Age differences in orienting to faces in dynamic scenes depend on face centering, not visual saliency

John M. Franchak & Kellan Kadooka
University of California, Riverside

Authors' preprint, manuscript accepted at *Infancy*

The current study investigated how infants (6-24 months), children (2-12 years), and adults differ in how visual cues—visual saliency and centering—guide their attention to faces in videos. We report a secondary analysis of Kadooka and Franchak (2020), in which observers' eye movements were recorded during viewing of television clips containing a variety of faces. For every face on every video frame, we calculated its visual saliency (based on both static and dynamic image features) and calculated how close the face was to the center of the image. Results revealed that participants of every age looked more often at each face when it was more salient compared to less salient. In contrast, centering did not increase the likelihood that infants looked at a given face, but in later childhood and adulthood centering became a stronger cue for face looking. A control analysis determined that the age-related change in centering was specific to face looking; participants of all ages were more likely to look in the center of the image, and this center bias did not change with age. The implications for using videos in educational and diagnostic contexts are discussed.

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

In one of the most popular *Sesame Street* segments, musician Leslie Feist sings a song about counting to four with a group of Muppet background singers. The video has been viewed over 60 million times per year since it was posted on YouTube in 2008. When interviewed by the *New York Times* about the popularity of her performance, Feist remarked that she is frequently stopped by parents to take photographs with child fans who rarely recognize her:

> [Parents] say, "Do you mind, my 3-year-old has watched it 7,000 times," Feist said. "And I say yes, but I always joke: You notice me because you're a grown-up—the 3-year-olds are really only interested in the puppets. And without fail, the kids are just sort of looking at me like, who is this weird lady at the airport?" (Ryzik, 2019)

The authors thank the members of the Perception, Action, and Developmental Laboratory for their assistance with collecting and coding data for this project. The authors declare no conflict of interest. This project is a secondary analysis of a previously collected data set (Kadooka & Franchak, 2020, *Developmental Psychology*).

Corresponding author: John Franchak, Department of Psychology, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, franchak@ucr.edu

Feist's intuition is correct: An eye tracking study using her *Sesame Street* video found that infants (6-9 months) mostly looked at Muppet faces, whereas adults mostly looked at Feist's face (Franchak, Heeger, Hasson, & Adolph, 2016). There are several potential reasons for this age difference and, more generally, for developmental differences in how observers prioritize whether and when to look at faces in video media: developmental changes in visual attention (Colombo, 2001; Oakes & Amso, 2018), disparities in media-specific knowledge (Abelman, 1990; Kirkorian & Anderson, 2017), and age differences in plot comprehension (Kirkorian & Anderson, 2018; Pempek et al., 2010).

In the current study, we investigated two visual cues that are predictive of adult face looking during media viewing, face saliency and face centering (Amso, Haas, & Markant, 2014; Xu, Liu, Hu, & He, 2018). As we will review in the following sections, the *visual saliency* of a face—the degree to which a face stands out from the surrounding scene based on low-level visual features—might differentially influence face looking in participants of different ages depending on the development of visual attention. In contrast, the *centering* of a face—whether the face is in the middle versus the periphery of the scene—may influence looking behavior depending on how observers have acquired knowledge about media conventions (i.e., important faces are centered in the frame by the cinematographer). Although past work has indicated that *overall* looking patterns of infants are guided by both saliency (Franchak et al., 2016; Frank, Vul,

& Johnson, 2009) and centering (Mahdi, Su, Schlesinger, & Qin, 2017; van Renswoude, van den Berg, Raijmakers, & Visser, 2019; van Renswoude, Visser, Raijmakers, Tsang, & Johnson, 2019), it is unknown whether those cues facilitate face looking in infants and children while viewing dynamic scenes.

Although visual attention to faces in screen-based media differs from face-looking in more naturalistic settings with mobile observers (e.g., Franchak, Kretch, & Adolph, 2018), understanding what influences attention to faces in video media has several applications. First, deficits in infants' and children's ability to transfer what they have learned from media to real life (Barr, 2013; Troseth, 2010) reduce the effectiveness of educational videos and television programs (Wartella, Richert, & Robb, 2010). Because the faces of characters contain important information, age differences in orienting could prevent children from looking to the face of the most important character at any given time (Amso et al., 2014; Franchak et al., 2016). Second, there is growing interest in how looking to faces and to facial features could serve as an early biomarker for identifying Autism Spectrum Disorder (Jones & Klin, 2013; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Zwaigenbaum, Bryson, & Garon, 2013), however, studies have found inconsistent results about whether face looking differs between typically-developing children and children with ASD (Guillon, Hadjikhani, Baduel, & Rogé, 2014; Papagiannopoulou, Chitty, Hermens, Hickie, & Lagopoulos, 2014; Yurkovic et al., 2021). One source of inconsistency may be the visual features of face stimuli used in different investigations. Determining the extent to which orienting to faces depends on face saliency and centering—and whether orienting based on those features changes with age—will help address what characteristics of face stimuli should be controlled in diagnostic tests.

Is visual saliency a cue for face looking?

To understand the development of face looking, we must consider how different mechanisms that influence where people distribute their overt visual attention (i.e., where people direct their eyes in a scene) change with age. Bottom-up influences refer to attentional capture based on the visual appearance of a stimulus. Areas of a scene that are visually salient (i.e., conspicuous) attract attention: When a region differs from its surroundings based on one or more visual features—intensity, color, orientation, and motion—it stands out and is more likely to be fixated (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998). For example, the moving lips of a person's face in an otherwise motionless image will make the face more salient compared to its surroundings. Top-down influences refer to how observers choose to look at areas that relate to observers' goals or knowledge (Castelhano, Mack, & Henderson, 2009; Tummeltshammer & Amso, 2018). For example, an observer might have a goal to attend to one particular face that belongs to a focal character in a video while inhibiting looking at other visually-salient (but less important) faces competing for attention.

Previous developmental research investigating eye movements during free viewing of images or videos has primarily asked whether bottom-up *versus* top-down influences account for where infants and children look. In these studies, bottom-up influences are measured by calculating looks to salient regions (as determined by a computational model that calculates which areas of a scene are visually conspicuous based on color, intensity, orientation, and motion), irrespective of what those regions contain. Looking at faces, irrespective of their visual saliency, is often considered evidence of top-down influences. The guiding assumption in using faces to measure top-down attention is that observers look at faces because they reflect knowledge about what is important in the scene regardless of whether they are visually salient (e.g. Birmingham, Bischof, & Kingstone, 2009). Using static image arrays that contained faces and objects, Kwon, Setoodehnia, Baek, Luck, and Oakes (2016) showed that infants 6 months and older preferentially look at faces even when more salient objects compete for attention, but 4-month-olds' attention is influenced more by visual saliency. Similarly, a saliency model better predicted 3-month-olds' eye movements while watching short animated video clips compared to a face model, whereas the face model outperformed the saliency model for 9-month-olds and adults (Frank et al., 2009). Other studies comparing eye movements during video viewing between adults and children (6-14 years) (Rider, Coutrot, Pellicano, Dakin, & Mareschal, 2018) or between adults and monkeys (*Macaca fascicularis)* (Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010) found that faces attracted attention more strongly than salient regions (despite some differences between age/species groups).

However, age differences in looking to salient areas versus faces should be interpreted cautiously. First, rates of looking to salient areas and faces are highly dependent on the particular video stimuli. When comparing across a wide range of ages (6 months to 12 years) and across a large, diverse set of video stimuli, there was no evidence for consistent age differences in either looking at salient areas or at faces across video clips (Kadooka & Franchak, 2020). Although a few clips showed age trends, age differences were not evident for most stimuli. Moreover, the influence of saliency or faces on eye movements varied widely between video clips and over time within each video clip, suggesting that observers adjust how they prioritize different features depending on the scene. Furthermore, infants' rates of face looking varies according to the number of faces in view in different stimuli (Franchak et al., 2016; Frank, Vul, & Saxe, 2012; Stoesz & Jakobson, 2014), suggesting that face looking is moderated by the content of the scene.

