Journal of Experimental Child Psychology 226 (2023) 105562

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

The effect of narrative coherence and visual salience on children's and adults' gaze while watching video

Mengguo Jing [a,*], Kellan Kadooka [b], John Franchak [b], Heather L. Kirkorian [a]

[a] *Department of Human Development and Family Studies, University of Wisconsin–Madison, Madison, WI 53705, USA*
[b] *Department of Psychology, University of California, Riverside, Riverside, CA 92521, USA*

A R T I C L E I N F O

Article history:
Received 6 July 2021
Revised 12 September 2022

Keywords:
Eye movements
Visual attention
Visual salience
Video viewing
Gaze prediction
Attention development

A B S T R A C T

Low-level visual features (e.g., motion, contrast) predict eye gaze during video viewing. The current study investigated the effect of narrative coherence on the extent to which low-level visual salience predicts eye gaze. Eye movements were recorded as 4-year-olds (n = 20) and adults (n = 20) watched a cohesive versus random sequence of video shots from a 4.5-min full vignette from *Sesame Street.* Overall, visual salience was a stronger predictor of gaze in adults than in children, especially when viewing a random shot sequence. The impact of narrative coherence on children's gaze was limited to the short period of time surrounding cuts to new video shots. The discussion considers potential direct effects of visual salience as well as incidental effects due to overlap between salient features and semantic content. The findings are also discussed in the context of developing video comprehension.

Introduction

Selective attention engages both scene-driven processes (e.g., stimulus-driven perceptual factors; Itti, 2000; Itti & Koch, 2001) and expectation-driven processes (e.g., internally driven cognitive factors; Birmingham et al., 2008; Castelhano et al., 2007). When watching dynamic scenes, such as excerpts

* Corresponding author.
E-mail address: jingme@bc.edu (M. Jing).

https://doi.org/10.1016/j.jecp.2022.105562
0022-0965/© 2022 Elsevier Inc. All rights reserved.

from television shows, children and adults alike are more likely to fixate visually salient regions than other less salient regions (Franchak et al., 2016; Frank et al., 2009). This tendency to fixate salient regions, which we call *salience-based gaze prediction* (SBGP), may be due to the overlap between visually salient features and meaningful semantic content. That is, visually salient features may signal important content or be clustered around meaningful regions of the scene (e.g., a character's face) (Henderson et al., 2007, 2019; Huston & Wright, 1983; Wass & Smith, 2014). Most of what is known about gaze allocation during video viewing, particularly among younger viewers, is based on short video excerpts. Thus, it is not clear whether SBGP is sensitive to a larger narrative context. The purpose of the current study was to determine the extent to which narrative coherence affects SBGP in young children and adults during free viewing of a complete television narrative.

Predicting eye gaze during dynamic scene viewing

Several studies have examined SBGP while adults view dynamic scenes (Itti, 2000; Itti & Koch, 2001). A relatively small number of studies have extended this work to younger viewers. For instance, Frank et al. (2009) observed infants (3–9 months) and adults watching 4-s clips from an animated movie. Frank et al. defined a predictive model using visual features such as temporal luminance contrast and spatial luminance contrast. They found that the extent to which infants fixated visually salient features on the screen was greater than would be expected by chance alone. In a similar study using a continuous 60-s clip from a children's television program, Franchak et al. (2016) calculated the salience of viewers' fixation areas based on five image channels: color, contrast, orientation, motion, and flicker (i.e., luminance differentials across frames). The average salience of the fixated areas was found to be in the top 20% of the salience of the whole screen in both infants (6–24 months) and adults. Together, these studies and others (e.g., Kadooka & Franchak, 2020; Mital et al., 2011; Pomaranski et al., 2021; Rider et al., 2018; Shepherd et al., 2010) demonstrate that visual salience predicts where eye gaze is directed in infants, school-age children, and adults alike.

Although many studies report that SBGP increases with age (Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018), this finding is not universal (Kadooka & Franchak, 2020). Moreover, some scholars posit that age differences in SBGP can be explained by superficial characteristics of the stimulus or data quality rather than underlying perceptual or cognitive abilities. For example, an age-related increase in SBGP may be explained by an increase in attentional synchrony (and thus gaze predictability) and the overlap between visually salient features and semantic content (e.g., faces), which capture attention to an increasing degree with age (Frank et al., 2009; Kiat et al., 2022; Pomaranski et al., 2021). Findings are mixed for cross-sectional studies comparing younger and older viewers. Thus, experimental research that manipulates stimulus comprehensibility while holding low-level visual properties constant could help to isolate expectation-driven processes within a given age group.

Many studies on eye gaze during video viewing use static images or brief video clips that lack narrative context, potentially reducing the extent to which expectation-driven processes drive eye gaze. Yet, even when viewing short videos with relatively little narrative context, viewers' gaze is at least partly driven by meaningful semantic features such as faces (e.g., Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018). In the case of narrative television, it is likely that a viewer's ongoing comprehension of the story underlies expectation-driven processes, allowing viewers to anticipate upcoming content (e.g., areas of interest, plot relevance). Indeed, the narrative coherence of video content influences the duration of overt looks toward the screen in viewers as young as 18 months (Anderson et al., 1981; Pempek et al., 2010). Moreover, narrative coherence affects attentional synchrony, or the consistency in eye gaze across observers, at least as early as 4 years (Kirkorian & Anderson, 2018; Wang et al., 2012).

Some studies examine the impact of comprehension on eye gaze by manipulating the narrative coherence of video while maintaining low-level perceptual features. For instance, Wang and colleagues (2012) cut movie clips into short segments at fixed intervals (e.g., every 0.5 s) and reordered the segments in a random sequence. Wang et al. found that intra-participant consistency in eye movements to normal versus random sequences depended on segment length. Specifically, intra-participant consistency was higher when segments were randomized at relatively long intervals (e.g., every 5 s) than at relatively short intervals (e.g., every 0.5 s). We found a similar effect of narrative coherence on the degree of attentional synchrony (i.e., inter- rather than intra-participant consistency), such that viewers were less likely to look at the same thing at the same time as each other when video shots were presented in a random (vs normal) sequence (Kirkorian & Anderson, 2018). It remains to be seen whether an experimental manipulation of narrative coherence would similarly affect the degree to which low-level visual features predict eye gaze. Nonetheless, the studies reviewed thus far are consistent with the hypothesis that a viewer's ongoing comprehension of the narrative influences eye gaze.

Narrative coherence may affect SBGP to the extent that visual salience signals the presence of meaningful content. For example, Taya et al. (2012) observed no systematic differences in the gaze patterns of adult participants who were instructed to free view a recording of a tennis match and those who were instructed to determine which player earned a point during the match. The similarity in eye movement patterns between the two viewing situations is likely due to the nature of the tennis match: The most visually salient regions (i.e., movement) overlapped with regions that provided task-relevant information (i.e., players earning points). As a direct test of this hypothesis, Hayes and Henderson (2017) identified regions of both perceptual salience and semantic informativeness over a scene. They found that both features predicted eye gaze. Although semantic informativeness explained more variance in eye gaze, there was a high degree of overlap in the spatial distribution of these features.

Similarly, visually salient features are likely to cluster around meaningful information (e.g., characters' faces) in children's television programs (Wass & Smith, 2014). In this way, perceptual salience may serve as a cue to draw attention toward meaningful areas in a scene. Indeed, infants' gaze patterns are more similar to those of adults when the visually salient features of a video clip are concentrated in a small area of the scene, especially when the scene also contains a face (Smith et al., 2021). Conversely, when viewing static images, children and adults are less likely to attend to faces when faces are less visually salient compared with when faces are more salient (Amso et al., 2014). Similarly, Franchak and Kadooka (2022) found that infants, children, and adults were more likely to attend to salient versus nonsalient faces when viewing short television clips. Together, these findings illustrate the interrelatedness of salient features and semantic information and that visually salient features can help viewers to orient toward semantic information.

The impact of scene changes on attention

Children's comprehension of television is partly driven by their comprehension of *filmic montage,* or video editing techniques that convey concepts through relations between shots. For example, television programs typically convey complex narratives through transitions across time, space, and character perspective. Comprehending filmic montage requires experience (Ildirar & Schwan, 2015) and well-developed cognitive abilities such as attention, memory, spatial representation, and language processing (Smith et al., 2012). As such, although television comprehension may feel like an automatic mindless process to experienced adult viewers, visually processing and comprehending video is more challenging for young children (e.g., Anderson et al., 2006; Lorch et al., 1987; Smith et al., 1985). Among the different montage techniques, of particular interest in the current study are the transitions from one video shot to another, often called *jump cuts,* such as from one camera angle to the next or from one scene to the next.

