PRE-PRINT

Annette Memorial Collection

Word count 7048 (ex. Refs) /6000

The debate on screen time:
an empirical case study in infant-directed video.

Tim J. Smith[1*], Parag K. Mital[2], & Tessa M. Dekker[3],

Running head: The debate on screen time

Key words: screen time, developmental cognition, TV, eye movements, infancy

---

[1] Birkbeck, University of London

[2] HyperSurfaces

[3]UCL

*Corresponding Author Dr. Tim J. Smith (tj.smith@bbk.ac.uk)

,

Annette Karmiloff-Smith cared deeply about engaging with the debate around screen time. As expressed in the title of her book chapter '*TV is bad for children' – less emotion, more science please!'*, she was outraged by the blinkered and emotive proclamations made in the press and by some academics, on the apparent damaging effect of screen time on child development (Christakis & Zimmerman, 2006; Sigman, 2012, Greenfield, 2015), and by the similarly unfounded claims from infant-directed media producers about the beneficial effects of their products on infant development (Karmiloff-Smith, 2012). With screen exposure occurring at increasingly younger ages (Anderson & Hanson, 2010; Courage & Howe, 2010; Cristia & Seidl, 2015; Bedford et al, 2016), parents are further confused by highly contradictory recommendations on what constitutes "healthy" screen time for infants and toddlers from national and international authorities (e.g. American Academy of Pediatrics, Radesky et al., 2016; WHO, 2019; The Royal College of Pediatricians and Child Health, Viner, Davie, & Firth, 2019).

Annette was a great pragmatist who understood the importance of working with parents, media producers, and policy makers, to find the best ways for infants to benefit from scientific insights into development. Recalling that some regions of human cortex are initially very immature (Hill et al., 2010) and that infants are born into a world filled with screen-based media (Anderson & Hanson, 2010; Courage & Howe, 2010; Wartella et al., 2010), she argued it is crucial to ensure that infant screen experiences are as developmentally and educationally appropriate and as scientifically informed as possible. She believed this required empirically investigating the bidirectional relationship between how the developing visual system responds to screen media and how designers intuitively tailor their media to developmental constraints. Her argument was that the discussion of whether screen time is good or bad for infants is unhelpful and simplistic; rather than debating whether screen exposure itself is intrinsically 'good' nor 'bad' we should investigate empirically what differentiates positive from negative screen experiences (see Courage, 2017 for a recent review). Annette therefore decided to collaborate with Abbey Media to develop infant-directed DVD content designed specifically with infant cognitive research in mind. Echoing her approach in other areas of research, Annette argued that understanding the effects of video on the developing infant requires careful investigation of how video content affects infant behaviour. For decades, infancy researchers have been using stimuli shown on screens, to carefully measure infant knowledge and learning at different post-natal ages. She believed that infant-directed videos and apps would do well to draw upon this extensive body of knowledge about the infant visual system and its changing limitations and demands. Whilst never a central focus of her research, Annette's work on screen time spanned many years and encompassed various approaches including public education (e.g. public lectures, TV including *Baby It's You!*, her work with Proctor & Gamble and Nursery World), collaboration with industry to develop scientifically-inspired infant directed video (with Abbey Home Media; more details below), understanding how infants watch screen media, and latterly studying the long-term associations between screen media use and infant development (see Discussion). This symbiotic approach to science, policy work and public engagement illustrates Annette's dedication to "impactful" research before the buzzword even appeared on the radar of most UK academics.

When and what can infants learn from video?

Infancy studies have used video to investigate, amongst others, which visual features (fine detail, motion, colour, depth) infants of different ages can detect, how infants learn to remember and retrieve hidden objects, recognise impossible or unexpected events, learn language, and imitate actions, with varying degrees of success depending on age and manipulations (Bremner & Wachs, 2010; Braddick & Atkinson, 2011). For example, after showing 11-month-old infants several hours of video that incorporated stimuli classically used to measure infant's cognitive control and flexibility, one study showed significant improvements in the ability to sustain and flexibly deploy attention 3 days later (Wass, Porayska-Pomsta, & Johnson, 2011). Unsurprisingly, several studies have reported that infants learn better from real life than identical content presented on a screen (DeLoache, 1991). However, this screen-learning deficit may be smaller in young infants (<6 months) who may possess fewer cognitive capacities needed to distinguish between real-life and screen presentations (Barr et al., 2007), and in older infants, learning from screens improves with repetition, spacing of learning trials, applying standard techniques of video production and, at least by two years of age, experience with television as a medium in general (Troseth, 2003; Anderson & Pempek, 2005; Barr et al., 2007).