Second, it is problematic to treat visual saliency and face

locations as *independent* influences on visual attention. Indeed, faces may be a privileged type of visual stimulus with a distinct neural mechanism supporting their detection (M. H. Johnson, Senju, & Tomalski, 2015; Mackay, Cerf, & Koch, 2012; Morton & Johnson, 1991). Furthermore, saliency models are designed to capture the likelihood of looking at meaningful areas in a scene based on their appearance, thus, there is considerable overlap between what bottom-up and top-down models predict (Einhauser, Spain, & Perona, 2008; Henderson, Brockmole, Castelhano, & Mack, 2007). A corpus analysis of children's television programs found that faces have high contrast features and are dynamic, so they are often more salient than the static background of a scene (Wass & Smith, 2015). In Frank et al. (2009), both saliency models and face models became more predictive with age (although the face model did so at a greater rate), possibly because face regions were more salient than non-face regions. This was explicitly tested in an investigation of Feist's *Sesame Street* clip: The relative saliency at the point of gaze was greater when observers fixated faces compared with non-face regions (Franchak et al., 2016). Of course, the saliency of faces changes from moment to moment and depends on the characteristics of the surrounding scene. When there are multiple faces in a scene, faces necessarily vary in their relative saliency. Complex interactions between the saliency of faces, the number of faces, and the importance of those faces to the plot narrative may explain why there are no consistent age differences in looking to salient areas or to faces when measured across a wide stimulus set (Kadooka & Franchak, 2020) or when comparing results across different investigations that used different stimuli (e.g., Franchak et al., 2016; Frank et al., 2009, 2012).

To better understand these interactions, we can investigate how saliency might serve as a *cue* to help guide attention towards faces. Rather than testing whether observers look at faces *versus* salient areas (e.g., Franchak et al., 2016; Frank et al., 2009; Kadooka & Franchak, 2020; Rider et al., 2018), we can instead ask whether observers look more often at a face when *it is more salient* compared to when *it is less salient*. Salient faces might help scaffold infants' scene viewing through the convergence of bottom-up (e.g., visual appearance) and top-down (e.g., role in the narrative) cues. However, the two prior developmental studies of bottom-up orienting to faces only tested how face saliency contributes to face looking in static images. Amso et al. (2014) found that young infants do not use saliency as a cue to guide face looking. Observers were presented with static images that contained faces; in half of the images the face was the most salient location, in the other the face was not the most salient location. Saliency facilitated face detection in observers 12 months and older, but infants under 12 months showed no difference in rates of orienting to salient and non-salient faces. A recent study (Kelly, Duarte, Meary, Bindemann, & Pascalis, 2019) found that salient faces were detected more reliably among 3- to 12-month-old infants compared with non-salient faces, suggesting that saliency could facilitate face looking early in infancy. However, no study has tested the extent to which infants' and toddlers' orienting to faces depends on different levels of face saliency in *dynamic stimuli* such as television videos. In particular, motion is a strong cue to faces in videos that might attract attention to faces (Wass & Smith, 2015).

Is centering a cue for face looking?

The *center bias* refers to the tendency for observers to direct their eyes towards the central region of a photograph, and has been found consistently in studies of adults' scene viewing (e.g., Tatler, 2007). The center bias is evident in eye tracking studies with infants as young as 3 months viewing photographs (Mahdi et al., 2017; van Renswoude, van den Berg, et al., 2019; van Renswoude, Visser, et al., 2019). In fact, a model that simply predicts that observers will look at the center of an image performs nearly as well at accounting for infant and adult eye movements as a state-of-the-art saliency model (Mahdi et al., 2017). There are several reasons for this bias. In part, viewers look to the center of images because of a "photographer's bias" (Parkhurst & Niebur, 2003; Tseng, Carmi, Cameron, Munoz, & Itti, 2009): Photographs and videos tend to center the subject in view. However, this explanation is not complete, because observers still fixate the center of photographs in which salient visual features are clustered outside of the center (Tatler, 2007). Tatler (2007) suggested two additional explanations. First, looking to the center is an ideal starting point for subsequent visual exploration—it allows observers to capture the "gist" of a scene and minimizes the distance the eyes need to travel in any direction to a location of interest. Second, viewers may look towards the center of a display to keep the eyes centered in their orbits ("orbital reserve"), which is motorically efficient and more comfortable compared with keeping the eyes rotated at an extreme angle.

Center bias is also evident in adults' viewing of dynamic scenes (Coutrot & Guyader, 2014; Kirkorian, Anderson, & Keen, 2012; Mital, Smith, Hill, & Henderson, 2011; Rider et al., 2018; Tseng et al., 2009; Wang, Freeman, Merriam, Hasson, & Heeger, 2012). In particular, adults tend to look to the center of a screen following cuts between shots, supporting the idea that the center of the screen is an ideal place to begin exploration. Indeed, the influence of the center bias weakens as time progresses after a cut. Despite the aforementioned center bias in infants' viewing of static images, infants do not re-center their gaze following scene cuts in videos (Kirkorian et al., 2012). Children 4 years and older tend to look to the center following scene cuts, but do so less reliably compared with adults (Kirkorian et al., 2012; Rider et al., 2018). These results suggest that infants' and children's lack of experience