Jump cuts may be a unique type of video editing feature that affects attention in specific ways. For instance, jump cuts have been shown to elicit overt looks from inattentive viewers (Alwitt et al., 1980), likely due to the abrupt change in visual and auditory cues. Jump cuts also have unique effects on the visual fixation of attentive viewers. Adult viewers tend to fixate the center of the screen immediately after a cut to a new scene (Le Meur et al., 2007; Mital et al., 2011; Tosi et al., 1997; Tseng et al., 2009; Wang et al., 2012). As a result, attentional synchrony in adults is higher immediately following cuts to new scenes than later in those scenes (Kirkorian & Anderson, 2018; Kirkorian et al., 2012). Fixating the center of the screen may be strategic, allowing viewers to orient quickly to new scene content. Indeed, adult viewers are more likely to fixate the center of the screen following a cut to a brand new unfamiliar scene than following a cut to a different camera angle within the same familiar scene (Kirkorian et al., 2012).

Although cuts clearly affect some aspects of gaze patterns in adults, the specific impact of cuts on SBGP remains unclear. A comparison across different studies suggests that the impact of cuts may depend on the semantic relation between consecutive video shots. When adult viewers watched a series of short (4.5- to 30-s) unrelated video shots, SBGP peaked within 250 ms of jump cuts and gradually decreased over the subsequent 2500 ms (Carmi & Itti, 2006). The authors posited that viewers first oriented to perceptually salient features in the new shot before identifying and attending to semantically meaningful objects. In contrast, in a study using longer (3-min) excerpts from movies, Rider et al. (2018) found an immediate decline in SBGP after jump cuts, followed by a recovery at around 500 ms. Together, this research suggests that the impact of cuts on SBGP may differ for short unrelated video clips versus a coherent sequence of shots representing a continuous action or story.

One reason why fixation patterns may differ for coherent shot sequences versus disconnected shots is adult viewers' tendency to anticipate the reappearance of an object based on its trajectory before the cut (Kirkorian & Anderson, 2017). If viewers comprehend a coherent action sequence, they have the opportunity to make anticipatory eye movements following a cut rather than make a reactive eye movement toward salient regions. Yet young viewers' comprehension of filmic montage emerges gradually throughout early and middle childhood (Calvert & Scott, 1989; Kirkorian & Anderson, 2017; Pempek et al., 2010; Smith et al., 1985, 2012; Smith & Henderson, 2008). Children as young as 18 months show emergent comprehension of cuts and shot sequences (Pempek et al., 2010), and even infants center fixations on the screen during free viewing of natural scenes (Franchak & Kadooka, 2022; van Renswoude et al., 2019); however, young children often fail to make inferences about scene continuity and discontinuity across a sequence of visually distinct shots (e.g., across space, time, action, character intention) (Calvert & Scott, 1989; Smith et al., 1985). Together, the research suggests that children begin to comprehend edited sequences of video shots during the second year of life, but this skill continues to improve through middle childhood. Given such protracted development of video comprehension, the impact of narrative coherence on eye movements may be markedly different in young children than in adults.

Overview of the current study

This study is based on a secondary analysis of a dataset described elsewhere (Kirkorian & Anderson, 2018) to address new research questions about SBGP. The original study examined the impact of narrative coherence on attentional synchrony (i.e., inter-participant similarity in gaze location) in 4-year-old children and adults. Our current aim was to examine the extent to which narrative coherence affects SBGP during adulthood and early childhood, a period of rapid development in cognitive skills in general as well as video comprehension in particular. To directly test the impact of narrative coherence on SBGP, we compared viewers watching a normal narrative sequence with those watching a random sequence of the same video shots.

Our first analysis focused on overall effects of narrative coherence (normal vs random sequence) across the whole video. The impact of narrative coherence on SBGP was an open research question. On the one hand, disrupting narrative coherence could reduce expectation-driven processes, thereby increasing viewers' reliance on visual salience to identify meaningful information, as evidenced by greater SBGP. On the other hand, disrupting narrative coherence could result in less systematic and less predictable eye gaze, yielding lower SBGP. Given that young children have relatively limited comprehension of edited video sequences (Calvert & Scott, 1989; Smith et al., 1985), we expected any effect of narrative coherence on SBGP (whether positive or negative) to be larger in adults than in children, as evidenced by an age-by-condition interaction. We also expected gaze to be more predictable in adults than in children, as evidenced by a main effect of age with higher SBGP in adults (Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018).

Our second analysis focused on SBGP immediately following jump cuts and similar transitions to new shots given the impact of these transitions on eye movements (e.g., Kirkorian et al., 2012; Mital et al., 2011; Rider et al., 2018) and on young children's comprehension of video (Anderson & Hanson, 2010). Based on prior research with cohesive video sequences (Rider et al., 2018), we expected an initial drop in SBGP immediately after cuts to new scenes accompanied by an increased likelihood of fixating the center of the screen (Kirkorian et al., 2012; Le Meur et al., 2007; Tseng et al., 2009; van Renswoude et al., 2019). We expected this time-bound effect to be more pronounced in adults due to greater predictability of gaze and sensitivity to cuts in adults than in children. However, the impact of video condition on this time-bound effect remained an open research question. To the extent that a random video sequence increases (vs decreases) SBGP in general, we expected this condition to produce a larger (vs smaller) drop in SBGP immediately following cuts to new scenes.

Method

Participants

The current study constitutes secondary data analysis. The original sample included 33 4-year-old children and 44 adults with normal or corrected-to-normal vision. Participants were assigned at random to view the normal or random shot sequence. From the original sample, 3 children and 12 adults were dropped due to inability to calibrate the eye-tracker (e.g., distortions from reflective glasses or head movements). From the remaining 62 participants, this secondary analysis included data from the 40 participants (10 per cell) with the lowest data missingness, as described later (see "Data preprocessing and inclusion criteria" section). Thus, the final sample for this secondary analysis included 20 children (7 girls and 13 boys; M_{age} = 4.51 years, *SD* = 0.10, range = 4.36–4.74) and 20 adults (15 women and 5 men; M_{age} = 20.46 years, *SD* = 1.14, range = 18.47–22.21) divided equally into normal and random video groups.

The original study was approved by the institutional review board at University of Massachusetts Amherst. Data were collected in 2008. Child participants were recruited through letters and phone calls based on a local database of birth records. The majority (90%) of the 4-year-olds were White/non-Hispanic. As a proxy for socioeconomic status, parents reported the number of years of education they completed, with 12 years typically indicating a high school diploma, 16 years typically indicating a 4-year college degree, and so on. The average number of years of education per parent was 17.73 (*SD* = 3.88, range = 12–25). Adult participants were recruited from undergraduate psychology courses.

Stimuli

The current analysis is based on a complete vignette from the children's television program *Sesame Street.* In the original study, participants viewed the 20-s opening scene for the show, followed by the 4.5-min *Journey to Ernie* vignette used for the current analysis. The video presents a full story arc from a recurring vignette in which the character Ernie hides and other characters search for him. In the specific vignette used in the current study, Ernie says he will hide behind something that grows. Another character, Big Bird, searches for Ernie behind several plants (e.g., flowers, pumpkin, acorn) before ultimately finding Ernie behind a leaf at the top of a beanstalk. Most of the video consists of live-action puppets superimposed on a computer-animated environment. The vignette contains 28 distinct shots with an average length of 6 s (range = 3.08–37.60) presented at 25 fps (frames per second). Most transitions between shots were abrupt jump cuts (e.g., shifting from one camera angle to another, shifting from one scene to another). As such, we use the term *cut* to refer to any transition between distinct shots.

The only difference between the two experimental conditions was the order of shots in the sequence. In the normal condition, the shots were played in their original order, presenting a cohesive story. In the random condition, the 28 shots (both video and audio) were reordered in a random sequence determined by a random number generator. To create the random sequence, the shots were separated either at the exact moment of an abrupt jump cut or at the midpoint of a wipe transition across the screen. Thus, to the extent possible, the visual and auditory characteristics within each shot remained, but the narrative coherence of events across the vignette was disrupted. This disrupted sequence renders the narrative less comprehensible, as evidenced by several prior studies examining associations between narrative coherence and young children's attention to television (e.g., Anderson et al., 1981; Hawkins et al., 1995; Pempek et al., 2010).