This large body of scientific research clearly shows that, while television content cannot and should not replace the richness of real-life, even young infants can learn from this medium. Moreover, the dynamic content of video and the interactive features of apps, offer clear benefits over static books when similar levels of parental engagement and interaction are involved. Annette therefore frequently suggested that video and apps might be used like a dynamic book (although this view did get her in hot water when misquoted and then retracted by the Sunday Times; June 14th, 2015).

To set an example for how other developmental scientists may directly shape the future direction of infant-directed video Annette extracted 7 key recommendations from research on infant development, which guided her collaboration with Abbey Home Media on Baby Bright (these recommendations are collated from Annette's various unpublished and published texts including Karmiloff-Smith, 2012):

1. **Allow clearer differentiation between those aspects that are required for the processing message (the 'signal') and those that are not (the 'noise')**. Infant brains are noisier, less sensitive, and more naive information processing systems. This need for simplification of the sensory environment is evident in the exaggeration and accentuation of speech by caregivers (i.e., Motherese; Ferguson 1964) and the same simplification is required for visual communication. Most adult screen media involves social interactions and action sequences that are incomprehensible to the infant brain as they assume knowledge and perceptual skills yet to be developed (Valkenburg & Vroone, 2004).
2. **Foreground shapes on screen need to "stand-out" from their background.** Our adult visual system has sufficient top-down knowledge to separate the two grounds easily even when borders are hard to decode, but the infant system finds this more difficult (Baker, Tse, Gerhardstein & Adler, 2008). Infant gaze is more driven by visual saliency than adult gaze with a rapidly increasing bias towards semantically-relevant features such as faces over the first year of life (Frank et al., 2009; Saez De Urabain, 2015; Franchak et al., 2016), increasing throughout childhood (Rider et al., 2018). Infants and young children do not look differently at video clips that have their semantics violated via shuffling

compared to unshuffled versions (Pempek et al., 2010; Kirkorian and Anderson, 2018) suggesting that designers cannot rely on semantics to guide infant attention and must instead use visual saliency.

3. **Avoid passive central fixation**. In adult TV, salient characters and action in most programmes feature at the screen centre, and movement of these characters is conveyed by movement of the background as a camera tracks their movement (Smith, 2013). Adult viewers re-interpret this as movement of the foregrounded central character, despite the fact that the viewers' eyes are fixated on the centre keeping eye movements to a minimum. By contrast, real movement rather than inferred movement is likely to attract the infant visual system; yet, to focus on the central character of most adult programmes, infants hardly need to move their eyes. This makes them very passive observers.
4. **Sound can be cognitively distracting.** What would seem like mere auditory decoration in the background of screen exposure may add significant cognitive load to the young children, particularly when it does not meaningfully connect visual and auditory content. A nice example of this is in the number domain. Background music while showing pairs or trios of objects to demonstrate the numbers '2' or '3', may be more disruptive than playing two drum beats or three drum beats, which match the number of visual objects on the screen, the latter procedure perhaps enhancing learning by providing multi-modality audio-visual representations of the same content.
5. **Use frequent repetitions.** Adult programming involves relatively few repetitions of event sequences; adults can learn from a single presentation, whereas infants' cognitive systems need repetition (Fiser & Aslin, 2002; Pelucchi, Hay & Saffran, 2009).
6. **Use exaggerated visual action.** A large proportion of new information in adult TV is conveyed through dialogue or nuanced facial expressions rather than clear visual action, which both require perceptual abilities that develop slowly over the first few years of life.
7. **Slow Down!** When new information is provided visually it is generally at a pace faster than can be processed by the sluggish attention/perceptual systems of infants, leading to missed comprehension or fatigue. The average time between overt attention shifts in infants whilst watching videos is almost twice as long

(M = 644 ms, SD = 142) as in adults (M = 327 ms, SD = 45; Saez De Urabain, Nuthmann, Johnson, & Smith, 2017).

