

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

Kellan Kadooka and John M. Franchak

University of California, Riverside

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.

Keywords: Eye movements, visual attention, saliency, faces, social attention, media

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 in-

formation 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 & Baldi, 2005; Itti & Koch, 2000; Itti, Koch, & Niebur, 1998). 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

All authors certify that they have no affiliation or involvement with any entity with a financial or non-financial interest in the subject matter discussed in this manuscript.

We thank the members of the UCR Perception, Action, and Development lab for their help in coding the data. We are especially grateful to the participants and their families for volunteering to take part in the study.

Correspondence should be sent to Kellan Kadooka, Department of Psychology, University of California, Riverside, 900 University Avenue, Riverside, CA 92521. E-mail: kellan.kadooka@ucr.edu.

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.

Faces and salient locations attract adults' attention

Adults distribute their attention to socially-relevant locations, such as people's bodies (Foulsham, Walker, & Kingstone, 2011), eyes (Birmingham, Bischof, & Kingstone, 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, Heeger, Hasson, & Adolph, 2016; Frank, Vul, & Johnson, 2009; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010). Heightened attention to faces and facial features is found for both static (photo) and dynamic (video) stimuli (Birmingham, Bischof, & Kingstone, 2008; Birmingham et al., 2009; Võ, Smith, Mital, & Henderson, 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, Ouerhani, Wartburg, Müri, & Hügli, 2005), high contrast (Parkhurst & Niebur, 2004; Reinagel & Zador, 1999), and contain motion (Mital, Smith, Hill, & Henderson, 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 be-

cause 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 & Baldi, 2005; Itti & Koch, 2000; Itti et al., 1998). 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, Iyer, Itti, & Koch, 2005; Smith & Mital, 2013).

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, Haas, & Markant, 2014; Gluckman & Johnson, 2013; Kwon, Setodehnia, Baek, Luck, & Oakes, 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, Amso, & Johnson, 2014; Frank et al., 2009; Rider, Coutrot, Pellicano, Dakin, & Mareschal, 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, Hou, & Norcia, 2012; Pascalis, de Haan, & Nelson, 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 (Açık, Sarvary, Schultze-Kraft, Onat, & König, 2010; Helo, Pannasch, Sirri, & Rämä, 2014; Kwon et al., 2016), others find that saliency models are *more* predictive of adults' gaze compared with infants and children (Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018). One explanation is that faces tend to have higher saliency than irrelevant locations in a scene (Henderson, Brockmole, Castelhano, & Mack, 2007; Torralba, Oliva, Castelhano, & Henderson, 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; Frank et al., 2014, 2009; Kirkorian, Anderson, & Keen, 2012). 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., 2014, 2012; 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.

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 (Figure 1) to determine whether changes in visual attention to specific feature types are global, that is, invariant across stimuli. 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