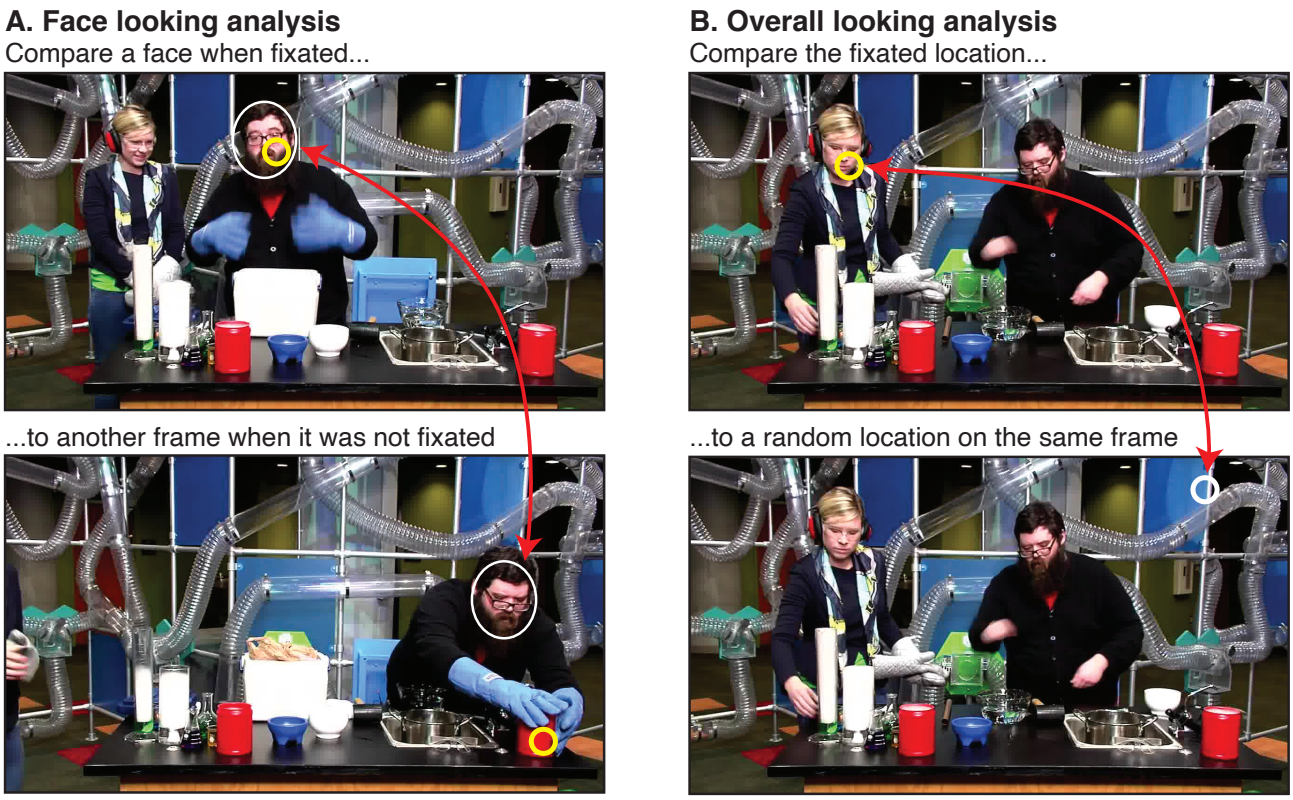

Figure 1. Two analytic strategies. A) The face looking analysis compared each face when it was looked at to the same face on other video frames when it was not looked at by each participant. For this analysis, the saliency and centering of the face area of interest (white ellipse) is compared across different video frames. B) The overall looking analysis compared the gazed location (yellow circle) on each video frame (regardless of whether it was a face or non-face region) to a randomly selected location (white circle) on the frame that was not looked at by the participant. In this analysis, the saliency, centering, and face content of the gazed and un-gazed locations (yellow vs. white circles) are compared within the same video frame.

with media may contribute to strategic differences in how centering guides visual exploration.

Adults use centering information to orient towards faces in dynamic videos (Xu et al., 2018). In particular, centering can be a useful cue to decide which face to look at when there are multiple faces in view because cinematographers often frame focal characters' faces in the center of a scene. To our knowledge, no previous study has tested whether centering influences infants' and children's tendency to look at faces—is a face fixated more during moments that it appears in the center as opposed to the edge of a screen? Given past research showing a protracted development of centering eye gaze following cuts (Kirkorian et al., 2012; Rider et al., 2018), we predict that the use of centering as a cue for face looking will increase with age.

Current Study

Much prior work in the development of viewing dynamic media has focused on how saliency, faces, and centering *independently* predict where infants and children look. The goal of the current study was to take a step further by asking how these factors *interact* by testing whether saliency and centering serve as cues to face looking, and whether those cues to face looking change with age. Although prior work examined saliency-based orienting to faces in photographs (Amso et al., 2014; Kelly et al., 2019), it is unknown whether those results generalize to dynamic videos that contain motion cues and narrative content. And although centering is a strong cue for adults' orienting to faces in videos (Xu et al., 2018), no study has tested whether centering relates to face looking in infants and children. Previous research found considerable variability in rates of looking at faces, rates of looking to salient areas, and the saliency of faces across video stimuli (Frank et al., 2012; Kadooka & Franchak, 2020; Stoesz & Jakobson, 2014; Wass & Smith, 2015). Accordingly, it was important for the current investigation to test saliency and centering cues for face looking across a large stimulus set with multiple faces that vary in appearance

and location over time.

Thus, these analyses were conducted on a previously collected data set that met those criteria (Kadooka & Franchak, 2020). Eye movement data from children (6 months to 12 years) and adults who watched five 2-minute video clips were used to determine the influence of saliency and centering as cues for face looking and whether the influence of those cues changed with age. We annotated the data set to define face areas of interest (AOIs) on every frame and calculated the saliency and centering of each face. We employed generalized mixed-effect models (GLMMs) to predict the likelihood of looking at each face based on saliency and centering by comparing frames in which the same face was or was not looked at by each participant (Figure 1A). This "face looking analysis" lets us take into account the nested random effects in complex stimuli—multiple faces nested within video frames nested within individual video exemplars—to determine how saliency, centering, and their interaction predicts face looking across age. Although past work with static images suggests an age-related increase in saliency-based orienting to faces, given inconsistent age differences in looking to salient areas and faces (when measured separately) across studies using dynamic stimuli (Franchak et al., 2016; Frank et al., 2009; Kadooka & Franchak, 2020; Rider et al., 2018), we had no strong basis to predict whether saliency-based orienting to faces would change with age. However, prior work is consistent in showing age differences in centering following scene cuts (Kirkorian et al., 2012; Rider et al., 2018), so we predicted that the influence of centering on face looking would increase with age.

If we find age-related changes in how face looking depends on saliency and/or centering, it is important to determine whether those changes in attention are specific to face looking or reflect general attention development. Although the past study using the data set (Kadooka & Franchak, 2020) found that there were no general age-related changes in how frequently infants, children, and adults looked at salient areas or at faces, the role of centering was not tested. Thus, we conducted a second "overall looking analysis" to assess whether there were age-related changes in looking at centered regions, salient regions, and faces. Following the approach of Rehrig et al. (2022), we used GLMMs to compare features at the region each participant looked at on each video frame compared with a randomly-selected un-gazed location on the same frame (Figure 1B). This analysis will distinguish whether the likelihood of looking at an area in the video versus a random area depend on saliency, centering, and the presence of a face, and whether that likelihood changes with age.

Method

Participants

Participants were drawn from a previous study that investigated eye movements during free viewing of video stimuli (Kadooka & Franchak, 2020) in children (6 months to 12 years) and college-aged adults. In the previous study, seven 2-min video clips were shown to participants. Data from five video clips were used in the current study (one video clip was excluded because it did not contain faces for a long enough duration to analyze, and another video was excluded because there was too much overlap in face regions in the image to disambiguate which face was looked at on each frame). The sample contained 79 children and 10 adults based on a power analysis and exclusion criteria reported in the original paper. Data from 10 additional adults were later collected to serve as a baseline group for calculating adult-like gaze scores, as in (Franchak et al., 2016), but were not available when the original paper was published. We made use of all available data in the current study, so the final sample included 79 children and 20 adults. Table 1 provides age, sex, number of participants, and calibration quality statistics for infants (0.5-1.5 years), toddlers (1.5-3 years), young children (3-6 years), older children (6-12 years), and adults (18-26 years).

Adult participants were recruited from the University of California, Riverside psychology participant pool and child participants were recruited from the Riverside County area. Adults participants received credit towards a course requirement, and families of child participants received $10 and a book or small toy. Caregivers identified participating children as Black/African American ($n = 1$), American Indian/Alaskan Native ($n = 4$), non-Hispanic White ($n = 38$), Hispanic or Latino(a)/White ($n = 12$), and more than one race ($n = 21$). Adult participants self-identified as Black/African American ($n = 2$), non-Hispanic White ($n = 5$), Asian ($n = 8$), Hispanic or Latino(a)/White ($n = 3$), and more than one race ($n = 1$). Race/ethnicity was not reported for 3 children and 1 adult. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. The study was approved by the Institutional Review Board of the University of California, Riverside and conforms to the standards of the US Federal Policy for the Protection of Human Subjects.

