-to-day experiences allow for transmission of culture-specific concepts indicative of socio-cultural learning processes for visual attention development. While a hierarchical account of meaning of visual features seems to fit well with past and current findings, this perspective is speculative. Closer examination, perhaps cross-culturally, would help to test this account.

Beyond theories of attention, the studies in this dissertation also made methodological contributions to the study of visual attention. Chapter 2 demonstrated that variability in stimuli should be the norm. After finding no global effect of faces or salient region on age, further examination of stimuli revealed patterns within stimuli that occurred when 10 second windows were calculated independently. Variations within and between stimuli were striking and revealed that attention was incredibly complex. Studies that use a single stimulus are not accurately representing the stimuli space in which the developmental phenomena occurs in. Furthermore, given the distribution of age correlations during 10 second windows that fell within the 95% confidence level in Figures 2.4 and 2.6, studies that continue to use singular or short duration stimuli are prone to find null effects even if there are developmental changes in prioritization of meaning.

Chapter 3 and 4 provided insight on how measures of attentional synchrony can provide a way to measure attention in complex, dynamic stimuli. Adult synchrony captures spatiotemporal variability in attention while also indexing development of visual attention. Because it is based on the eye movements of adult participants, it avoids selective biases that researchers introduce by choosing which features to measure (e.g., faces). Most importantly, adult synchrony is a way of capturing what is meaningful in a scene without defining what meaning is. The consistent age-related increase in adult synchrony provides a powerful way to simultaneously test how different features account for the changes in prioritization of meaningful information. A general trend in the visual attention literature towards understanding meaning by having adults inform researchers about what is meaningful draws a clear parallel between adult synchrony in the dissertation and meaning maps described earlier. Although the stimuli are different for meaning maps and adult synchrony (images and videos, respectively), the contributions of both may indicate that less informative feature-based approaches will lose popularity in favor of understanding more complex influences, like meaning.

I also extended the synchrony measure approach to measure attention to hands and hand actions. Basic descriptive statistics, such as proportion of time looking at faces, cannot account for the spatiotemporal variability in face looking over a stimulus. However, the new hand synchrony and hand-object synchrony measures did predict changes in adult-synchrony. In determining other features that may influence attention, I provide two considerations 1) considering the meaning that the feature cues and 2) the times at which the feature is most meaningful.

Across the three studies, findings suggest that the development of visual attention in dynamic scenes involves improvements in moment-to-moment prioritization of attention to meaningful information. Evidence from Chapter 4 about the mechanism of change is inconclusive, however, I propose that perceptual experiences reveal regularities about actions that improves comprehension. While still technically compatible with theoretical approaches that categorize the influence of bottom-up and top-down features, I suggest that moving away from features as singular stable factors towards consideration of *when* and *where* these features convey meaning is a step towards a more comprehensive understanding of visual attention development.

# References

Ambrosini, E., Reddy, V., De Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PLoS ONE*, *8*(7), e67916. doi: 10.1371/journal.pone.0067916

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PLoS ONE*, *9*, 1–7. doi: 10.1371/journal.pone.0085701

Anderson, D. R., Lorch, E. P., Field, D. E., & Sanders, J. (1981). The effects of tv program comprehensibility on preschool children's visual attention to television. *Child Development*, *20*, 151–157. doi: 10.2307/ 1129224

Aring, E., Grönlund, M. A., Hellström, A., & Ygge, J. (2007). Visual fixation development in children. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*(11), 1659–1665.

Açik, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Frontiers in Psychology*, *1*. doi: 10.3389/fpsyg.2010.00207

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, *17*, 1185–1204. doi: 10.1080/13506280902978477

Barrett, T. M., Davis, E. F., & Needham, A. (2007). Learning about tools in infancy. *Developmental Psychology*, *43*(2), 352.