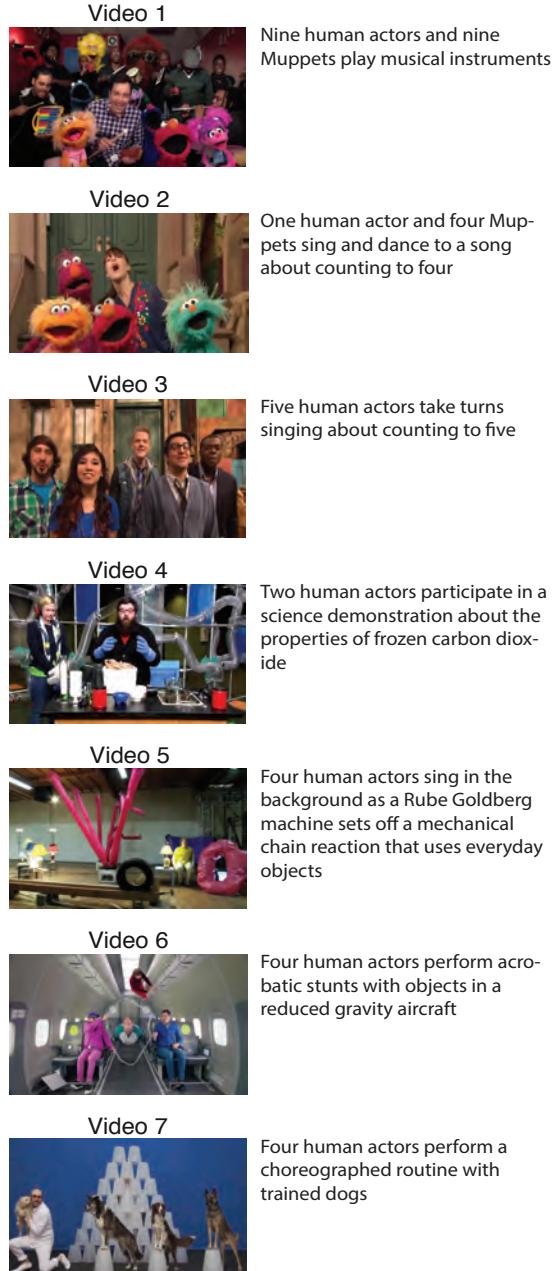


Figure 1. Screenshots and descriptions of the seven stimuli used in this study.

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 per-

formance 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.

Method

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 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.

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

Table 1

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

	<i>n</i>	<i>r</i>
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

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).

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 (Figure 1). 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-agentic movement. Each stimulus video, overlaid with data from infants and adults, are 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 chil-

dren's instrumental music. Every participant received the same pairing of music and video.

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 $31^\circ \times 19^\circ$. 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.

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 $> 1.5^\circ$ 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 $M = 0.54^\circ$ ($SD = 0.26$) 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 ($M = 0.66^\circ$, $SD = .29$) and adults ($M = 0.43^\circ$, $SD = .15$), 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 $M = 1.68^\circ$ ($SD = .32$) 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 stim-

ulus 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.

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 (red ellipses in Figure 2). *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; Smith & Mital, 2013; Tatler, Hayhoe, Land, & Ballard, 2011). 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. In Figure 2, an overlaid heatmap shows the saliency of different regions. 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.



Figure 2. Example video frames showing face AOIs (red ellipses) that were adjusted in size and location to accommodate moving agents. Heatmap overlay shows the relative saliency of different areas of the image (yellow = more salient, blue = less salient). In the top image, the actor's hand is the most salient region in the scene, whereas the actor's face is the most salient region in the bottom image.

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.

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 3), 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 \times 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 \times 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). 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 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 \times video interaction (Wald's $\chi^2 =$