Annette was hopeful that the incorporation of these recommendations in infant-directed TV (I-DTV) might result in a screen experience more suitable to the developing mind. However, she was keenly aware that the translation of these recommendations into a commercial product comes with challenges, because other factors beyond the experience of the infant, such as attractiveness to the parent who buys and also watches the DVD, may also play a role in the design process. She also noted that failures of adult-directed TV may also be addressed intuitively by I-DTV creators with no scientific consultant on board. Indeed, I-DTV content analyses by Wass and Smith (2015) reveal that much I-DTV optimizes the audiovisual stimulus to simplify their infant viewers' task of deciding what to attend to (i.e., the signal) over the irrelevant background features (e.g., noise). This is done, for example, by aligning peaks in low-level visual features (e.g., luminance, colors, edges, flicker, and motion) with the location of a speaking face (the *signal*), increasing shot length, and reducing edits to give young children more time to locate the focal object within a frame. Interestingly, these formal differences between I-DTV and Adult TV are not entirely due to the I-DTV being animated, as they persisted when comparing animations directed at infants versus those directed at adults, suggesting that their creators understood the need to tailor the flow of audiovisual information to the age of their respective audiences. Being an empiricist, Annette therefore viewed the incorporation of these scientifically-informed considerations as hypotheses about how infants may respond to I-DTV and, as such she endeavoured to test these hypotheses in an experimental study with author TD and later, TS.

The following section presents some detailed empirical work that illustrates the kinds of methods Annette believed were needed to ground an evidence-based approach to understanding the impact of screen time on infant development. For readers who do not want to descend into this level of detail, feel free to skip ahead to the end of this section for the take-home points.

In a novel empirical case study, Annette aimed to use eye tracking to test whether a clip derived from an infant-directed DVD designed according to earlier versions of the recommendations above (*Baby Bright*, Abbey Home Media) would hold infant attention, simplify the viewing process and guide infant gaze to the intended concepts, in this case the numbers "2" and "3" better than a matched control clip derived from a different infant-directed DVD covering the same concepts. Drafts of the study account included below were written by Annette and the present authors (TD, PKM and TS) from 2011 to 2015. For this festschrift we welcome the opportunity to help Annette's empirical contribution on the screen time debate finally reach its audience.

Empirical Case Study in I-DTV: Methods

Participants: Sixteen 6-month-olds (mean age=27.4 weeks, SD=2 weeks; 9 girls), 16 12-month-olds (mean age=50 weeks, SD=6 weeks; 6 girls), and 16 adults were tested. The age of the infants studied here (6 and 12-months) is considerably younger than the officially recommended age at which children should be first exposed to screens (AAP suggest 18 months; Radesky et al., 2016) but at which the majority of UK infants are actually receiving daily exposure (Bedford et al., 2016).

Stimuli: Two 130-second clips were extracted from 30-minute, commercially available DVDs, specifically targeted at infants. A scientifically informed clip (video-SI) was extracted from Baby Bright (Abbey Home Media), developed with the deliberate aim of incorporating earlier-listed recommendations. A control clip matched on topic (video-C) was taken from Baby Einstein (Disney). Both clips were on the numbers two and three, and used a combination of brightly colored infant animation, simple photographed sequences, and naturalistic scenes. In each clip, number was represented as the quantity of objects presented on the screen e.g. three cups (Figure 1, shot 15) or three lambs (Figure 2, shot 2). Video-SI used movement across the screen and an accompanying narration to count objects. Video-C showed pairs or triplets of objects, accompanied by a classical music soundtrack. The two clips also differed slightly on other stylistic dimensions such as the amount of naturalistic photographed sequences used (video-C> video-SI), use of classical music (video C > video SI), the pacing (video C > video SI), repetition of visually similar objects (video-SI>video-C), and the use of multimodal counting (video-SI > video-C).