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

Bertenthal, B., & Boyer, T. (2012). Developmental changes in infants' visual attention to pointing. *Journal of Vision*, *12*(9), 480–480.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *Quarterly Journal of Experimental Psychology*, *61*, 986–998. doi: 10.1080/17470210701410375

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, *49*, 2992–3000. doi: 10.1016/j.visres.2009.09.014

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 185–207. doi: 10.1109/TPAMI.2012.89

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology*, *103*(6), 3389–3397.

Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development*, *65*, 1243–1261. doi: 10.2307/1131497

Bruce, V. (1993). What the human face tells the human mind: Some challenges for the robot-human interface. *Advanced Robotics*, *8*, 341–355. doi: 10.1163/156855394X00149

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15*(2), 292–298.

Cashon, C. H., & Cohen, L. B. (2004). Beyond U-shaped development in infants' processing of faces: An information-processing account. *Journal of Cognition and Development*, *5*, 59–80. doi: 10.1207/s15327647jcd0501$_4$

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*, 1–15. doi: 10.1167/9.3.6

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

Colombo, J., Mitchell, D. W., Coldren, J. T., & Atwater, J. D. (1990). Discrimination learning during the first year: Stimulus and positional cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 98–109. doi: 10.1037/0278-7393.16.1.98

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*, 1-17. doi: 10.1167/14.8.5

Darby, K. P., Deng, S. W., Walther, D. B., & Sloutsky, V. M. (2021). The development of attention to objects and scenes: From object-biased to unbiased. *Child Development*, *92*(3), 1173–1186.

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

de Villiers Rader, N., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture*, *10*(2-3), 202–221.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*, 28. doi: 10.1167/10.10.28

Farmer, H., Bevan, C., Green, D., Rose, M., Cater, K., & Stanton Fraser, D. (2021). Did you see what i saw?: Comparing attentional synchrony during 360° video viewing in head mounted display and tablets. *Journal of Experimental Psychology: Applied*, *27*(2), 324.

Farzin, F., Hou, C., & Norcia, A. M. (2012). Piecing it together: Infants' neural responses to face and object structure. *Journal of Vision*, *12*, 6–6. doi: 10.1167/12.13.6

Farzin, F., Rivera, S. M., & Whitney, D. (2011). Time crawls: the temporal resolution of infants' visual attention. *Psychological Science*, *22*, 1004–1010. doi: 10.1177/0956797611413291

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. doi: 10.1016/j.cognition.2016.03.005

Filippi, C. A., & Woodward, A. L. (2016). Action experience changes attention to kinematic cues. *Frontiers in Psychology*, *7*, 19. doi: 10.3389/fpsyg.2016.00019

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. doi: 10.1038/nature01861

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

Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free viewing gaze behavior in infants and adults. *Infancy*, *21*, 262–287. doi: 10.1111/infa.12119

Franchak, J. M., & Kadooka, K. (2022). Age differences in orienting to faces in dynamic scenes depend on face centering, not visual saliency. *Infancy*.

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

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology*, *118*, 13–26. doi: 10.1016/j.jecp.2013.08.012

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170. doi: 10.1016/j.cognition.2008.11.010

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy*, *17*, 355–375. doi: 10.1111/j.1532-7078.2011.00086.x

Gannon, E. T., & Grubb, M. A. (2022). How filmmakers guide the eye: The effect of average shot length on intersubject attentional synchrony. *Psychology of Aesthetics, Creativity, and the Arts*, *16*(1), 125.

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00527

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in biology and medicine*, *37*(7), 957–964.

Green, D., Li, Q., Lockman, J. J., & Gredebäck, G. (2016). Culture influences action understanding in infancy: Prediction of actions performed with chopsticks and spoons in chinese and swedish infants. *Child Development*, *87*(3), 736–746. doi: 10.1111/cdev.12500

Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition*, *47*(1), 17–32. doi: 10.3758/s13421-018-0847-z

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. In *Proceedings of the 19th International Conference on Neural Information Processing Systems* (pp. 545–552). Cambridge, MA: MIT Press. doi: 10.7551/mitpress/7503.003.0073

Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., Konig, P., ... Einhäuser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*, 1132–1158. doi: 10.1080/13506280902812304

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication monographs*, *85*(1), 4–40.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 6–6.