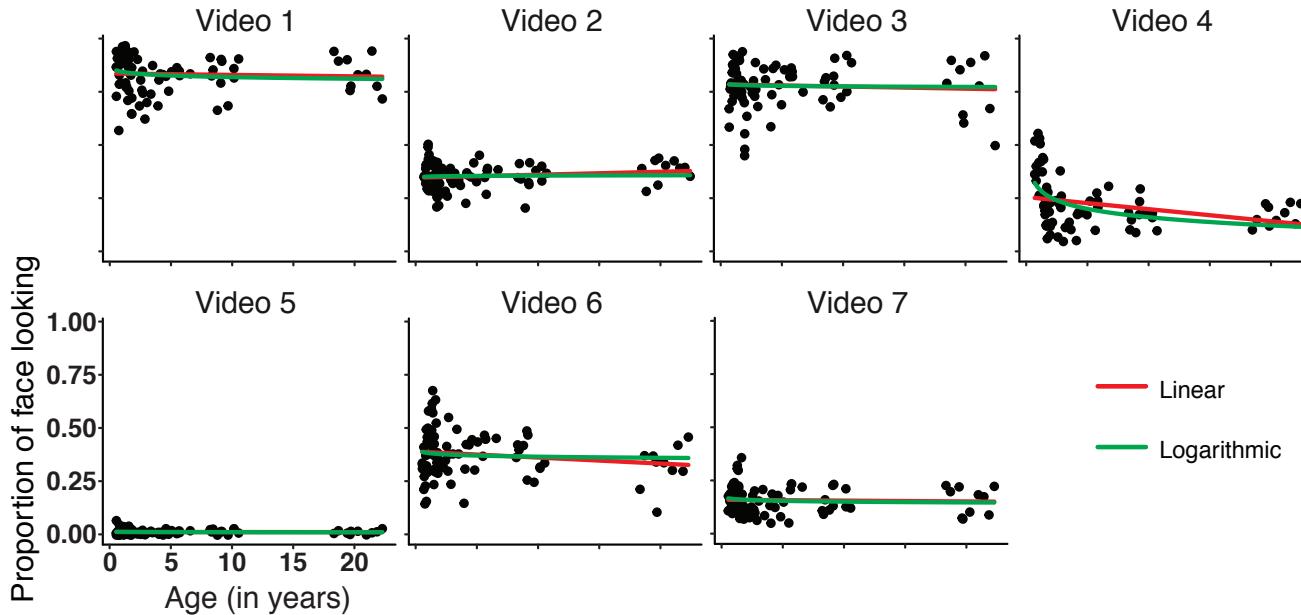


Figure 3. Changes in face looking as a function of age for all seven videos. Linear and logarithmic functions are plotted for each stimulus video.

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 \times 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 4, 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 \times 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 \times 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). 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 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 \times 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 \times video interaction (Wald's $\chi^2 = 14.50$, $p = .024$).

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

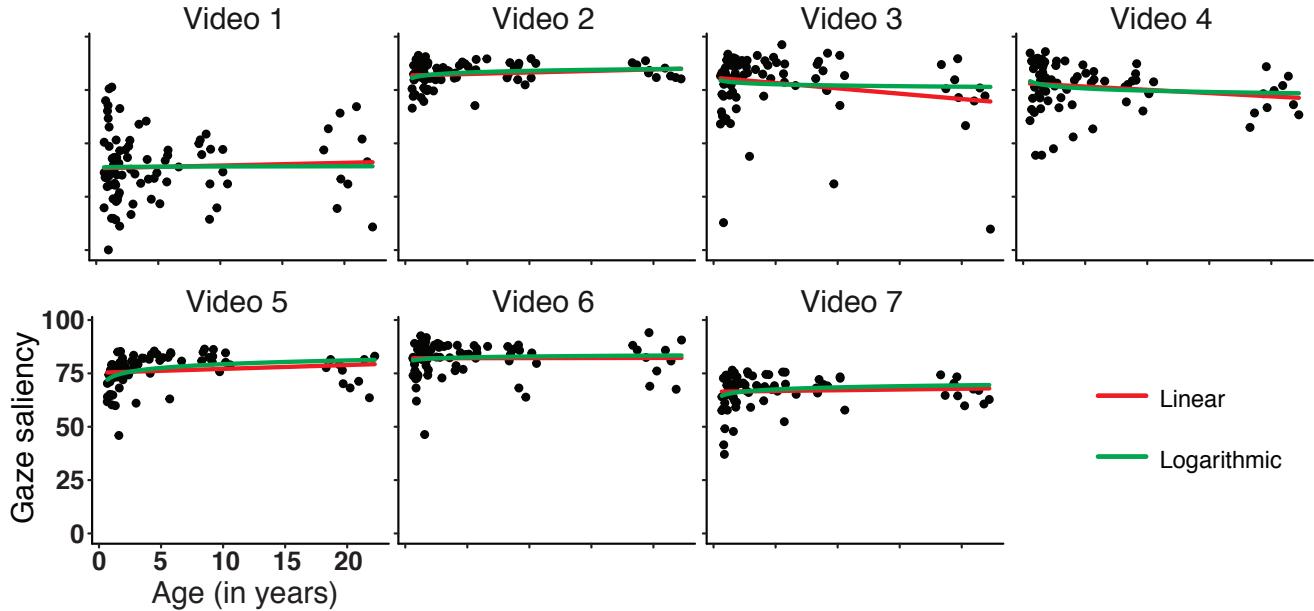


Figure 4. Changes in gaze saliency as a function of age for all seven videos. Linear and logarithmic functions are plotted for each stimulus video.