Setting and apparatus

The study took place on a university campus in an eye-tracking laboratory room. The image display on the television set was presented in 4:3 aspect ratio with a resolution of 720 × 576 pixels. Based on the typical viewing distance of about 100 cm, the image subtended approximately 23° × 30.5° visual angle. The eye-tracking cameras sat on a table approximately 65 cm in front of the participant. The eye camera was an Applied Science Laboratories (ASL) Eye-Trac 6000, a near-infrared corneal reflection system with remote pan-tilt optics. Temporal resolution was 60 Hz. An ASL VH2 head-tracking camera used face recognition software to locate and track the viewer's head. An ASL Digital Frame Overlay was also used to insert a digital frame number from the ASL Control Unit onto the video recordings of the sessions, allowing the experimenter to sync gaze data and video stimuli.

General procedure

Upon entering the study room, the participant was seated in front of the video display screen. Children sat on a booster seat to approximate the height and viewing angle of adults. Parents sat in a chair to the right of the child participants. Parents remained in the room during the session but were asked to refrain from directing their children's attention to any particular area on the screen once the stimulus video began.

A two-point calibration procedure was used for all participants. Small animated characters appeared on the screen for 4 s each, alternating between the top-left and bottom-right corners of the screen. Adult participants were asked to look at each character; child participants were asked to "play a guessing game" by identifying the characters (e.g., mouse, robot) as they appeared. Calibration accuracy was not determined for each participant in the original study (Kirkorian & Anderson, 2018). However, given that the same calibration procedure was used for all participants, and the eye-tracker was calibrated before participants watched the stimulus video, the condition effects reported here are not likely to be explained by systematic differences in the quality of calibration for those in the normal versus random video groups.

After calibration, the experimenter indicated that it was time to watch the show. There were no additional instructions. The experimenter then started the stimulus video and began recording the gaze file and the digital video. Throughout the session, the experimenter ensured that the eye-tracker remained focused on the participant's right eye.

Parent survey

Parents of child participants completed a questionnaire on demographic information (e.g., parent's education, child's race and ethnicity). To gain a general sense of children's household television exposure, the parents also completed a retrospective viewing diary, recording their children's television exposure for each day (Monday through Sunday) in a typical week. Parents were asked to report on typical television exposure in the foreground (watching child-directed television) and background (being in the room with the television on but not watching a child-directed program). Television exposure data for adults were not collected because we expected them to be experienced television viewers with relatively little variability in the adults' ability to comprehend the *Sesame Street* vignette.

Data preprocessing and inclusion criteria

The raw eye-tracking data contained horizontal and vertical gaze coordinates originally recorded at 60 Hz. To smooth the raw data and reduce noise, gaze coordinates were down-sampled to match the frame rate of the video (25 Hz) by taking the average gaze coordinate within each frame.

Participants with high data missingness were excluded to minimize the impact of systematic data loss. Data exclusion occurred in two steps. In the first step, we excluded individual video shots within each participant if that participant was missing data for more than 50% of the shot. The first step excluded 48% of the shots across all children and 25% of the shots across all adults. In the second step, we retained 10 participants per group (20 children and 20 adults) with the fewest number of excluded shots to minimize the impact of systematic data loss and to equate sample size across groups. The latter was particularly important for analyses that pooled data across participants (Analysis 2 described later). The mean percentages of included shots were 95% (range = 89–100) for the adult–normal group, 91% (range = 86–100) for the adult–random group, 74% (range = 50–89) for the child–normal group, and 69% (range = 50–93) for the child–random group. By comparison, the percentages of shots that would have met our inclusion criteria for the excluded participants were 81% (range = 71–89) for the adult–normal group, 74% (range = 68–90) for the adult–random group, 55% (range = 18–68) for the child–normal group, and 42% (range = 11–46) for the child–random group.

We ran robustness checks in three ways: (a) including shots with at least 25% valid data (rather than 50% valid data), (b) including all shots for included participants (i.e., shots with > 0% valid data), and (c) randomly sampling 10 participants per group rather than choosing the 10 per group with the most included shots. In all cases, the qualitative pattern of results was the same as that reported here. See Tables S1–S3 in the online supplementary material for the results of robustness checks.

Data processing and reduction

The scripts for data processing were written in MATLAB (MathWorks, Natick, MA, USA). Image frames were extracted from the video as JPEG files at the rate of presentation (25 Hz). For each frame image, a salience map was generated to calculate the relative salience of each pixel using the Itti and Baldi (2005) salience algorithm as implemented in the GBVS (graph-based visual saliency) MATLAB toolbox (Harel et al., 2006). The relative salience of each pixel was determined using a combination of five feature maps that capture low-level image characteristics: color, intensity, orientation, flicker, and motion. Flicker and motion were calculated by comparing each frame with the previous frame. Feature maps were weighted equally to create an overall salience map for each frame. Each pixel within a frame was ranked with a value from 0 to 1 that represented the salience relative to all pixels in this frame, with the most salient pixel ranked as 1. For every frame, an individual participant's gaze salience was calculated as the average salience value within a 24-pixel radius of the point of gaze, which was equivalent to a visual angle of 1°. Gaze salience values for each participant were calculated for all frames with a valid gaze coordinate.

Salience-based gaze prediction was estimated using the receiver operating characteristic (ROC) curve as proposed for eye-tracking analysis by Tatler et al. (2011). The aim was to test the extent to which salience can discriminate actual gazed regions (i.e., where eye gaze lands) and actual ungazed regions (i.e., where eye gaze does not land). In the current study, on each video frame there was a gazed region (i.e., the actual observed gaze coordinate) and a corresponding ungazed region (i.e., a randomly sampled gaze coordinate from the participant's actual gaze coordinates across all other frames in the video). By randomly sampling from each participant's own gaze data rather than all possible coordinates on the screen to create "ungazed" locations, we ensured that our SBGP estimates control for systematic biases that drive attention to screens such as the tendency to fixate the center and the horizon line (Tatler, 2007; Tatler & Vincent, 2008).

Next, for a certain salience threshold, a region was classified as "gazed" if its salience was larger than the threshold and as "ungazed" if its salience was below the threshold. By comparing this classification with actual gaze coordinates, we extracted the hit rate (i.e., classifying actual gazed regions as "gazed") and the false alarm rate (i.e., classifying actual ungazed regions as "gazed") for each frame. By varying the threshold from 0 to 1 at a.001 interval, an ROC curve was plotted. The resulting area under this curve (AUC) indicated how well salience discriminated actual gazed regions from randomly selected ungazed regions, in other words, how well the observed gaze coordinates were predicted by salience. Thus, AUC was the metric for our main dependent variable, SBGP. As an advantage over regression-based prediction, this approach minimized any influence of individual- or group-level differences in data quality (e.g., calibration accuracy, amount of data captured) by comparing each individual with himself or herself rather than with others in the sample. That is, the ROC analysis determined, for each individual, how well visual salience predicted where the individual looked in that moment versus where the same individual looked at a different moment; thus, the model predication is not subject to the impact of variance of gaze data across participants.

Critically, visual salience is not a mutually exclusive video feature (Henderson et al., 2019; Smith et al., 2021; Wass & Smith, 2014). As such, AUC should be interpreted as the degree to which gaze was correlated with visual salience rather than caused by visual salience at the exclusion of other features (e.g., semantic information, inference). Uncentered, AUC had a possible range of 0 to 1, with.5 as the chance level indicating that salience equally predicted gazed and ungazed regions. In the current study, we centered AUC at the chance level, such that the centered range was −5 to.5 and values greater than 0 indicated that gazed regions were on average more salient than ungazed regions.

In addition to SBGP, distance to center (DTC) was calculated to quantify the extent to which viewers fixated the center of the screen following cuts. We computed the Euclidean distance between the screen center and each gaze coordinate.

Fig. 1 depicts how dependent variables were reduced for statistical analysis. Analysis 1 examined overall age and condition effects on SBGP across the entire vignette. We calculated SBGP using the ROC plot based on gaze locations from all the frames within a shot for every individual participant; that is, the unit of analysis was each video shot, with each participant receiving one SBGP measurement per shot. Thus, the dependent variable was calculated at the individual participant level.