Stimuli

The stimuli came from different sources. Two clips were from *Sesame Street* (including Feist's counting video), two were music videos designed for an adult audience (but still appropriate for child viewers), and one clip was a children's science demonstration. Clips were selected to be continuous shots containing no cuts and to present live-action

Table 1
Characteristics of participants in each age group: Number of participants, sex of participants (M = male, F = female), age (months), mean calibration error (degrees), mean precision (degrees), and percentage of video frames containing valid data. Standard deviations are shown in parentheses.

	n	M	F	Age	Error	Precision	% Valid Gaze
Infants (0.5-1.5 yr)	33	13	20	0.97	0.64 (0.29)	1.71 (0.38)	77.9 (17.3)
Toddlers (1.5-3 yr)	20	11	9	2.00	0.54 (0.15)	1.69 (0.30)	82.5 (15.0)
Young Children (3-6 yr)	14	8	6	4.52	0.57 (0.33)	1.70 (0.33)	85.2 (13.7)
Older Children (6-11 yr)	12	9	3	9.01	0.33 (0.10)	1.59 (0.22)	88.5 (9.0)
Adults (18-26 yr)	20	13	7	19.83	0.45 (0.16)	1.53 (0.31)	91.8 (6.8)

scenes with limited graphical elements. The audio track from each video was replaced with instrumental children's music so that eye movements would be based solely on visual information in each scene. A Databrary repository (https://nyu.databrary.org/volume/1007) contains each of the stimulus videos overlaid with exemplar data from infants and adults (for more stimulus details, see Kadooka & Franchak, 2020).

Apparatus and procedure

Participants sat 60 cm from an Eyelink 1000 Plus eye tracker (SR Research) mounted beneath a 43.2 cm (diagonal) computer monitor. Monocular eye movements (right eye) were recorded at 500 Hz. Infants sat in a high chair with a harness to reduce body movement and older participants were seated on a chair. A sticker was placed on each participant's forehead to allow the eye tracker to detect participant faces because a chinrest was not used. A 5-point calibration was performed for each participant followed by a 5-point validation used to estimate calibration accuracy. The calibration process was repeated until validation indicated errors 1.5º or less. After a satisfactory calibration, videos were shown in a randomized order. Videos were presented at 30 Hz at the maximal size allowed by the monitor, subtending a visual angle of 31º (horizontal) × 19º (vertical).

Calibration accuracy (spatial error of estimated point of gaze compared to actual validation target) averaged M = 0.53º (SD = 0.25) across age groups and was correlated with age (r = -.27, p = .006). However, the size of the age difference in accuracy was minimal, with infants' accuracy only 0.22º worse than adults'. Spatial position was reliable over successive samples throughout testing, with precision averaging M = 1.65º (SD = 0.33) using the metric described by Wass, Forssman, and Leppänen (2014). Precision was negatively correlated with age (r = -0.20, p = .048), but differed by only 0.18º between infants and adults. Calibration accuracy, precision, and the percentage of valid gaze samples are summarized by age group in Table 1. Although the small differences in accuracy and precision are unlikely to have an influence in the resulting analyses, to be cautious we refrained from using fixation-detection algorithms, which can be susceptible to differences in calibration quality (Wass et al., 2014). Instead, we analyzed raw data samples to avoid differences in fixation detection creating an age-related confound.

Data processing

Dataviewer software (SR Research) was used to draw an area of interest (AOI) ellipse around each face exemplar on every frame (white ellipses in Figure 2). There were 21 different face exemplars coded across the 5 videos. Faces varied in size and location from frame to frame due to character movement. For each participant, we calculated face looking for each face exemplar on each video frame when the point of gaze fell within the face AOI. Frames for which a participant had valid gaze data (i.e., looking on screen) but did not look at a particular face exemplar were considered non-looking events. Because the face-looking analysis modeled each face separately, we did not need to address overlapping face AOIs. If the point of gaze was located within two overlapping face AOIs, that frame counted as a face-looking event for both of the faces. The following sections describe how we calculated each cue—saliency and centering—to determine whether each cue could be used to accurately distinguish between face-looking events and non-looking events.

Saliency of face AOIs. A saliency map for every pixel in each video frame was calculated using the GBVS toolbox (Harel, Koch, & Perona, 2006) implementation of the Itti and Koch saliency algorithm (Itti et al., 1998). Each video frame was converted to an image, from which five separate low-level feature maps (intensity, color, orientation, flicker, and motion) were generated. Flicker and motion maps used information from successive video frames. Each feature map represented the degree to which each pixel was unique in its feature value relative to the other pixels in the image. The five feature maps were evenly weighted and combined into an overall saliency map that contained the relative saliency of every pixel in the image. Based on the overall map, each pixel was given a percentile score ranging from 0 to 1, with 1 being most salient pixel in the image.

We determined the saliency of each face AOI on every video frame as the value of the most salient pixel contained

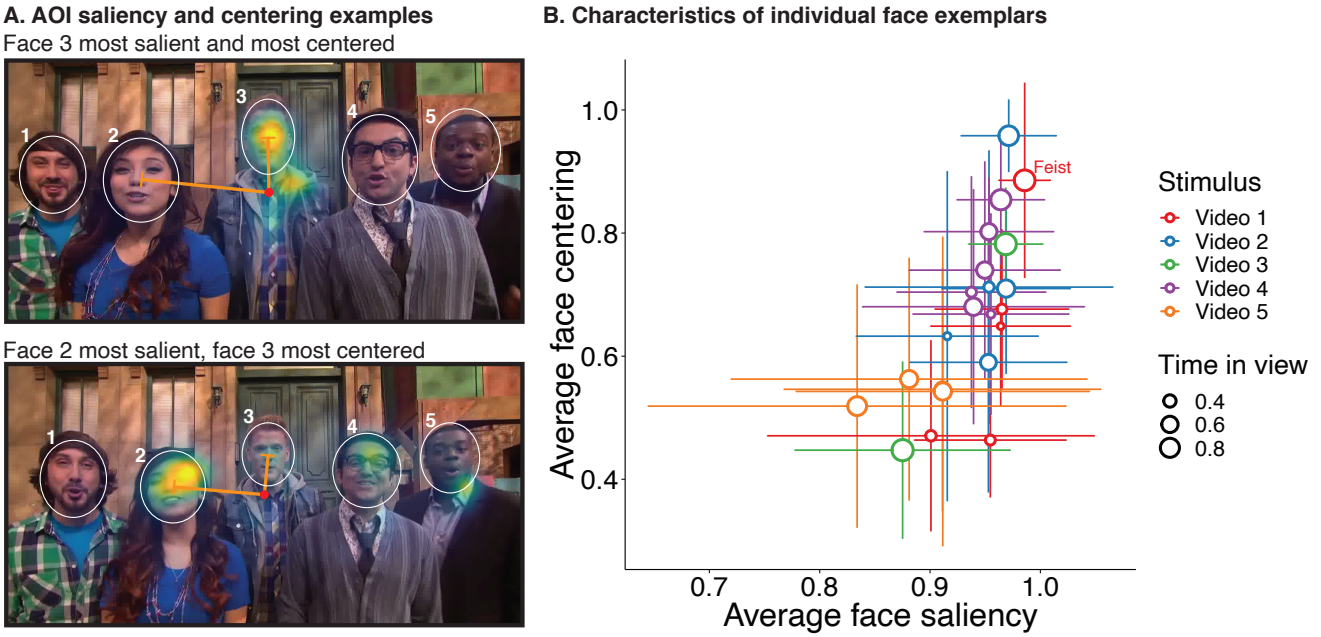

Figure 2. (A) Example of face areas of interest (AOIs) on two example video frames. White ellipses were drawn by hand to indicate face locations on each frame. The saliency map is overlaid onto each frame to indicate which regions had higher saliency (bright yellow) compared to little saliency (no colored overlay). The centering of each face is indicated by the orange line measuring the distance of each AOI to the center of the image (red dot). (B) Characteristics of the 21 face exemplars. Each symbol indicates one face AOI by plotting its average saliency value against its average centering value. Vertical error bars indicate the SD of saliency and horizontal error bars indicate the SD of centering (SDs calculated across all frames in which the face appeared in the video). The color of the point refers to which of the five stimulus videos the face appeared in. The size of the point indicates the proportion of time that the face appeared in the 2-min video. Feist's face is labelled, showing that it was both more salient and more centered on average than most faces in the data set.