Table 2

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

Video	Face looking				Gaze saliency			
	Linear		Log		Linear		Log	
	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

* $p < .05$

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 \times 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 \times 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 direc-

tion 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 5 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 5, 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 5 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 3 shows that for all seven videos, there were significant age \times 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 \times window interactions by measuring the correlation between logarithmic age and face looking for each time window (e.g., the r values depicted in Figure 5). Figure 6A 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 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 priori-

tized 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 6B). 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 6B 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 6A). 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.

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 7 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.

Similar to face looking, seven GEE models were calculated to test for effects of age, window, and age \times window interactions on gaze saliency. As shown in Table 3, significant age \times 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

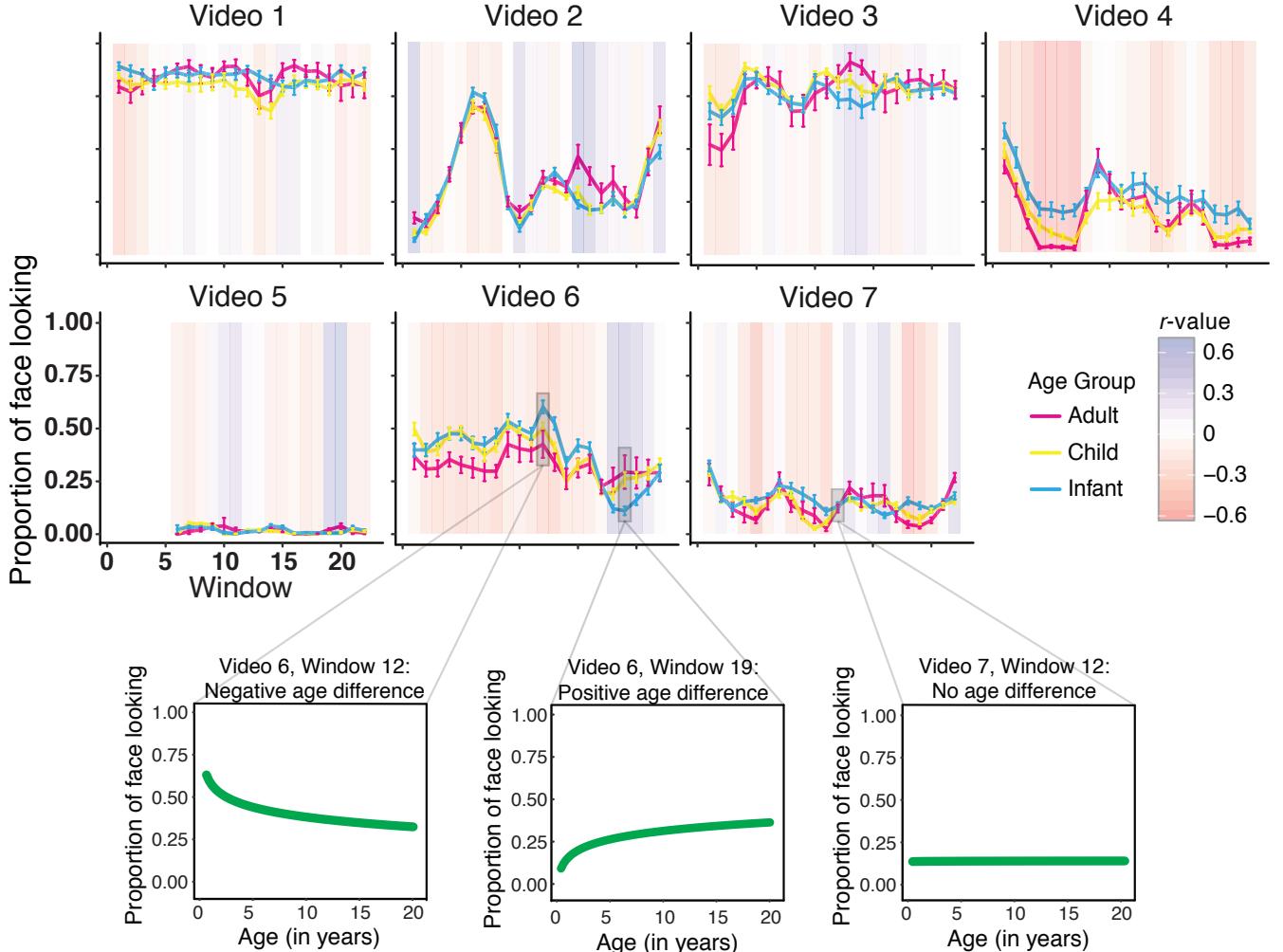