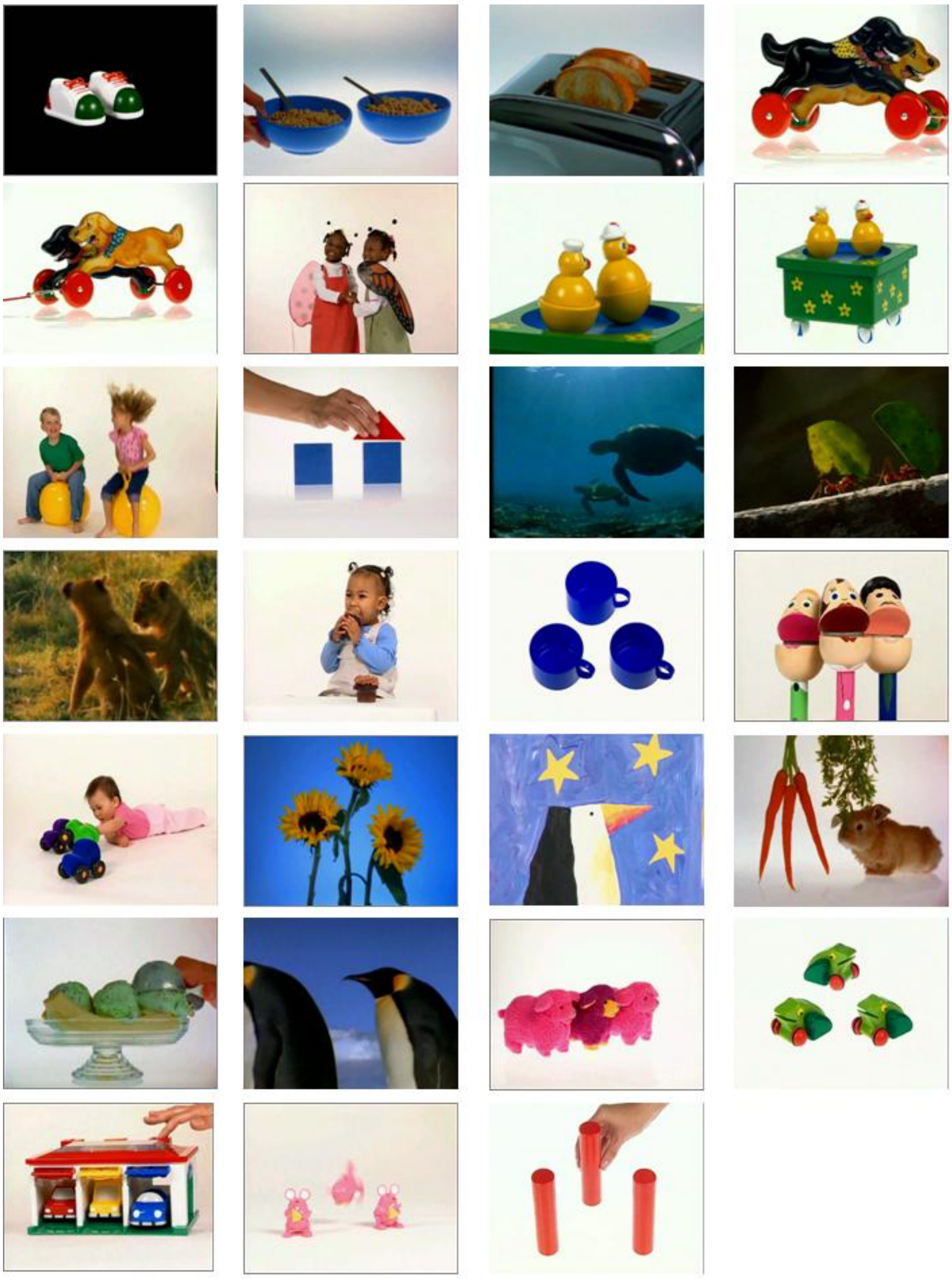

Figure 1: Individual shots presented in video C (control) taken from Baby Einstein (Disney). Shots were presented in order with an accompanying baby friendly music. Shots are intended to represent quantities of two or three as represented by the number of objects on the screen.