Analysis 2 addressed age and condition effects on SBGP and DTC during the specific period of time surrounding cuts to new shots. Based on prior research on SBGP after cuts in older children and adults (Rider et al., 2018), we focused on time periods that began 160 ms (4 video frames) before each cut and ended 480 ms (12 video frames) after each cut. To capture the temporal change within each time

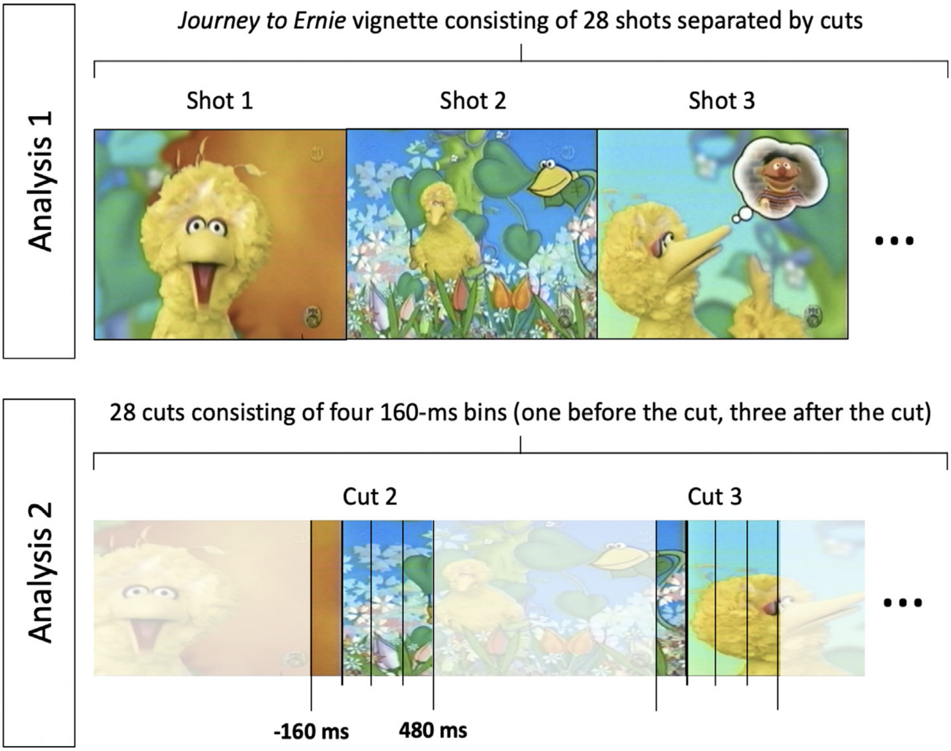
Analysis 1
Journey to Ernie vignette consisting of 28 shots separated by cuts
Shot 1
Shot 2
Shot 3
...
Analysis 2
28 cuts consisting of four 160-ms bins (one before the cut, three after the cut)
Cut 2
Cut 3
...
-160 ms
480 ms

Fig. 1. Graphical representation of the stimuli and the data structure in data preprocessing. Analysis 1 examined eye gaze for each shot (n = 28) nested within each participant (n = 40). Analysis 2 examined change over time surrounding cuts to new shots, including 160 ms before each cut to 480 ms after each cut, divided into four bins. Analysis 2 examined bins (n = 4) nested within each cut (n = 28), pooling across all 10 participants within each of the four groups.

period, we calculated SBGP and DTC within each of four 160-ms bins (see Fig. 1), resulting in four data points surrounding each cut. Note that to ensure a sufficient amount of data samples for the SBGP calculation, we pooled the gaze locations of all participants within each group. As such, the dependent variables in Analysis 2 were at the group level rather than at the individual level.

Statistical analysis

We first conducted preliminary analyses, including randomization checks and bivariate correlations, to identify potential covariates (e.g., gender, exact age within each age group, naturalistic television exposure in the child group). Next, analyses were conducted to examine the age and condition effects on gaze across the entire vignette (Analysis 1) and during the period of time surrounding cuts (Analysis 2). All the models were estimated using the function *glmer* from the package lme4 (Version 1.1–12; Bates et al., 2015) in the R software environment (Version 3.3.0; R Core Team, 2016).

In Analysis 1, the dependent variable was SBGP measured as AUC, which was calculated at the individual participant level. To account for the potential clustered standard errors at both the participant level (i.e., an individual participant may display similar eye movement patterns across the shots) and the shot level (i.e., participants may show similar patterns with each other when watching the same shots), we used multilevel modeling with shots and participants as the random effects. Given that participants and shots were crossed factors nested within each other, a crossed-mixed-effects model was fitted (Raudenbush, 1993) with observations (Level 1) nested within the participant ID and shot ID (Level 2). The proportion of variance in the outcome variable explained by the participant-level clustering (i.e., between shots, within participants) and shot-level clustering (i.e., between participants, within shots) added up to 20% of the total variance, as indicated by the intraclass correlation. The model specification was as follows:

Level 1 model:

$$\text{auc}_{ij} = \beta_{0i} + \beta_{0j} + \varepsilon_{ij}$$

Level 2 model:

$$\beta_{0i} = \gamma_{00.1} + \gamma_{01}(\text{age}_i) + \gamma_{02}(\text{condition}_i) + \gamma_{03}(\text{age}_i \cdot \text{condition}_i) + \eta_{0i}$$

$$\beta_{0j} = \gamma_{00.2} + \theta_{0j}$$

Combined model:

$$\text{auc}_{ij} = \gamma_{00} + \gamma_{01}(\text{age}_i) + \gamma_{02}(\text{condition}_i) + \gamma_{03}(\text{age}_i \cdot \text{conditio}n_i) + \eta_{0i} + \theta_{0j} + \varepsilon_{ij}$$

In these models, γ_{00} represents the (intercept) grand mean of the reference group (adult–normal) across participants and shots. γ_{0q} represents the fixed effect of variable q (q = 1, 2, 3 for age, condition, and age-by-condition interaction) on the participant-level intercept, β_{0i}, and η_{0i} adds a random effect to β_{0i}. θ_{0j} denotes the random intercept (i.e., β_{0j}) effect at the shot level. The residual at the participant-cross-shot level is represented by ε_{ij}. We also conducted exploratory analyses that tested effects of shot order (i.e., time into the vignette) and its interaction with the group variables (i.e., age, condition). Given that the model fit was not improved, we did not include this predictor in the final model (see Table S4 in supplementary material).

In Analysis 2, the dependent variables were SBGP and DTC at the group level. To account for the clustering in the nested data at the bin level (i.e., participants may display similar eye movement patterns when watching the same 160-min bins) and the shot level (i.e., participants may display similar eye movement patterns when watching different bins surrounding the same cut), we again used multilevel modeling with bins and shots as the random effects. Specifically, a three-level hierarchical mixed-effects model with groups of participants (Level 1) nested within bins (Level 2) nested within shots (Level 3) was fitted to test the fixed effect of age and video condition on SBGP and DTC change following a cut. Two splines were used to fit a piecewise linear regression with a knot fixed at the time of the cut. That is, the first spline (Time1) included Bin 1 (the last 160-ms bin before the cut, coded as −1) and Bin 2 (the first 160-ms bin after the cut, coded as 0) to model change across the cut. The second spline (Time2) included Bins 2 to 4 (the last three 160-ms bins after the cut, coded as 0–2) to model change immediately following the cut. Both splines were centered at Bin 2, which represents the intercept against which these time variables can be compared. The total intraclass correlations at the shot and bin level were 22% for SBGP and 26% for DTC. The model specification was as follows:

Level 1 model:

$$\text{outcome}_{ijk} = \beta_{0jk} + \beta_{1jk}(\text{age}_{ijk}) \; + \; \beta_{2jk}(\text{condition}_{ijk}) \; + r_{ijk}$$

Level 2 model:

$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k}(\text{time1}_{jk}) \; + \gamma_{02k}(\text{time2}_{jk}) \; + u_{0ik}$$

$$\beta_{1jk} = \gamma_{10k}$$

$$\beta_{2jk} = \gamma_{20k}$$

Level 3 model:

$$\gamma_{00k} = \delta_{000} + v_{00k}$$

$$\gamma_{01k} = \delta_{100}$$

$$\gamma_{02k} = \delta_{200}.$$

Combined model:

$$\text{outcome}_{ijk} = \delta_{000} + \delta_{100}(\text{time1}_{jk}) \; + \delta_{200}(\text{time2}_{jk}) \; + \gamma_{10k}(\text{age}_{ijk}) \; + \gamma_{20k}(\text{condition}_{ijk}) \; + v_{00k} + u_{0ik} + r_{ijk}$$

In these models, outcome denoted AUC or DTC, depending on the specific dependent variable of the analysis. δ_{000} represents the intercept for the reference group (adult–normal). δ_{100} and δ_{200} represent the linear effects of the bin variable, which model the change in SBGP over time during two time phases separated by the cut, on β_{0jk}. The time variable, which was indexed by the time order of bins, was treated as a continuous predictor centered at the cut. The time variables for the first and second pieces of the curve were denoted as Time1 and Time2, respectively. γ_{q0k} represents the fixed effect of variable q (q = 1, 2 for age and condition). u_{0ik} and v_{00k} add random effects, at the shot level and bin level respectively, to the intercept. The residual in AUC at the group level is denoted by r_{ijk}.