Figure 5. Windowed analyses of age differences in face looking over the duration of all 7 videos. Age was analyzed as a continuous variable, but for illustration purposes age was averaged into three groups (infants: 6-24 months; children: 2-10 years; adults: 18-22 years). Colored vertical bars represent 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 represented as a continuous predictor.

over time irrespective of age. Finally, as seen in the previous analyses, Videos 2 and 5 showed main effects of log age.

We explored the age \times window interactions by examining the distribution of all correlations between age and gaze saliency (Figure 8A). 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 8B 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

8A, 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).

Table 3

Generalized Estimating Equation Wald's χ^2 for effects of window, age, and age \times window for each video stimulus.

Video	Face looking			Gaze saliency		
	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*

* $p < .05$

A. Distribution of observed age by face looking correlations

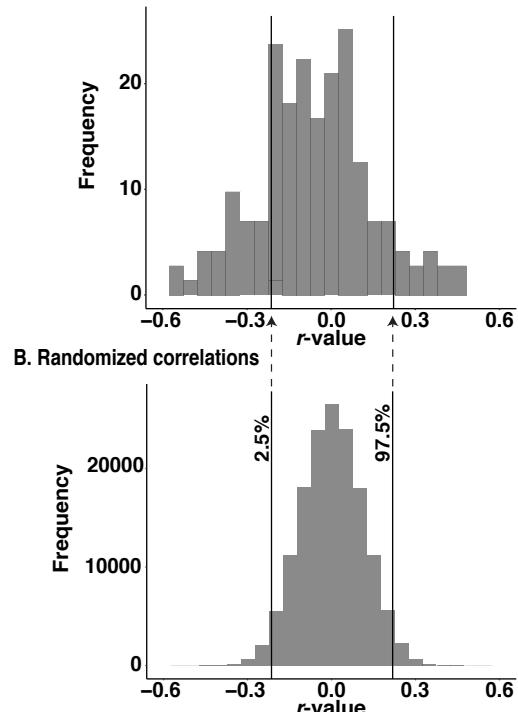


Figure 6. (A) Observed distribution and (B) randomized null distribution of correlations between age and face looking for each window aggregated across videos. Vertical black lines mark the 95% range of expected correlations in the null distribution.

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.

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 (Amso et al., 2014; Franchak et al., 2016; Frank et al., 2009; Kwon et al., 2016), decreases in looking to salient areas (Açık et al., 2010; Helo et al., 2014; Kwon et al., 2016), or increases in looking to salient areas (Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018). 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