in the AOI ellipse. Figure 2A displays the overall saliency map overlaid onto two example video frames. In the top image of Figure 2A, face 3 is the most salient, but face 2 is the most salient in the bottom image. An AOI value of 1 indicated that the AOI contained the most salient pixel in the image; an AOI value of .75 indicated that the most salient pixel in the AOI was more salient than 75% of the pixels in the image. Note, we chose to use the most salient pixel as a metric for saliency as opposed to the mean of the saliency values to capture the degree to which a face "pops out" from the scene. A face that uniformly contains moderately-salient pixels may have a higher mean value than a face that contains the most salient pixel in the image. However, the face with the most salient pixel is more likely to stand out in the image due to the extreme value. This choice in operationalizing saliency also makes our results more comparable to the only other prior studies of bottom-up orienting to faces in complex images (Amso et al., 2014; Kelly et al., 2019), which defined salient faces as those that contained the most salient pixel. The resulting saliency values were z-scored for use in statistical models.

Centering of face AOIs. The centering of each face AOI was calculated for every video frame based on the Euclidean distance of the AOI to the center of the image (based on the distance from the center of the screen to the center pixel of the AOI). The yellow lines on Figure 2A show the distance from center for two faces (2 and 3); in both cases, face 3 is the most centered in the image. Raw centering distances in pixels were z-scored and reversed, so that positive values indicate AOIs that are closer to the center of the image.

Results

As in recent work (Nuthmann, Einhäuser, & Schütz, 2017; Rehrig et al., 2022; van Renswoude, Visser, et al., 2019), we used generalized linear mixed-effect models (GLMMs) with binomial link functions (i.e., logistic regression) to model how the likelihood of looking depended on the visual features of gazed and un-gazed locations. Analyses were conducted in *R* (R Core Team, 2018) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2014). The first analysis ("face looking") tested whether each face was more or less likely

to be looked at by participants of different ages depending on the saliency and centering of that face (Figure 1A). This analysis addresses the primary aim of determining whether face saliency, face centering, and their interaction result in a higher likelihood of looking at a given face. The second analysis ("overall looking") determined whether there were baseline differences in looking at salient, face, and centered locations across age. This secondary analysis was used to determine whether the results from the face looking analysis derived from general age-related changes in attention to these cues, or whether they were specific to face looking.

Age moderated the influence of cues on face looking

First, the "face looking analysis" compared the saliency and centering of each face on frames where it was looked at by a participant compared with the *same face on frames where it was not looked at by that participant*. We gathered each face AOI on each frame that was looked at by each participant (face looking = 1). To model what factors increased the likelihood that each participant looked at that face, we created a comparison set of faces by randomly sampling an equal number of video frames that each face was not looked at by each participant (face looking = 0). For example, if a participant looked at a particular character's face on 800 video frames, we randomly selected 800 frames during which that participant did not look at that character's face and instead looked elsewhere to compare the saliency and centering of that face on looking versus non-looking frames. The final data set for this analysis contained 1,128,513 rows of data, determined by the variation in how many frames each of the 99 participants looked at each face, how many faces were contained in each video (2-5 faces), and how many video frames each face was present within each video. As in previous work (Amso et al., 2014; Kadooka & Franchak, 2020), log-transformed age was used to model changes that are initially rapid in infancy but slow in later childhood and adulthood. Additionally, better fits based on AIC values were obtained when using log-transformed age compared with non-transformed age. The fixed effects (centering, saliency, and log-transformed age) were z-scored to aid in GLMM calculations. Age was centered at the minimum value so that parameter estimates would indicate performance for the youngest infants tested in the study.

A benefit of mixed-effect modeling is that it allows specification of multiple crossed and nested random effects. In this design, individual faces were nested within video frames, and individual video frames were nested within the five videos. Each participant was a random effect that was crossed with the nested face-within-frame-within-video effect, since each participant viewed each of the five videos. The most maximal model converged, which included random intercepts and random slopes for centering, saliency, and their interaction. A second benefit of modeling faces and frames as random effects is that it helps address the differential missingness in gaze data observed across age groups; the random effect structure captures baseline differences in the degree to which different faces vary in the influence of saliency and centering on different video frames. The final model tested was: $Face\ Looking \sim Centering \times Saliency \times log(Age) + (1 + Centering \times Saliency|Subject) + (1 + Centering \times Saliency|Video/Frame/Face)$. Table 2 shows the fixed and random effects of the GLMM.

Because age was centered on the youngest participant (6 months), the significant effect of saliency (*beta* = 0.18) and non-significant effect of centering (*beta* = -0.04) suggest that saliency but not centering served as a cue that distinguished between times that infants would look at versus not look at a given face. However, the role of saliency and centering changed with age, as evidenced by a significant Centering×Saliency×Age 3-way interaction (*beta* = 0.03) and a significant 2-way Centering×Age (*beta* = 0.06) interaction. To further illustrate this 3-way interaction, we plotted model estimates and confidence intervals of the likelihood of looking at a face given that face's saliency and centering separately for each age group (each individual plot in Figure 3 shows marginal estimates set at the mean age of each age group listed in Table 1). As the figure shows, younger infants looked more at faces when they were more salient (the solid line is above the dashed line), however, the likelihood of face looking did not change according to the centering value (lines are flat with respect to the x-axis). By adulthood, both face saliency and face centering increased the likelihood of face looking, and the two cues showed an interactive effect. Increasing saliency had a minimal effect on increasing the likelihood of looking at a less-centered face, but had a larger effect on increasing the likelihood of looking at a more-centered face.

To investigate this another way, we re-centered age on adults and then recalculated the model. Like infants, adults showed a significant effect of saliency (*beta* = 0.17, $p = .004$). Unlike infants, adults showed a significant effect of centering (*beta* = 0.15, $p = .028$) and a significant Saliency×Centering interaction (*beta* = 0.07, $p = .014$). In summary, infants' face looking patterns suggest that they attended to faces when they were more salient, but not when they were more centered. In contrast, with age, both saliency and centering increased the likelihood of looking at a face, in particular when the two cues were present in combination (a face that was both centered and salient).

The random effect structure of the model provides a way to investigate what aspects of the stimuli created variations in face looking. By-subject random effects varied less compared to by-item random effects (at every level of nesting), suggesting that faces varied substantially across video frames in ways affected how each cue might influence face looking. Furthermore, random slopes of centering

Table 2

Generalized linear mixed-effect model predicting the likelihood of face looking from face centering, face saliency, and participant age. Centering and saliency were scaled and centered at the mean; age was log-transformed, scaled, and centered at the minimum (6 months). Fixed-effect parameter estimates are log(odds); random effects indicate the standard deviation (SD) of parameters according to each random effect grouping. Statistically-significant model terms are indicated in bold.