Results

Preliminary analyses

As a randomization check, we tested for experimental condition differences with respect to participant gender and exact age (i.e., age in years within the child or adult group). No significant differences were found in the child group [exact age: $t(16) = 1.66$, p =.116; gender: $\chi^2(1, N = 20) = 0.88$, p =.348] or the adult group [exact age: $t(17) = -0.35$, p =.730; gender: $\chi^2(1, N = 20) < 0.001$, p = 1.00].

To identify potential participant-level covariates, we calculated bivariate correlations between the dependent variables and exact age, gender, and children's typical television exposure at home. See Table 1 for descriptive statistics, including bivariate correlations. Given that these participant-level characteristics did not differ significantly by condition and were not correlated with the dependent variables, they were not considered further.

Analysis 1: Salience-based gaze prediction across the entire vignette

The distribution of salience-based gaze prediction by age and condition is illustrated in Fig. 2. The three fixed effects from the final model are reported in Table 2. Because the dependent variable was centered at chance, the intercept effect compares SBGP with chance in the reference group

Table 1
Descriptive statistics and zero-order correlations for participant-level variables.

Variable	Descriptives		Correlations		*t* Test			
	Frequency	*M* (*SD*)	SBGP	DTC	SBGP		DTC	
					M (*SD*)	*t* Ratio (*p* value)	*M* (*SD*)	*t* Ratio (*p* value)
Adult group								
SBGP	–	0.21 (0.12)						
DTC	–	104.51 (48.38)						
Exact age	–	20.46 (1.14)	−.22	−.14				
Male	15%	–			0.17 (0.07)	1.22	127.16 (6.05)	0.49
Female	85%	–			0.22 (0.05)	(.327)	129.23 (10.06)	(.650)
Child group								
SBGP	–	0.14 (0.12)						
DTC	–	137.49 (18.90)						
Exact age	–	4.51 (0.10)	.03	.34				
Male	65%	–			0.15 (0.04)	−0.60	143.90 (10.54)	1.36
Female	35%	–			0.14 (0.02)	(.560)	152.62 (17.20)	(.441)
BTV	–	0.81 (1.25)	−.37	−.01				
FTV	–	2.11 (1.64)	−.19	−.17				
Total TV	–	2.92 (2.78)	−.28	−.10				

Note. Salience-based gaze prediction (SBGP) and distance to center (DTC) were aggregated across all shots for each participant for bivariate correlations. Exact age was measured in years. The television (TV) variables represent the hours of exposure per day averaged across a typical week separately for background TV (BTV) and foreground TV (FTV) exposure as well as total TV exposure (Total TV). Data on TV exposure was available for only 14 of the children.

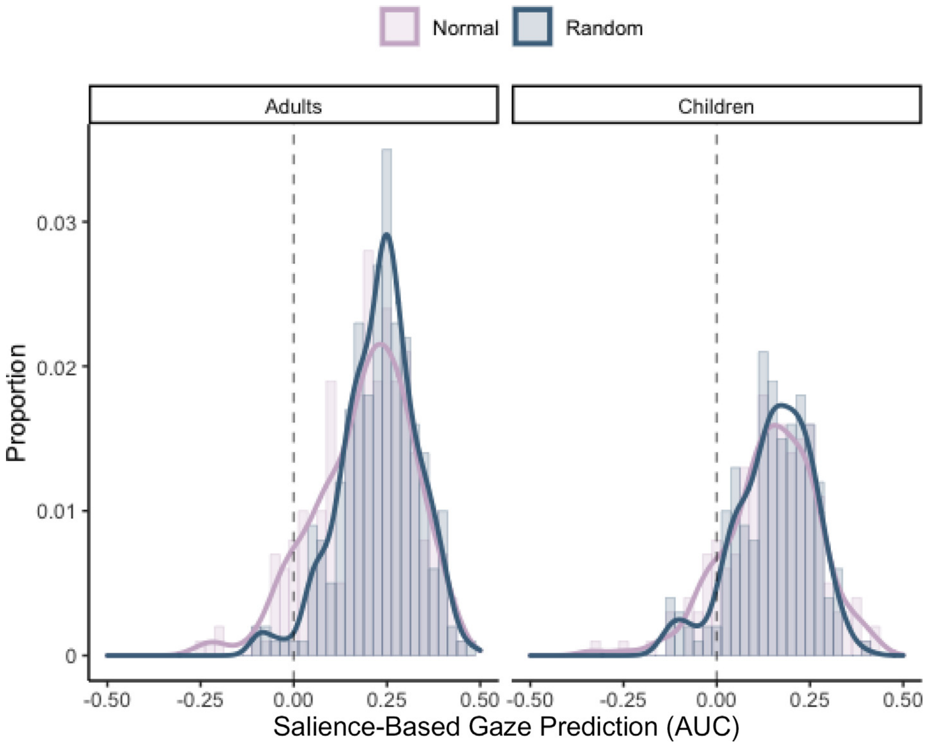

Fig. 2. Histogram and density plot of salience-based gaze prediction (SBGP) by age group and video condition. The dashed line represents the chance level, so the area to the right of the line corresponds to SBGP above chance. AUC, area under the curve.

Table 2
Fixed effects from the final mixed-effects model predicting salience-based gaze prediction for all shots in the vignette.

Predictor	β	*SE*	*t* Ratio
Intercept (γ_{00})	0.20	0.02	12.34***
Age: Child (γ_{01})	−0.05	0.01	−4.68***
Condition: Random (γ_{02})	0.04	0.01	4.32***
Age × Condition (γ_{03})	−0.03	0.01	−2.45*

Note. Age was a binary variable with adult group coded as the reference group and child group coded as the contrast group. Condition was a binary variable with normal condition coded as the reference group and random video condition coded as the contrast group.
* p <.05.
*** p <.001.

(adult–normal). SBGP for adults in the normal condition was significantly above the chance level, γ_{00} = 0.20, *SE* = 0.01, *t*(43) = 12.34, *p* <.001. Subgroup analyses not shown in Table 2 confirmed that SBGP was above the chance level in both age groups and both video conditions [child–normal: γ = 0.15, *SE* = 0.02, *t*(43) = 10.88, *p* <.001; child–random: γ = 0.16, *SE* = 0.02, *t*(43) = 9.56, *p* <.001; adult–random: γ = 0.24, *SE* = 0.03, *t*(43) = 15.86, *p* <.001].

The age effect tests the difference between children and adults in the reference condition (normal). In the normal video condition, adults (*M* = 0.20, *SE* = 0.01) had higher SBGP than children (*M* = 0.15, *SE* = 0.02), γ_{01} = −0.05, *SE* = 0.01, *t*(1039) = −4.68, *p* <.001. The condition effect tests the difference between random and normal video in the reference age group (adults). Adults' SBGP was higher in the random video condition (*M* = 0.24, *SE* = 0.03) than in the normal condition (*M* = 0.20, *SE* = 0.01), γ_{02} = 0.04, *SE* = 0.01, *t*(1039) = 4.32, *p* <.001.

Including the age-by-condition interaction significantly improved the fit of this model, χ^2(1) = 5.89, *p* <.05. The interaction effect was significant, indicating that the condition effect was moderated by age, γ_{03} = −0.03, *SE* = 0.01, *t*(1050) = −2.45, *p* <.05. In a follow-up model presented in Table S6 of the supplementary material, we rotated the reference groups to examine this interaction effect. In the random video condition, SBGP was again significantly higher in adults (*M* = 0.74, *SE* = 0.03) than in children (*M* = 0.65, *SE* = 0.02), as indicated by a significant age effect, γ_{01} = 0.08, *SE* = 0.01, *t* (1039) = 7.96, *p* <.001. However, the condition effect was not significant where children were the reference group, γ_{02} = 0.00, *SE* = 0.01, *t*(1039) = −0.43, *p* >.400. That is, for children, SBGP did not differ for the normal video condition (*M* = 0.65, *SE* = 0.02) versus the random video condition (*M* = 0.66, *SE* = 0.02). Although SBGP was lower in children than in adults, children's SBGP was still significantly greater than chance, as indicated by a significant intercept effect with children as the reference group, γ_{00} = 0.15, *SE* = 0.02, *t*(43) = 9.56, *p* <.001.