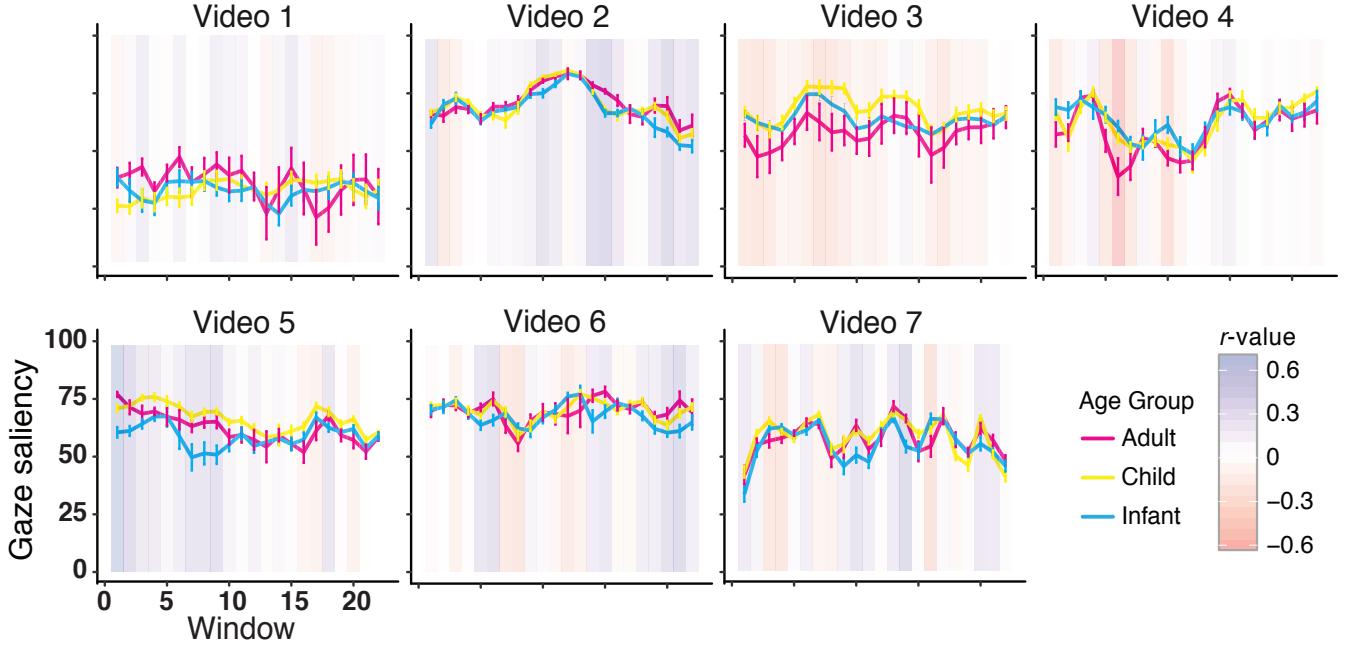


Figure 7. Windowed analyses of age differences in gaze saliency by video stimulus. Age was analyzed as a continuous variable, but for illustration purposes age was averaged into three groups (infants: 6-24 months; children: 2-10 years; adults: 18-22 years). Colored vertical bars represent strength and direction of correlation between age and gaze saliency for every window. Darker colors indicate stronger correlations.

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 (Kirkorian et al., 2012; Mital et al., 2011; Wang, Freeman, Merriam, Hasson, & Heeger, 2012). This has been attributed to an adult viewing strategy that expects screen-based media to center relevant information in the image frame (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

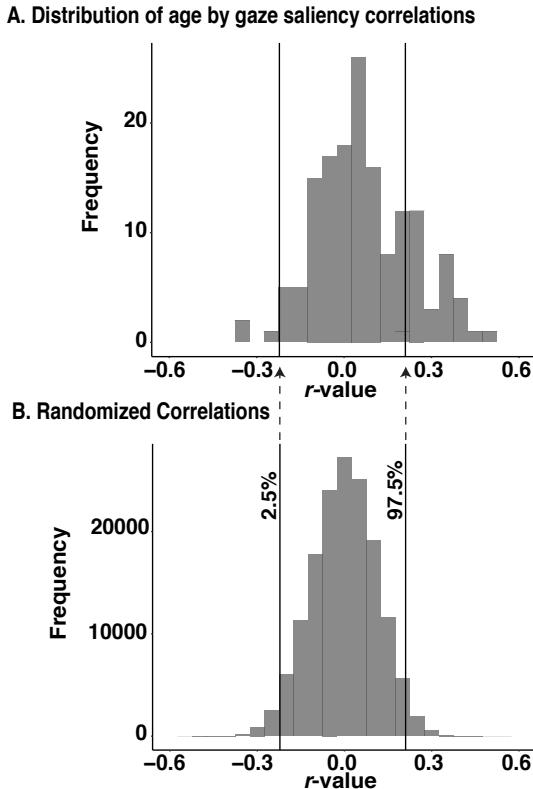


Figure 8. (A) Observed distribution and (B) randomized null distribution of correlations between age and gaze saliency for each window aggregated across videos. Vertical black lines mark the 95% range of correlations in the null distribution.