	Fixed Effects				Random Effects (*SD*)			
Predictor	*beta*	*SE*	*z*	*p*	Subject	Face[1]	Frame[2]	Video
Intercept	-0.20	0.053	-3.72	**<0.001**	0.06	1.00	0.18	0.11
Centering	-0.04	0.069	-0.56	0.57	0.11	0.29	0.27	0.15
Saliency	0.18	0.061	3.05	**0.002**	0.09	0.11	0.25	0.13
log(Age)	-0.01	0.007	-0.90	0.37				
Centering×Saliency	-0.01	0.030	-0.38	0.70	0.06	0.14	0.10	0.06
Centering×log(Age)	0.06	0.011	5.40	**<0.001**				
Saliency×log(Age)	0.00	0.010	-0.32	0.75				
Centering×Saliency×log(Age)	0.03	0.007	4.00	**<0.001**				

[1]Face nested in Frame nested in Video
[2]Frame nested in Video

according to Face/Frame/Video were more variable (*SD* = 0.29) compared with random slopes of saliency according to Face/Frame/Video (*SD* = 0.11). This suggests that saliency played a more constant role over time within a video for each given face, whereas centering was a more variable cue over time. This is consistent with the greater underlying variability in face centering compared with face saliency seen in Figure 2.

Did overall looking depend on face presence, saliency, and centering?

The prior section indicates that face looking depends on saliency across age, but increasingly depends on both saliency and centering through childhood and adulthood. This raises the question of whether the age-related increase in centering as a cue for face looking is specific to face looking, or whether it simply reflects an age-related increase in the center bias. To address this, we conducted an "overall looking analysis" to determine whether gazed versus un-gazed locations on every video frame differed in saliency, centering, and/or the presence of a face. Whereas the previous analysis selected data based on face looking events and calculated the saliency and centering of an entire face AOI (Figure 1A), the current analysis analyzed every video frame, regardless of face looking, to compare image features within that video frame (Figure 1B).

We gathered every video frame where the participant had valid gaze data (i.e., excluding blinks or looks offscreen). For each frame for each participant, we calculated the saliency and centering within a circular region 1.2º in radius around the point of gaze (looking = 1). As in the face looking analysis, the saliency value used was the maximum of saliency values within the gazed region, and centering was the scaled distance between the center of the gazed region and the center of the video. Using the AOI information, we coded whether any part of the gazed region intersected with a face AOI or not. Next, we randomly selected a second circular region of the same size from within the video frame (looking = 0) that was not permitted to overlap the gazed region, as in Rehrig et al. (2022). We then calculated saliency, centering, and face presence for the un-gazed location. Note that truly random comparison locations (as opposed to gaze-inspired random locations) were necessary to reveal whether participants show a centering effect. The resulting data set contained 2,802,128 rows of data based on how many frames/videos contained valid gaze data for each of the 99 participants.

As before, we calculated a GLMM to model the likelihood of looking at a region of the video depending on fixed effects of log(age), saliency, and centering, but with the addition of face (face versus non-face region) as an additional categorical fixed factor. We included random effects for individual video frames nested within the five videos, but individual faces were not considered in this model, just the overall tendency to look at any face region. Thus, each participant was a random effect that was crossed with the nested frame-within-video effect. Because our goal was simply to test whether each of the three cues predicted overall looking and whether the strength of those cues changed with age, we centered age on the youngest participant and included interaction terms for each cue against age (more complex interaction terms were not necessary to answer our question). The final model tested was: $Looking \sim Centering \times log(Age) + Saliency \times log(Age) + Face \times log(Age) + (1 + Centering + Saliency + Face|Subject) + (1 + Centering + Saliency + Face|Video/Frame)$. Table 3 shows the fixed and random effects of the final GLMM model.

The results of the GLMM rule out the possibility that the

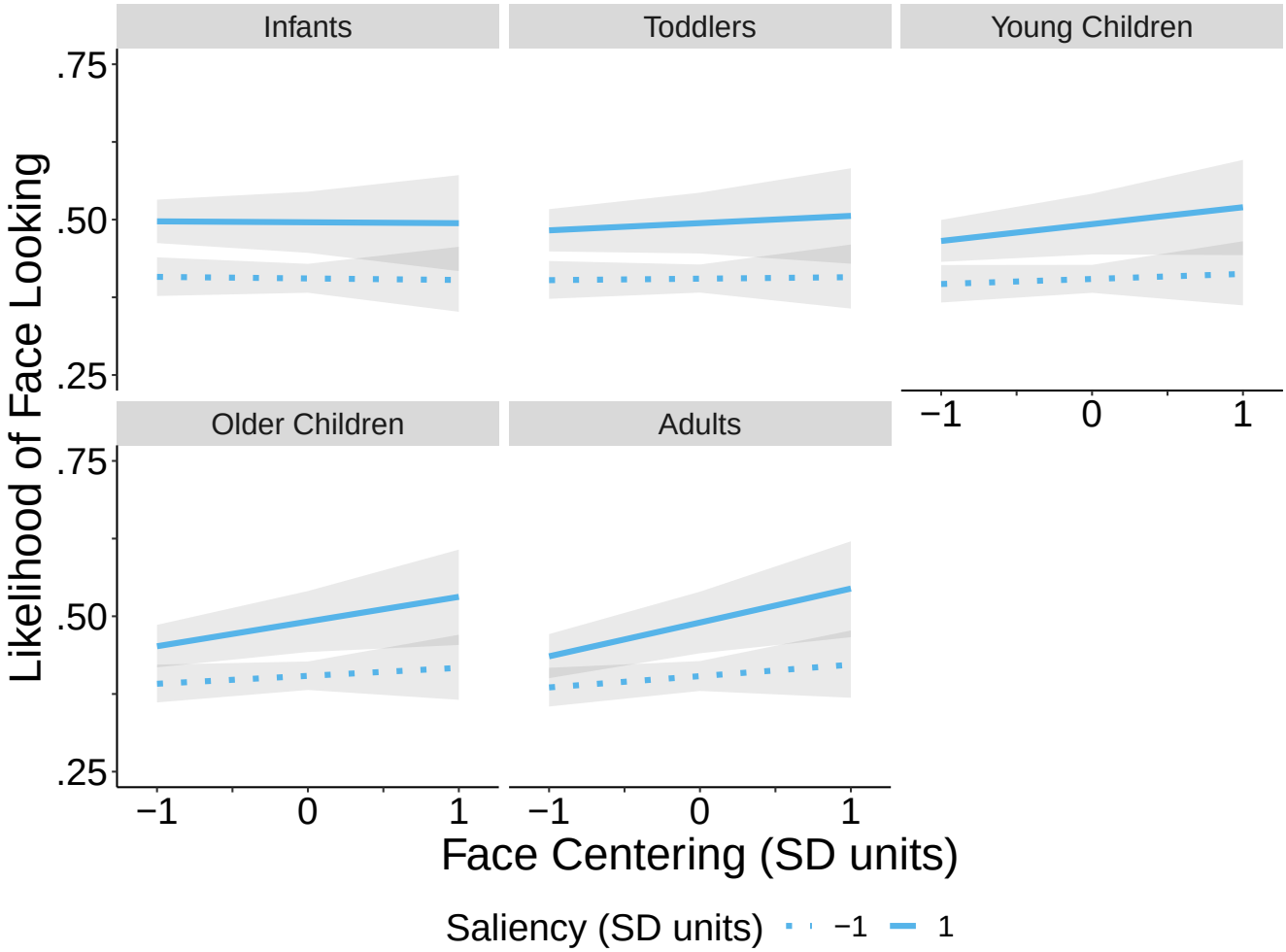

Figure 3. Estimated likelihood of face looking (y-axis) according to z-scored face centering (x-axis; larger values = more centered) and face saliency (solid lines = more salient). Shaded regions show 95% confidence intervals (darker gray regions indicate slight overlaps between confidence intervals). Plots are shown individually according to age group to indicate how the interaction between centering and saliency emerged from infancy to adulthood.

face-looking effects in the previous section are due to general age-related changes in attention to each type of cue. All three cues showed significant positive effects in the model (face *beta* = 1.58, centering *beta* = 0.86, and saliency *beta* = 0.77). Because age was centered at the youngest participant, we can conclude that centering predicted overall behavior in the youngest infants even though it did not significantly influence face looking in young infants. More importantly, none of the three cues interacted with age (*p*s > .073). Thus, we failed to find evidence that saliency, centering, and face presence changed in their overall influence on looking behavior from infancy to adulthood. Figure 4 shows the likelihood of looking at a region according to each cue; flat lines with respect to participants' age (x-axis) suggest that the influence of each cue was consistent at every age. As in the previous analysis, random slopes for each type of cue were more variable over frames compared to between videos (and between participants), suggesting that the video stimuli contained substantial variation in how each type of cue influenced looking over time.