Exploratory analyses were conducted to test the extent to which naturalistic television exposure moderated the effect of condition on SBGP across the entire vignette. The effect of video condition on SBGP did not differ between children with high versus low television exposure whether considering background, foreground, or total television. See Supplementary Material S5.

Analysis 2: Gaze patterns surrounding cuts to new shots

Salience-based gaze prediction

The temporal evolution of SBGP following a cut is plotted in Fig. 3 as a function of age group and video condition. The plot depicts a decrease in SBGP at the cut in all four groups, followed by a recovery of around 300 ms into the shots.

The mixed-effects model examined SBGP as a function of age group, condition, and time measured as Time1 and Time2. A full report of fixed effects can be found in Table 3. Consistent with the patterns shown in Fig. 3, there was a significant negative effect of Time1, δ_{100} = −0.17, *SE* = 0.03, *t*(4) = −6.06, *p* <.010, and a significant positive effect of Time2, δ_{200} = 0.22, *SE* = 0.04, *t*(4) = 6.04, *p* <.010, suggesting a U-shaped pattern such that SBGP decreased as the cut occurred and then recovered following the cut. A main condition effect was found such that SBGP was higher for the random video (*M* = 0.57,

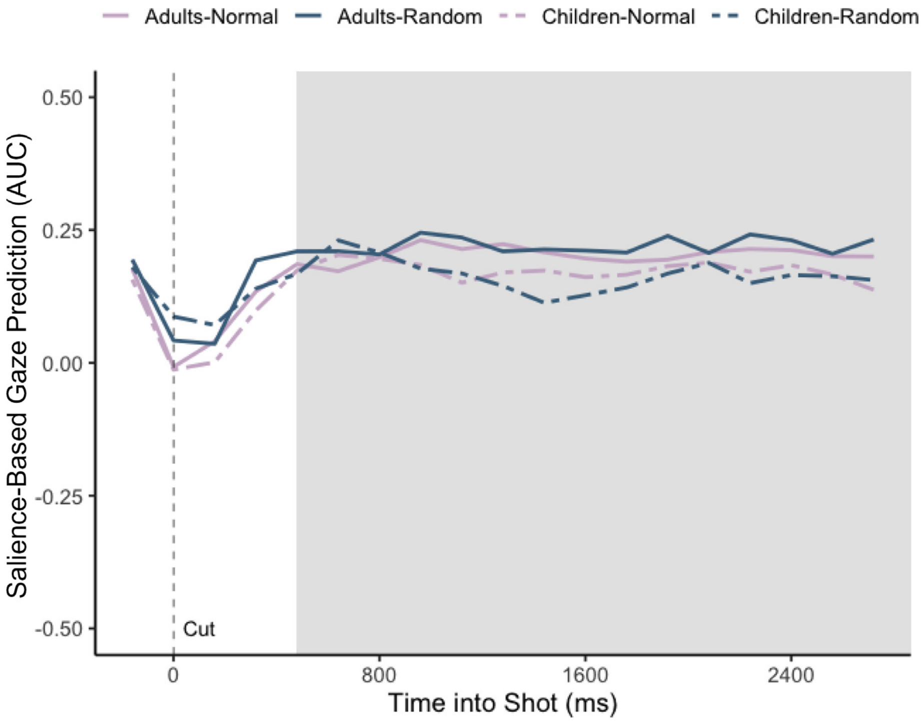

Fig. 3. Temporal evolution of salience-based gaze prediction (SBGP) by age group and video condition surrounding the cut to a new shot. Values are averaged over all shots. Analysis 2 focused on the period of time immediately surrounding the cut, as indicated by the unshaded area from −160 to 480 ms. AUC, area under the curve.

Table 3
Fixed effects from the final mixed-effects model predicting salience-based gaze prediction surrounding the cut to a new shot.

Predictor	β	*SE*	*t* Ratio
Intercept (δ_{000})	0	0.02	−0.23
Time1 (δ_{100})	−0.17	0.03	−6.06**
Time2 (δ_{200})	0.22	0.04	6.04**
Age: Child (γ_{10k})	−0.01	0.01	−0.80
Condition: Random (γ_{20k})	0.04	0.01	3.26**

Note. Age was a binary variable with adult group coded as the reference group and child group coded as the contrast group; Condition was a binary variable with normal condition coded as the reference group and random video condition coded as the contrast group.
**p <.01.

SE = 0.17) than for the normal video (*M* = 0.62, *SE* = 0.16), δ_{20k} = 0.04, *SE* = 0.01, $t(416)$ = 3.26, p <.010. However, there was no difference in SBGP between the children (*M* = 0.59, *SE* = 0.16) and the adults (*M* = 0.60, *SE* = 0.17), δ_{10k} = −0.01, *SE* = 0.01, $t(416)$ = −0.80, p >.05. The age-by-condition interaction did not significantly improve the fit of the model, $\chi^2(1)$ = 1.08, p =.298, and therefore was not included. Together, the model results suggest that SBGP was higher for the random video than for the normal video during this short period of time following cuts to new shots regardless of age group.

Distance to center

We examined DTC immediately following cuts because the tendency to center gaze at these shot transitions may contribute to a drop in SBGP following cuts. Fig. 4 plots the temporal evolution of

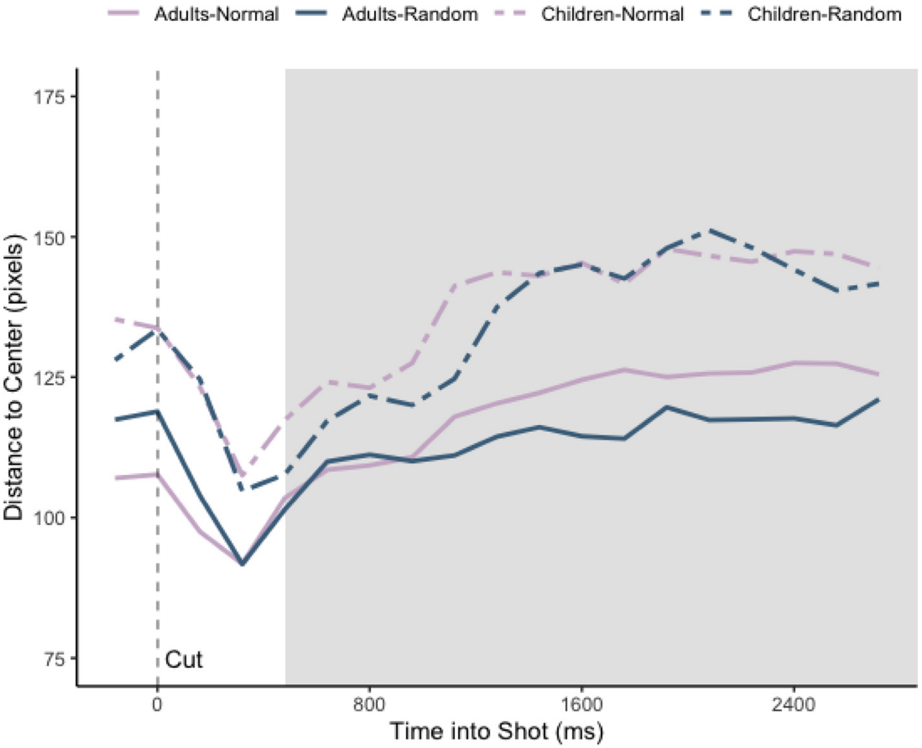

Fig. 4. Temporal evolution of distance to center by age group and video condition surrounding the cut to a new shot. Values are averaged over all shots. Analysis 2 focused on the period of time immediately surrounding the cut, as indicated by the unshaded area from −160 to 480 ms.

DTC averaged over all shots as a function of age group and video condition. Similar to SBGP, the plot depicts a decrease in DTC immediately after transitions to new shots in all four groups.

Multilevel modeling was used to test the effect of age group and video condition on DTC following a cut to a new shot, paralleling the model examining the temporal evolution of SBGP after cuts. The results are reported in Table 4. There was a nonsignificant effect of bin number at Time1, δ_{100} = 1.88, *SE* = 5.00, *t*(420) = 0.38, *p* =.707, and a significant negative effect of bin number at Time2, δ_{200} = −12.25, *SE* = 2.58, *t*(420) = −4.76, *p* <.001, suggesting a decrease of DTC after a cut. An age effect was found such that children displayed larger DTC (*M* = 123.37, *SE* = 14.62) than adults (*M* = 106.14, *SE* = 14.65) immediately following a cut, δ_{10k} = 19.35, *SE* = 3.69, *t*(420) = 5.25, *p* <.001.The condition effect was not significant, δ_{20k} = 2.38, *SE* = 3.69, *t*(420) = 0.52, *p* =.518. Moreover, the

Table 4
Fixed effects from the final mixed-effects model predicting distance to center surrounding the cut to a new shot.