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, *103*, 83–91. doi: 10.1016/j.visres.2014.08.006

Helo, A., Rämä, P., Pannasch, S., & Meary, D. (2016). Eye movement patterns and visual attention during scene viewing in 3-to 12-month-olds. *Visual Neuroscience*, *33*, 1–7. doi: 10.1017/S0952523816000110

Helo, A., van Ommen, S., Pannasch, S., Danteny-Dordoigne, L., & Rämä, P. (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behavior and Development*, *49*, 248–266. doi: 10.1016/j.infbeh.2017.09.008

Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, *21*, 15–23. doi: 10.1016/j.tics.2016.11.003

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window* *on mind and brain* (pp. 537–562). Oxford: Elsevier. doi: 10.1016/b978-008044980-7/50027-6

Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, *18*, 1–18. doi: 10.1167/18.6.10

Ildirar, S., & Schwan, S. (2015). First-time viewers' comprehension of films: Bridging shot transitions. *British Journal of Psychology*, *106*(1), 133–151.

Itti, L., & Baldi, P. (2005). A principled approach to detecting surprising events in video. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 631–637). IEEE.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*, 1295–1306. doi: 10.1016/j.visres.2008.09.007

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. doi: 10.1016/s0042-6989(99)00163-7

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

Jing, M., Kadooka, K., Franchak, J., & Kirkorian, H. L. (in press). The effect of narrative coherence and visual salience on children's and adults' gaze while watching video. *Journal of Experimental Child Psychology*.

Jost, T., Ouerhani, N., Wartburg, R. V., Müri, R., & Hügli, H. (2005). Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, *100*, 107–123. doi: 10.1016/j.cviu.2004.10.009

Kadooka, K., & Franchak, J. M. (2020). Developmental changes in infants' and children's attention to faces and salient regions vary across and within video stimuli. *Developmental Psychology*, *56*(11), 2065.

Kadooka, K., & Franchak, J. M. (in prep). Attention to hands during manual actions account for developmental increases in attentional synchrony.

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature communications*, *2*(1), 1–6.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, *7*(5), e36399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child development*, *85*(5), 1795–1804.

Kirkorian, H. L., & Anderson, D. R. (2017). Anticipatory eye movements while watching continuous action across shots in video sequences: A developmental study. *Child Development*, *88*(4), 1284–1301.

Kirkorian, H. L., & Anderson, D. R. (2018). Effect of sequential video shot comprehensibility on attentional synchrony: A comparison of children and adults. *Proceedings of the National Academy of Sciences*, *115*, 9867–9874. doi: 10.1073/pnas.1611606114

Kirkorian, H. L., Anderson, D. R., & Keen, R. (2012). Age differences in online processing of video: An eye movement study. *Child Development*, *83*, 497–507. doi: 10.1111/j.1467-8624.2011.01719.x

Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? the role of motor representations in semantic sensibility judgments. *Journal of memory and language*, *28*(1), 56–77.

Klin, A., & Jones, W. (2008). Altered face scanning and impaired recognition of biological motion in a 15-month-old infant with autism. *Developmental Science*, *11*, 40–46. doi: 10.1111/j.1467-7687.2007.00608.x

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*, 809–816. doi: 10.1001/archpsyc.59.9.809

Kochukhova, O., & Gredebäck, G. (2010). Preverbal infants anticipate that food will be brought to the mouth: An eye tracking study of manual feeding and flying spoons. *Child Development*, *81*(6), 1729–1738.

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

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, *26*, 51–62. doi: 10.1017/S0952523808080899

Land, M. F., & Fernald, R. D. (1992). The evolution of eyes. *Annual Review of Neuroscience*, *15*, 1–29. doi: 10.1146/annurev.ne.15.030192.000245

Land, M. F., & McLeod, P. (2000). From eye movements to actions: Batsmen hit the ball. *Nature Neuroscience*, *3*, 1340–1345. doi: 10.1038/81887

Libertus, K., & Hauf, P. (2017). *Motor skills and their foundational role for perceptual, social, and cognitive development* (Vol. 8). Frontiers Media SA.