Discussion

Complex scenes—whether static photographs or dynamic videos—contain rich, structured information that can guide observers' attention. The faces of important characters are often salient (Wass & Smith, 2015) and centered in view (Xu et al., 2018). Our new analysis of a previously-collected data set (Kadooka & Franchak, 2020) addresses a gap in the developmental literature concerning whether infants, children, and adults are differentially sensitive to these cues. Our find-

Table 3

Generalized linear mixed-effect model predicting overall looking from centering, saliency, face presence, and participant age. Centering and saliency were scaled and centered at the mean; age was log-transformed, scaled, and centered at the minimum (6 months). Fixed-effect parameter estimates are log(odds); random effects indicate the standard deviation (SD) of parameters according to each random effect grouping. Statistically significant model terms are shown in bold.

	Fixed Effects				Random Effects (*SD*)		
Predictor	*B*	*SE*	*z*	*p*	Subject	Frame[1]	Video
Intercept	-0.56	0.206	-2.70	**0.007**	0.20	0.48	0.45
Centering	0.86	0.103	8.39	**<0.001**	0.14	0.61	0.22
Saliency	0.77	0.051	15.13	**<0.001**	0.10	0.56	0.11
Face	1.58	0.199	7.93	**<0.001**	0.34	0.99	0.42
log(Age)	-0.02	0.020	-1.18	0.24			
Centering×log(Age)	0.02	0.015	1.40	0.16			
Saliency×log(Age)	-0.02	0.010	-1.79	0.073			
Face×log(Age)	0.01	0.034	0.22	0.83			

[1]Frame nested in Video

ings revealed that whereas face saliency cued face looking in participants of all ages, face centering emerged as a cue later in development. Furthermore, the overall looking analysis revealed that this age-related change in centering as a cue for face looking cannot be attributed to a general age difference in the center bias: Participants of all ages were similarly sensitive to face presence, saliency, and centering.

Demonstrating that saliency is a positive cue for face looking in both infants and children (6 months to 12 years) adds to previous work that tested static images (Amso et al., 2014; Kelly et al., 2019). Whereas Amso et al. (2014) found that young infants (4-12 months) did not orient to faces based on saliency, Kelly et al. (2019) found that face saliency aided face detection. In the current study, we found that saliency was a significant and equally-strong cue for face looking across age when viewing dynamic scenes. In addition to static features (i.e., intensity, color, orientation), videos contain dynamic features (i.e., flicker, motion). Motion may be a particularly informative cue for face detection because infants see agents (and their faces) move in daily life. An advantage of the current approach compared to previous studies of saliency-based orienting to faces was that we modeled whether *continuous* increases in saliency predicted face looking using GLMMs, whereas Amso et al. (2014) and Kelly et al. (2019) made *binary* comparisons between faces that either contained the most salient region or did not. The saliency of 'non-salient' faces relative to other targets in the photographs was not reported in those investigations. As our analysis showed (Figure 2B), faces were among the most salient regions in most frames of the videos.

Our analysis was novel in testing whether face centering, and the interaction of face centering and face saliency, were predictive of face looking in infant and child samples. The youngest infants we tested did not show evidence of using face centering as a cue for face looking. With age, the role of centering increased; by adulthood, centering was an equally strong predictor compared with saliency. Moreover, when we centered the age variable on adults, a significant Saliency×Centering interaction revealed that adults used the two cues in combination (Figure 3). For salient faces, centering was a stronger predictor of face looking, but centering had a more modest effect among less salient faces. Because face saliency and face centering were strongly correlated across face exemplars (Figure 2B), at first glance it would seem that the two cues might be redundant. Yet, adults behaved as if the combination of saliency and centering was a particularly strong cue for face looking, suggesting that centering and saliency might carry different information. This is consistent with the random effect component of the model; faces varied more in centering compared with saliency across videos frames.

The different age trajectories of orienting based on saliency versus centering shed light on the developmental mechanisms that underlie changes in face looking during media viewing. Bottom-up attention develops rapidly in the first months of life (Colombo, 2001; Oakes & Amso, 2018). Improvements in the perception of low-level features, such as motion, color, and orientation, depend on the maturation of visual pathways (S. P. Johnson, 2011). This maturation is not independent of experience. For example, deprivation studies show that a complete lack of visual input to one eye disrupts visual development in kittens (Hubel & Wiesel, 2004). More subtle effects of experience have been found in studies with human infants: Amso et al. (2014) found that bottom-up orienting to faces relates to environmental factors such as family size and socioeconomic status. Regardless, because saliency-based face orienting was stable from 6 months through childhood and early adulthood, protracted improvements in attention through early childhood (Oakes & Amso, 2018) do not lead to an appreciable difference in the role of saliency as a

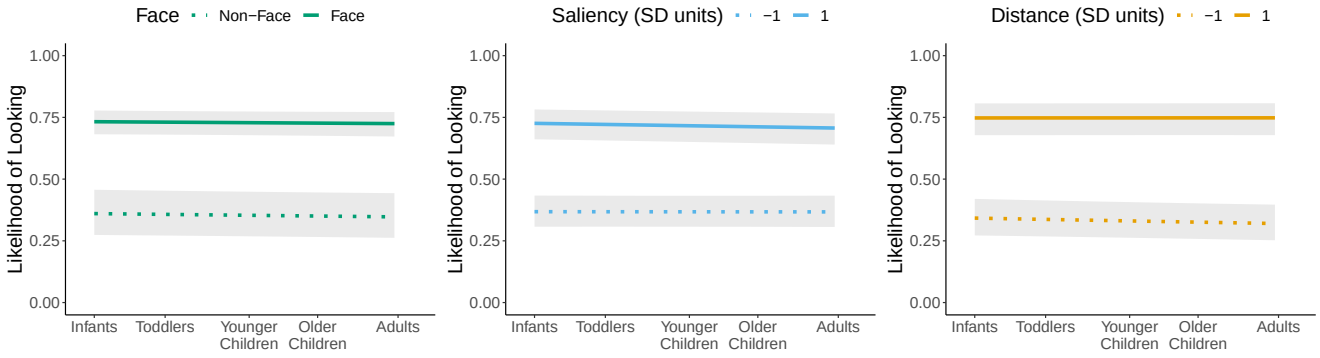

Figure 4. Estimated overall likelihood of looking (y-axis) by age (x-axis) according to face presence, saliency, and centering. Shaded regions show 95% confidence intervals. Each plot shows that participants of all ages looked more frequently at faces compared to non-faces, salient areas compared to non-salient areas, and centered areas compared with non-centered areas. Moreover, the flat lines indicate that these overall trends were unrelated to participant age.

cue for face looking.