Predictor	*β*	*SE*	*t* Ratio
Intercept (δ_{000})	112.97	6.22	15.15***
Time1(δ_{100})	1.88	5.00	0.38
Time2 (δ_{200})	−12.25	2.58	−4.76***
Age: Child (γ_{10k})	19.35	3.69	5.25***
Condition: Random (γ_{20k})	2.38	3.69	0.52

Note. Age was a binary variable with adult group coded as the reference group and child group coded as the contrast group; Condition was a binary variable with normal condition coded as the reference group and random video condition coded as the contrast group.
***p <.001.

age-by-condition interaction did not significantly improve the fit of the model, $\chi^2(1) = 1.57$, p =.210, and therefore was not included. Thus, the DTC was higher in children than in adults, regardless of condition, during this short period of time immediately following cuts.

Discussion

The purpose of this study was to determine (a) whether narrative coherence affected the degree to which gaze was predicted by low-level visual features during naturalistic video viewing and (b) how salience-based gaze prediction changed immediately following cuts to new video shots. We found that both 4-year-olds' and adults' eye movements were predicted by visual salience given that the discrimination of gazed versus randomly selected ungazed regions based on visual salience was greater than would be expected by chance alone. Overall, across the entire video, visual salience was a stronger predictor in adults than in 4-year-olds. However, whereas visual salience predicted 4-year-olds' eye movements equally in both video conditions, adults were more likely to direct their gaze at visually salient regions while watching the shots in a random sequence than in their original sequence.

Effects of narrative coherence on adults' salience-based gaze prediction

Why might SBGP increase when adults watch random video sequences? Presumably, when expectation-driven processes are disrupted, adults may rely more heavily on salient visual features to guide their attention to the most informative parts of the scene. For example, prior research on eye gaze patterns before and after cuts demonstrated that adults sometimes anticipate the reappearance of an object after a cut (Kirkorian & Anderson, 2017). This expectation-driven approach could lead viewers to look at parts of the screen with low salience such as an empty part of the screen where the object is about to reappear. By contrast, without coherent plot and action sequences, adults watching a random shot sequence may rely more heavily on salient visual features to quickly locate meaningful content given that perceptual salience tends to coincide with meaningful regions in video (Henderson et al., 2019; Smith et al., 2021; Wass & Smith, 2014). Indeed, in the current study, the most salient regions in each shot often (but not always) overlapped with meaningful information such as characters' faces (see Fig. S1 in the supplementary material).

Whether automatic or intentional, salience-based eye gaze could help adults to become more efficient viewers in different viewing scenarios. Just as Smith and Mital (2013) argued that an explicit task goal (i.e., identifying the locations depicted in a video clip) could direct viewers' attention away from visually salient features, so too might a coherent narrative direct comprehending viewers to look toward otherwise uninteresting parts of the screen. For instance, adults are more likely than children to make anticipatory eye movements if they can predict where an object will reappear. In the absence of such predictability, centering gaze after a cut or fixating the most salient region might be a useful strategy. Together, these findings indicate that disrupting expectation-driven processes in turn increases SBGP in adult viewers. These findings also extended prior work on the role of narrative coherence in dynamic attention to video, which demonstrated narrative coherence effects on viewers' reaction time (Hinde et al., 2018; Lorch & Castle, 1997) and on attentional synchrony both within viewers (Wang et al., 2012) and between viewers (Kirkorian & Anderson, 2018).

Effects of narrative coherence on children's salience-based gaze prediction

Children had lower SBGP than adults, particularly when watching a random video sequence. This finding replicates several prior studies (Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018). Lower SBGP in children might reflect less advanced comprehension than in adults given that salient regions are more likely to be meaningful and informative (e.g., Henderson et al., 2019; Taya et al., 2012; Wass & Smith, 2014). Indeed, adults tend to fixate regions they understand to be semantically relevant or where they anticipate something interesting to occur (e.g., Henderson et al., 1999; Kirkorian & Anderson, 2017; Land et al., 1999; Morgante et al., 2008). Eye movements may be less predictable in younger viewers due to less viewing experience or still-developing cognitive skills such as attentional control (van Renswoude et al., 2016, 2019). However, age differences in SBGP are not always found (Kadooka & Franchak, 2020). Moreover, such age differences may be due to a wide range of confounding factors such as age differences in data quality and attentional synchrony (Frank et al., 2014; Kiat et al., 2022; Pomaranski et al., 2021), as described later.

Given that there are many plausible explanations for overall age differences in SBGP—many of which are unrelated to our main questions here—it may be more fruitful to focus on condition effects *within* the child age group. Like adults, 4-year-olds showed slightly higher SBGP when watching the randomly edited video. However, unlike adults, the condition effect was isolated to the short period of time immediately following cuts (Analysis 2). The condition effect was not significant in children when considering the overall main effect across the entire video (Analysis 1). On the surface, this might be due to 4-year-olds' insufficient comprehension of the normal video sequence in the first place. Prior research demonstrates that children at this age have limited understanding of video content, with adult-like comprehension emerging at around 12 years of age (Anderson & Hanson, 2010; Collins & Wellman, 1982). Thus, the degree to which their visual attention is controlled by content comprehension may be relatively low even when watching a normal video sequence. This could explain why children's overall SBGP was similar in the two conditions given that randomizing the shot sequences does not change the location of salient regions within each shot.

Although video comprehension develops gradually across childhood, accumulated evidence demonstrates that young children are sensitive to the narrative coherence of video sequences, reflecting some—albeit limited—comprehension of video (e.g., Pempek et al., 2010; Richards & Cronise, 2000; Smith et al., 1985). Indeed, our Analysis 2 shows some evidence for a condition effect in children when examining the short period of time surrounding transitions to new video shots. Like adults, 4-year-olds' SBGP dropped immediately following cuts and was higher for the random shot sequence than for the normal video sequence. This adult-like eye movement pattern suggests that differences existed in processing shot transitions while the children were viewing the normal versus random video. Moreover, a prior study, using the same random sequence manipulation as the current study, found that children as young as 18 months made longer looks toward normal sequences as compared with random sequences of the same shots (Pempek et al., 2010). Such a looking preference for coherent video sequences indicates that a perception of shot relations emerges during late infancy. Thus, expectation-driven processes based on a narrative sequence may become another meaningful feature that guides attention during early childhood, building on those that emerge during infancy (Kiat et al., 2022).

Given that children are sensitive to random video manipulations by 18 months of age, why did 4-year-olds' SBGP show relatively little change while viewing the incomprehensible video in the current study? It is possible that the earliest beginnings of video comprehension are evident in overt looks *toward* the screen as in Pempek et al. (2010) but not in their eye movement patterns *within* the screen as observed in the current study. In this sense, differences in the degree of SBGP might be more subtle than differences in overt gaze toward the screen, requiring a more sophisticated understanding of the narrative. Alternatively, given that eye movement patterns tend to be more idiosyncratic in children than in adults (Frank et al., 2011; Franchak et al., 2016; Pomaranski et al., 2021), perhaps a general effect of condition across the video was less detectable in children. Indeed, we previously reported lower attentional synchrony among children than among adults in this data set, which could at least partly explain lower SBGP and a smaller condition effect for children (Kirkorian & Anderson, 2018).

Limitations and future directions

The current study has some limitations that should be considered when interpreting the results and considering future research directions. First, as a secondary data analysis, we are limited by the quantity, quality, and representativeness of the existing data set. Our samples were not representative of the general population with respect to many characteristics, including gender, race, ethnicity, and parent education. Eye-tracking technologies and computational methods have improved since these data were collected. The findings should be replicated in a larger and more diverse sample of participants with less data loss and participant-level measures of calibration accuracy. Such data are necessary to disentangle true age differences in SBGP from age differences due to data quality such as calibration accuracy and the number of data points meeting inclusion criteria.