Libertus, K., Landa, R. J., & Haworth, J. L. (2017). Development of attention to faces during the first 3 years: Influences of stimulus type. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.01976

Libertus, K., & Needham, A. (2010). Teach to reach: The effects of active vs. passive reaching experiences on action and perception. *Vision Research*, *50*(24), 2750–2757. doi: 10.1017/S0952523808080899

Lorch, E. P., & Castle, V. J. (1997). Preschool children's attention to television: Visual attention and probe response times. *Journal of Experimental Child Psychology*, *66*, 111–127. doi: 10.1006/jecp.1997.2372

Mahdi, A., Su, M., Schlesinger, M., & Qin, J. (2017). A comparison study of saliency models for fixation prediction on infants and adults. *IEEE Transactions on Cognitive and Developmental Systems*, *10*, 485–498. doi: 10.1109/tcds.2017.2696439

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, *3*, 5–24. doi: 10.1007/s12559-010-9074-z

Monroy, C., Gerson, S., & Hunnius, S. (2017). Infants' motor proficiency and statistical learning for actions. *Frontiers in Psychology*, *8*, 2174. doi: 10.3389/fpsyg.2017.02174

Moore, C. G., & Müller, S. (2014). Transfer of expert visual anticipation to a similar domain. *Quarterly Journal of Experimental Psychology*, *67*(1), 186–196. doi: 10.1080/17470218.2013.798003

Mulder, H., Oudgenoeg-Paz, O., Verhagen, J., van der Ham, I. J., & Van der Stigchel, S. (2022). Infant walking experience is related to the development of selective attention. *Journal of Experimental Child Psychology*, *220*, 105425.

Napier, J. R. (1956). The prehensile movements of the human hand. *The Journal of bone and joint surgery. British volume*, *38*(4), 902–913.

Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills. *Infant Behavior and Development*, *25*(3), 279–295. doi: 10.1016/S0163-6383(02)00097-8

Oakes, L. M., & Amso, D. (2018). The development of visual attention. In J. Wixted (Ed.), *The Steven's Handbook of Experimental Psychology and Cognitive Neuroscience* (4th ed., Vol. 4, pp. 1–33). New York: Wiley. doi: 10.1002/9781119170174.epcn401

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*(2), 125–154.

Parkhurst, D. J., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience*, *19*, 783–789. doi: 10.1111/j.0953-816x.2003.03183.x

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*, 1321–1323. doi: 10.1126/science.1070223

Paus, T., Babenko, V., & Radil, T. (1990). Development of an ability to maintain verbally instructed central gaze fixation studied in 8-to 10-year-old children. *International journal of Psychophysiology*, *10*(1), 53–61.

Pempek, T. A., Kirkorian, H. L., Richards, J. E., Anderson, D. R., Lund, A. F., & Stevens, M. (2010). Video comprehensibility and attention in very young children. *Developmental Psychology*, *46*, 1283–1293. doi: 10.1037/a0020614

Pereira, E. J., Birmingham, E., & Ristic, J. (2019). The eyes do not have it after all? attention is not automatically biased towards faces and eyes. *Psychological Research*, 1–17. doi: 10.1007/s00426-018-1130-4

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*, 2397–2416. doi: 10.1016/j.visres.2005.03.019

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rehrig, G., Peacock, C. E., Hayes, T. R., Henderson, J. M., & Ferreira, F. (2020). Where the action could be: Speakers look at graspable objects and meaningful scene regions when describing potential actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(9), 1659.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, *10*, 341–350. doi: 10.1088/0954-898x/10/4/304

Rideout, V. (2017). The common sense census: Media use by kids age zero to eight. *San Francisco, CA: Common Sense Media*.