More work is needed to understand what drives the emergence of centering as a cue for face looking. It is unlikely infants could learn that centering cues faces from real-life social attention experiences (outside of screens), because faces are only centered in view when infants choose to orient their heads to bring faces in view. Prior studies examining real-life social attention demonstrate faces become less frequent in view with age (Fausey, Jayaraman, & Smith, 2016), possibly as a consequence of: 1) infants spending time on the ground in body positions that are not conducive to orienting towards adult faces (Franchak et al., 2018; Kretch, Franchak, & Adolph, 2014), and 2) infants increasingly centering toys rather than faces in view (Franchak, 2020). Possibly, the development of centering as a cue for faces is related to media knowledge, similar to the age-related changes seen in re-centering gaze following scene cuts (Kirkorian et al., 2012; Rider et al., 2018). Moreover, given that our "overall looking" analysis failed to find a change in center bias with age, the current study suggests that the early presence of a general center bias in infants (Mahdi et al., 2017; van Renswoude, van den Berg, et al., 2019; van Renswoude, Visser, et al., 2019) does not carry over to centering serving as a significant cue for face looking. Just like adults, infants in the study showed a robust center bias, and both saliency and centering were strong predictors of overall looking. The latter result contrasts with previous research of the center bias in static images (van Renswoude, van den Berg, et al., 2019), in which the center bias was stronger in adults compared with infants.

The current study highlights the methodological importance of testing free-viewing behavior across a larger, more varied stimulus set than is typical in the literature. Previously-reported age differences in attention to faces that used only a single video stimulus (Franchak et al., 2016; Frank et al., 2009) do not generalize to a large, diverse set of videos (Kadooka & Franchak, 2020). In the current study, variations within the stimulus set also played an important role. Had we analyzed only a single face in a single video to study (such as Feist's face in the *Sesame Street* clip), we would have missed the larger range of face saliency and centering values that occurred across the varied set of face exemplars (Figure 2B) that made it possible to detect the effects of each cue. As in Nuthmann et al. (2017), we found that between-item variation in random effects surpassed between-participant variation in random effects. Beyond variation in the faces' features (between faces and over time), there was likely considerable un-measured variation in how those features corresponded with "importance" within the scene. Future work could address the latter more directly by systematically varying whether salient/centered faces are more or less relevant in the scene's narrative to assess how observers prioritize different cues for face looking.

We acknowledge some limitations in our approach. First, visual attention to faces on screens is unlike visual attention to faces in everyday life when observers are free to move and interact. Regardless, we believe the question of what governs looking to faces in screens is important in its own right, given the applications for understanding educational media and diagnostic testing that uses faces on screens. Furthermore, the use of replicable stimuli—videos that can be shown to multiple participants—is a methodological necessity for such a question. In a naturalistic setting, it would be impossible to equate the frequency with which participants saw the same faces at the same levels of saliency and centering. Second, as with any study that uses "found" stimuli, such as videos, we acknowledge that the specific characteristics of the videos could not be perfectly controlled nor manipulated. We hope

that by detailing the characteristics of the face exemplars in the video stimuli, these results can provide a basis for comparison to any future work that uses different stimuli with different face exemplars. Third, as is common in developmental studies that use eye tracking, the data set contained differences in eye tracking accuracy and amount of useable data between participants of different ages. Although the differences in accuracy were modest, we cannot completely rule out their effect on the current analyses. However, given that we only found age differences in face centering with respect to face looking—no other analysis revealed effects of age—it seems less likely that the developmental finding is confounded by eye tracking quality or differences in valid gaze data.

Conclusion

A primary strength of the current study was moving beyond treating saliency and faces as separable influences on the development of looking behavior. Face looking during video viewing is not a pure measure of top-down attention or of social attention because infants' and children's decisions to look at faces vary according to their visual features with respect to the scene—face saliency and centering. Developmental differences in face looking likely depend on a variety of factors: attention development, experience with media, and age differences in plot comprehension. Crucially, the particular correspondence between the *visual characteristics of faces* and *importance of faces* in a scene—which is often unknown or unreported in prior research—may tap into different underlying developmental influences on face looking behavior. This can have significant consequences on the interpretation of face looking and its application. For researchers who seek to use face looking as a diagnostic tool for early identification of ASD, care should be taken to describe the visual features of face stimuli (saliency and centering). Future research should endeavor to understand how differences in the correspondence between visual features, faces, and narrative content relate to free viewing behavior over development.

References

Abelman, R. (1990). You can't get there from here: Children's understanding of time-leaps on television. *Journal of Broadcasting & Electronic Media*, *34*, 469–476.

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

Barr, R. (2013). Memory constraints on infant learning from picture books, television, and touchscreens. *Child Development Perspectives*, *7*, 205–210.

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

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

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

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

Einhauser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*, 1-26.

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

Franchak, J. M. (2020). Looking with the eyes and head. In J. B. Wagman & J. J. C. Blau (Eds.), *Perception as Information Detection: Reflections on Gibson's Ecological Approach to Visual Perception* (pp. 205–221). Routledge.

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., 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., 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

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, *42*, 279–297.

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

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

Hubel, D. H., & Wiesel, T. N. (2004). *Brain and visual perception: The story of a 25-year collaboration*. Oxford University Press.

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

Johnson, M. H., Senju, A., & Tomalski, P. (2015, March). The two-process theory of face processing: modifications based on two decades of data from infants and adults. *Neuroscience and biobehavioral reviews*, *50*, 169–179.

Johnson, S. P. (2011). Development of visual perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 515–528.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431.

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), 2065-2079.

Kelly, D. J., Duarte, S., Meary, D., Bindemann, M., & Pascalis, O. (2019). Infants rapidly detect human faces in complex naturalistic visual scenes. *Developmental Science*, *22*(6), e12829.

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*, 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

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

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

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

Mackay, M., Cerf, M., & Koch, C. (2012, April). Evidence for two distinct mechanisms directing gaze in natural scenes. *Journal of vision*, *12*(4), 9.

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

Morton, J., & Johnson, M. H. (1991, April). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological review*, *98*(2), 164–181.

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017, October). How well can saliency models predict fixation selection in scenes beyond central bias? a new approach to model evaluation using generalized linear mixed models. *Frontiers in human neuroscience*, *11*, 491.

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

Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B., & Lagopoulos, J. (2014). A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience*, *9*, 610–632.

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

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

R Core Team. (2018). *R: A language and environment for statistical computing; 2015.*

Rehrig, G., Barker, M., Peacock, C. E., Hayes, T. R., Henderson, J. M., & Ferreira, F. (2022). Look at what I can do: Object affordances guide visual attention while speakers describe potential actions. *Attention, Perception & Psychophysics*, doi: 10.3758/s13414-022-02467-6.

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

Ryzik, M. (2019, August 25). How Feist's '1234' turned into a 'Sesame Street' blockbuster. *New York Times*, AR13.

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

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. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*, 1-17. doi: 10.1167/7.14.4

Troseth, G. L. (2010). Is it life or is it Memorex? Video as a representation of reality. *Developmental Review*, *30*, 155–175.

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

Tummeltshammer, K., & Amso, D. (2018). Top-down contextual knowledge guides visual attention in infancy. *Developmental Science*, *21*, e12599.

van Renswoude, D. R., van den Berg, L., Raijmakers, M. E., & Visser, I. (2019). Infants' center bias in free viewing of real-world scenes. *Vision Research*, *154*, 44–53.

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*, 693–717.

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

Xu, M., Liu, Y., Hu, R., & He, F. (2018). Find who to look at: Turning from action to saliency. *IEEE Transactions on Image Processing*, *27*, 4529–4544.

Yurkovic, J. R., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., … Yu, C. (2021). Using head-mounted eye tracking to examine visual and manual exploration during naturalistic toy play in children with and without autism spectrum disorder. *Scientific Reports*, *11*(1), 3578. doi: 10.1038/s41598-021-81102-0

Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioural Brain Research*, *251*, 133–146.