Related to the issue of data quality, we applied relatively conservative data exclusion criteria to minimize the impact of systematic data loss. This approach may limit the generalizability of our findings to high-quality datasets and potentially more compliant children. Notably, we ran several robustness checks suggesting that the amount or quality of data alone is unlikely to explain the full set of results in our sample, particularly as they relate to condition effects within each age group. Nonetheless, the findings presented here require replication in a new sample with higher data quality as well as cognitive assessments that could be included as covariates (e.g., attentional control).

Future research should also consider individual differences as they relate to children's prior experience with television and other edited video as a moderator of eye gaze (e.g., SBGP, anticipatory eye movements). We did not observe a correlation between SBGP and naturalistic television viewing in the child sample. However, our sample was small, and the media exposure measure did not capture the content of video exposure. Future work could examine the impact of video viewing experience, with a focus on video content, on the effects observed in the current study.

In addition, the current study did not have a direct measure of comprehension. Although we believe that viewers' comprehension was successfully manipulated by rearranging the video shots in a random order, future research should directly examine the relation between SBGP and viewers' comprehension of the video.

Finally, this study focused on gaze prediction by visually salient features, which are likely to overlap with meaningfulness and other features (Henderson et al., 2019; Pomaranski et al., 2021; Wass & Smith, 2014). As such, effects observed in the current study may be due to passive orientation toward salient features, active selection of semantic information, anticipation of meaningful content, or any combination of these things. Nonetheless, evidence based on infants' attention to static images (Pomaranski et al., 2021) suggests that perceptual salience alone does not fully explain age-related increases in attentional synchrony. Thus, future research could isolate the effects of narrative coherence on expectation-driven processes, perhaps using measures that do not rely on visual salience (e.g., anticipatory eye movements, comprehension posttests).

Conclusion

We found that disrupting narrative coherence increased overall SBGP in adults. This effect is likely due to a greater reliance on visually salient cues to locate potentially meaningful content when expectation-driven processes were less useful. In 4-year-old children, this effect was limited to the short period of time immediately following cuts to new shots. Prior research suggests that age differences may be due to age-related increases in video comprehension, attentional synchrony (and thus predictability), or both. Together, the current findings underscore the complex relation between visual attention and video comprehension.

Data availability

The authors do not have permission to share data.

Acknowledgments

Aspects of this research were presented at the 2019 biennial meeting of the Cognitive Development Society, Louisville, KY. This work was supported by the National Science Foundation (BCS-0623888). Findings and opinions expressed in this manuscript do not reflect endorsement by the National Science Foundation. We are grateful to Daniel R. Anderson for approving use of the dataset for secondary analysis.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2022.105562.

References

Alwitt, L. F., Anderson, D. R., Lorch, E. P., & Levin, S. R. (1980). Preschool children's visual attention to attributes of television. *Human Communication Research, 7*(1), 52–67.

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), e85701.

Anderson, D. R., Fite, K. V., Petrovich, N., & Hirsch, J. (2006). Cortical activation while watching video montage: An fMRI study. *Media Psychology, 8*(1), 7–24.

Anderson, D. R., & Hanson, K. G. (2010). From blooming, buzzing confusion to media literacy: The early development of television viewing. *Developmental Review, 30*(2), 239–255.

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, 52*, 151–157.

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

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*(7), 986–998.

Calvert, S. L., & Scott, M. C. (1989). Sound effects for children's temporal integration of fast-paced television content. *Journal of Broadcasting & Electronic Media, 33*, 233–246.

Carmi, R., & Itti, L. (2006). The role of memory in guiding attention during natural vision. *Journal of Vision, 6*(9), Article 4.

Castelhano, M. S., Wieth, M., & Henderson, J. M. (2007). I see what you see: Eye movements in real-world scenes are affected by perceived direction of gaze. In *International Workshop on Attention in Cognitive Systems* (pp. 251–262). Springer.

Collins, W. A., & Wellman, H. M. (1982). Social scripts and developmental patterns in comprehension of televised narratives. *Communication Research, 9*(3), 380–398.

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., & Kadooka, K. (2022). Age differences in orienting to faces in dynamic scenes depend on face centering, not visual saliency. *Infancy*. https://doi.org/10.1111/infa.12492. Advance online publication.

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*(2), 160–170.

Frank, M. C., Vul, E., & Saxe, R. (2011). Measuring the development of social attention using free-viewing. *Infancy, 17*, 355–375.

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). MIT Press.

Hawkins, R. P., Tapper, J., Bruce, L., & Pingree, S. (1995). Strategic and non-strategic explanations for attentional inertia. *Communication Research, 22*, 188–206.

Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *Journal of Vision, 17*(5), Article 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 movement research: Insights into mind and brain* (pp. 537–562). Elsevier.

Henderson, J. M., Hayes, T. R., Peacock, C. E., & Rehrig, G. (2019). Meaning and attentional guidance in scenes: A review of the meaning map approach. *Vision, 3*(2), Article 19.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228.

Hinde, S. J., Smith, T. J., & Gilchrist, I. D. (2018). Does narrative drive dynamic attention to a prolonged stimulus? *Cognitive Research: Principles and Implications, 3*(1), 1–12.

Huston, A. C., & Wright, J. C. (1983). Children's processing of television: The informative functions of formal features. In J. Bryant & D. R. Anderson (Eds.), *Children's understanding of television: Research on attention and comprehension* (pp. 35–68). Academic Press.

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

Itti, L. (2000). *Models of bottom-up and top-down visual attention*. California Institute of Technology. Doctoral dissertation.

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). Institute of Electrical and Electronics Engineers.

Itti, L., & Koch, C. (2001). Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging, 10*(1), 161–169.

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.

Kiat, J. E., Luck, S. J., Beckner, A. G., Hayes, T. R., Pomaranski, K. I., Henderson, J. M., & Oakes, L. M. (2022). Linking patterns of infant eye movements to a neural network model of the ventral stream using representational similarity analysis. *Developmental Science, 25*(1). Article e13155.

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

Kirkorian, H. L., & Anderson, D. R. (2018). Effect of sequential video shot comprehensibility on attentional synchrony: A comparison of children and adults. *Proceedings of the National Academy of Sciences of the United States of America, 115*(40), 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.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311–1328.

Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research, 47*(19), 2483–2498.

Lorch, E. P., Bellack, D. R., & Augsbach, L. H. (1987). Young children's memory for televised stories: Effects of importance. *Child Development, 58*(2), 453–463.

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*(1), 111–127.

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.

Morgante, J. D., Haddad, J. M., & Keen, R. (2008). Preschoolers' oculomotor behavior during their observation of an action task. *Visual Cognition, 16*(4), 430–438.

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*(5), 1283–1293.

Pomaranski, K. I., Hayes, T. R., Kwon, M. K., Henderson, J. M., & Oakes, L. M. (2021). Developmental changes in natural scene viewing in infancy. *Developmental Psychology, 57*(7), 1025–1041.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics, 18*(4), 321–349.

Richards, J. E., & Cronise, K. (2000). Extended visual fixation in the early preschool years: Look duration, heart rate changes, and attentional inertia. *Child Development, 71*(3), 602–620.

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.

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*(7), 649–656.

Smith, R., Anderson, D. R., & Fischer, C. (1985). Young children's comprehension of montage. *Child Development, 56*(4), 962–971.

Smith, T., & Henderson, J. (2008). Attentional synchrony in static and dynamic scenes. *Journal of Vision, 8*(6). Article 773.

Smith, T. J., Levin, D., & Cutting, J. E. (2012). A window on reality: Perceiving edited moving images. *Current Directions in Psychological Science, 21*(2), 107–113.

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*(8), Article 16.

Smith, T. J., Mital, P. K., & Dekker, T. M. (2021). The debate on screen time: An empirical case study in infant-directed video. In M. S. C. Thomas, D. Mareschal, & V. Knowland (Eds.), *Taking development seriously: A Festschrift for Annette Karmiloff-Smith* (pp. 257–279). Routledge.

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*(14), Article 4.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5), Article 5.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2*(2), Article 5.

Taya, S., Windridge, D., & Osman, M. (2012). Looking to score: The dissociation of goal influence on eye movement and meta-attentional allocation in a complex dynamic natural scene. *PLoS One, 7*(6). Article e39060.

Tosi, V., Mecacci, L., & Pasquali, E. (1997). Scanning eye movements made when viewing film: Preliminary observations. *International Journal of Neuroscience, 92*(1–2), 47–52.

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*(7), Article 4.

van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E. J., & Visser, I. (2016). Do infants have the horizontal bias? *Infant Behavior and Development, 44*, 38–48.

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

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), Article 16.

Wass, S. V., & Smith, T. J. (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. *Infancy, 19*(4), 352–384.