Rider, A. T., Coutrot, A., Pellicano, E., Dakin, S. C., & Mareschal, I. (2018). Semantic content outweighs low-level saliency in determining children's and adults' fixation of movies. *Journal of Experimental Child Psychology*, *166*, 293–309. doi: 10.1016/j.jecp.2017.09.002

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*, 1–20. doi: 10.1167/7.14.16

Rotman, G., Troje, N. F., Johansson, R. S., & Flanagan, J. R. (2006). Eye movements when observing predictable and unpredictable actions. *Journal of Neurophysiology*, *96*(3), 1358–1369. doi: 10.1152/jn.00227.2006

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*, 454–464. doi: 10.1037/0012-1649.34.3.454

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, *69*, 181.

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye–hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience*, *25*, 8833–8842. doi: 10.1523/jneurosci.2658-05.2005

Schiller, P. H. (1998). The neural control of visually guided eye movements. In *Cognitive neuroscience of attention* (pp. 13–60). Psychology Press.

Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in integrative neuroscience*, *4*, 5.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology*, *20*, 649–656. doi: 10.1016/j.cub.2010.02.032

Sherman, R. A., & Serfass, D. G. (2015). The comprehensive approach to analyzing multivariate constructs. *Journal of Research in Personality*, *54*, 40–50. doi: 10.1016/j.jrp.2014.05.002

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.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*, 1–24. doi: 10.1167/13.8.16

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11. doi: 10.1016/j.cognition.2004.07.004

Spelke, E. S. (1990). Principles of object perception. *Cognitive science*, *14*(1), 29–56.

Stechler, G., & Latz, E. (1966). Some observations on attention and arousal in the human infant. *Journal of the American Academy of Child Psychiatry*.

Stoesz, B. M., & Jakobson, L. S. (2014). Developmental changes in attention to faces and bodies in static and dynamic scenes. *Frontiers in Psychology*, *5*, 193. doi: 10.3389/fpsyg.2014.00193

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*, 1–23. doi: 10.1167/11.5.5

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722. doi: 10.1111/j.1467-8624.2007.01025.x.

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

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

Tseng, P.-H., Carmi, R., Cameron, I. G., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, *9*, 1–16. doi: 10.1167/12.13.3

van Renswoude, D. R., Visser, I., Raijmakers, M. E., Tsang, T., & Johnson, S. P. (2019). Real-world scene perception in infants: What factors guide attention allocation? *Infancy*, *24*(5), 693–717.

Veer, I. M., Luyten, H., Mulder, H., van Tuijl, C., & Sleegers, P. J. (2017). Selective attention relates to the development of executive functions in 2, 5-to 3-year-olds: A longitudinal study. *Early childhood research quarterly*, *41*, 84–94.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*(3), 24–24.

Võ, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*, 1-14. doi: 10.1167/14.8.5

Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, *12*, 1–27. doi: 10.1167/12.1.16

Wartella, E., Richert, R. A., & Robb, M. B. (2010). Babies, television and videos: How did we get here? *Developmental Review*, *30*, 116–127. doi: 10.1016/j.dr.2010.03.008

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy*, *19*, 427–460. doi: 10.1111/infa.12055

Wass, S. V., & Smith, T. J. (2015). Visual motherese? Signal-to-noise ratios in toddler-directed television. *Developmental Science*, *18*, 24–37. doi: 10.1111/desc.12156

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*, 229–250. doi: 10.3758/s13428-012-0245-6

Westheimer, G. (1982). The spatial grain of the perifoveal visual field. *Vision Research*, *22*, 157–162. doi: 10.1016/0042-6989(82)90177-8

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied*, *8*(4), 259. doi: 10.1037//1076-898x.8.4.259

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34.

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current directions in psychological science*, *18*(1), 53–57.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum.

Yoshida, H., & Fausey, C. M. (2019). Visual objects as they are encountered by young language learners. *International Handbook of Language Acquisition*, 115–127.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180.

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

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin*, *127*(1), 3.

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the new York Academy of Sciences*, *1339*(1), 154–164.