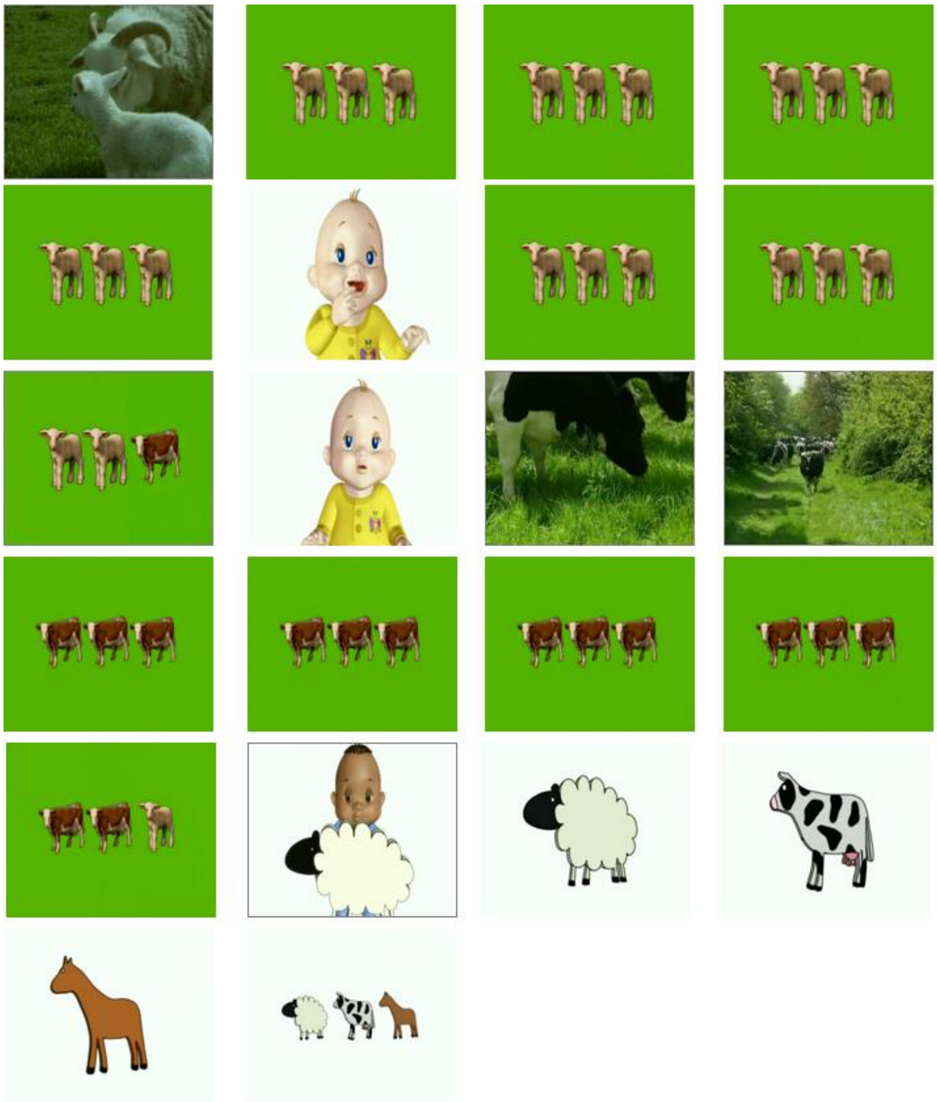

Figure 2: Individual shots presented in video SI (science inspired) taken from Baby Bright (Abbey Home Media). Shots were presented in order with an accompanying narration. Individual lambs and cows (shots against green background) spun across the screen as the narrator counted them.

Procedure: The infants sat on their caregiver's lap in front of a 17-inch LCD screen at a set height and at 60 centimeters distance from the screen. The infants' looking behaviour was registered with a Tobii 1750 Infrared Eye Tracker. Before the experiment was run, an infant-friendly five-point calibration was run. Each clip was presented separated by a short break of 1-3 minutes. Order of clip presentation was counter-balanced across infants. Volume was kept constant. All parents gave informed consent before the study commenced and received detailed debriefing afterwards.

The procedure and setup was identical for the adults except for the fact that they sat on a seat positioned with their heads in a similar position to that of the infants. Adults were instructed to freely view the video clips.