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 (Figure 1), 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 proper-

ties 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 (Amso et al., 2014; Açı̄k et al., 2010; Helo et al., 2014; Kwon et al., 2016), whereas studies that found inconsistent effects of face looking (Franchak et al., 2016; Frank et al., 2009) 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 (Franchak, Kretch, & Adolph, 2018; Yu & Smith, 2013)—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.

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 (Ballard & Hayhoe, 2009; Franchak et al., 2016; Henderson, 2017; Henderson & Hayes, 2018; Rothkopf, Ballard, & Hayhoe, 2007; 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 5 shows that face looking jumps from 13% to 72% and then back to 13% in a short time for participants of ev-

ery 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 \times 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 (Franchak et al., 2016; Frank et al., 2012). 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, van Ommen, Pannasch, Danteny-Dordogne, & Rämä, 2017; Kirkorian & Anderson, 2018; Kirkorian et al., 2012). 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, Lorch, Field, & Sanders, 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).

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 (Henderson & Hayes, 2018; Land & McLeod, 2000; Pereira, Birmingham, & Ristic, 2019; Sailer, Flanagan, & Johansson, 2005; Tatler et al., 2011). 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 (Amso et al., 2014; Wass & Smith, 2015) 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.

References

- 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.
- 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.
- Açık, A., Sarvary, 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.
- Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, 17, 1185–1204.
- 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.
- Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, 49, 2992–3000.
- Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 185–207.
- Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development*, 65, 1243–1261.
- Bruce, V. (1993). What the human face tells the human mind: Some challenges for the robot-human interface. *Advanced Robotics*, 8, 341–355.
- 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.
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, 52, 337–367.
- 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.
- Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, 14, 1–17.
- 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.
- Farzin, F., Rivera, S. M., & Whitney, D. (2011). Time crawls: the temporal resolution of infants' visual attention. *Psychological Science*, 22, 1004–1010.
- 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.
- Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free viewing gaze behavior in infants and adults. *Infancy*, 21, 262–287.
- 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.

- 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.
- Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, 110, 160–170.
- Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy*, 17, 355–375.
- Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, 4.
- 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.
- 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.
- 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.
- Helo, A., van Ommen, S., Pannasch, S., Danteny-Dordogne, 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.
- Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, 21, 15–23.
- 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.
- 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.
- 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.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Kwon, M.-K., Setodehnia, 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.
- Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, 26, 51–62.
- Land, M. F., & Fernald, R. D. (1992). The evolution of eyes. *Annual Review of Neuroscience*, 15, 1–29.
- Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, 3, 1340–1345.
- 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.
- 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.
- 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.
- 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.
- 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.
- Parkhurst, D. J., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience*, 19, 783–789.
- Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296, 1321–1323.
- 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.
- 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.
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, 45, 2397–2416.
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10, 341–350.

- 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.
- Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*, 1–20.
- 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.
- 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.
- 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.
- Sherman, R. A., & Serfass, D. G. (2015). The comprehensive approach to analyzing multivariate constructs. *Journal of Research in Personality, 54*, 40–50.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Wartella, E., Richert, R. A., & Robb, M. B. (2010). Babies, television and videos: How did we get here? *Developmental Review, 30*, 116–127.
- 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.
- Wass, S. V., & Smith, T. J. (2015). Visual motherese? Signal-to-noise ratios in toddler-directed television. *Developmental Science, 18*, 24–37.
- 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.
- Westheimer, G. (1982). The spatial grain of the perifoveal visual field. *Vision Research, 22*, 157–162.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.
- 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.