RESULTS

To test how actively engaged infants were with the videos, we first examined how each video held infants interest, by quantifying overall looking time to the screen. We then tested how successful the videos were in directing infants' attention to the intended number of objects, a prerequisite of subsequent comprehension. Finally, we examined how these differences may have been influence by compositional differences between the two videos by looking how well gaze of all infants was attracted successfully to key focal points in the scenes. We did this by comparing infant gaze distributions to that of an adult control group, whose gaze patterns we took as ground-truth for active visual processing of meaningful scene-content.

1) Interest in the videos

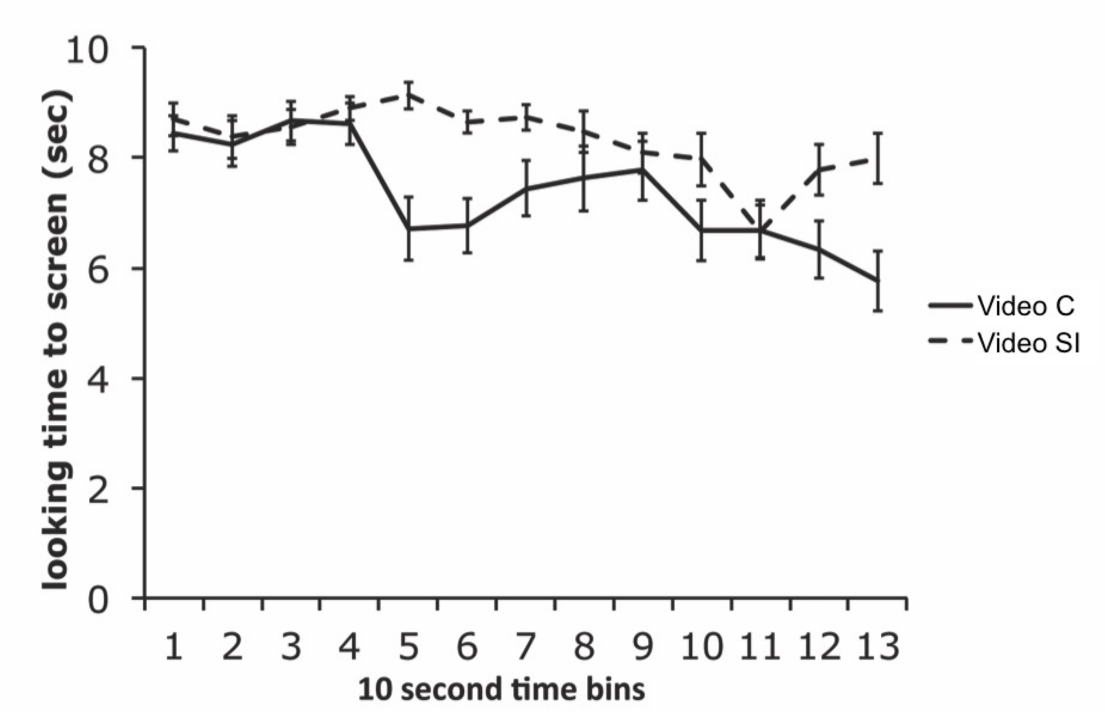

Figure 3 Time spent looking **at the** screen **over time** (split into 10 second bins) for each video (Video SI vs. Video C). Error bars +/-1 SE. Decreasing looking time can be interpreted as an increase in blinking/looking away from the screen.

We can ascertain whether the infants are interested in the videos by using an eye tracking proxy of looking time: whether the eye tracker can detect they are looking at the screen. To examine the loss of interest over time the two 130-second clips were divided into 13 10-second time blocks. Overall, infants looked significantly more at Video SI than at Video C, ($F(1,30) = 7.314$, $p=0.011$; see Figure 3). This pattern held at both 6 and 12 months of age (not shown). Three interesting patterns emerged: 1) infants *started* by looking for an equally long time at both Video SI and Video C (during first quarter of clips: $F(1,30)=0.085$, $p=0.773$); 2) looking decreased more over the course of Video C than over the course of Video SI; 3) by the final quarter infants looked significantly longer at Video SI than Video C ($F(1,30)=12.773$, $p = 0.001$).

2) Looking at Number

To explore what drives differences in infant's looking at the two videos we next explore how successfully infant's gaze was directed to the intended number of objects in each scene.

-

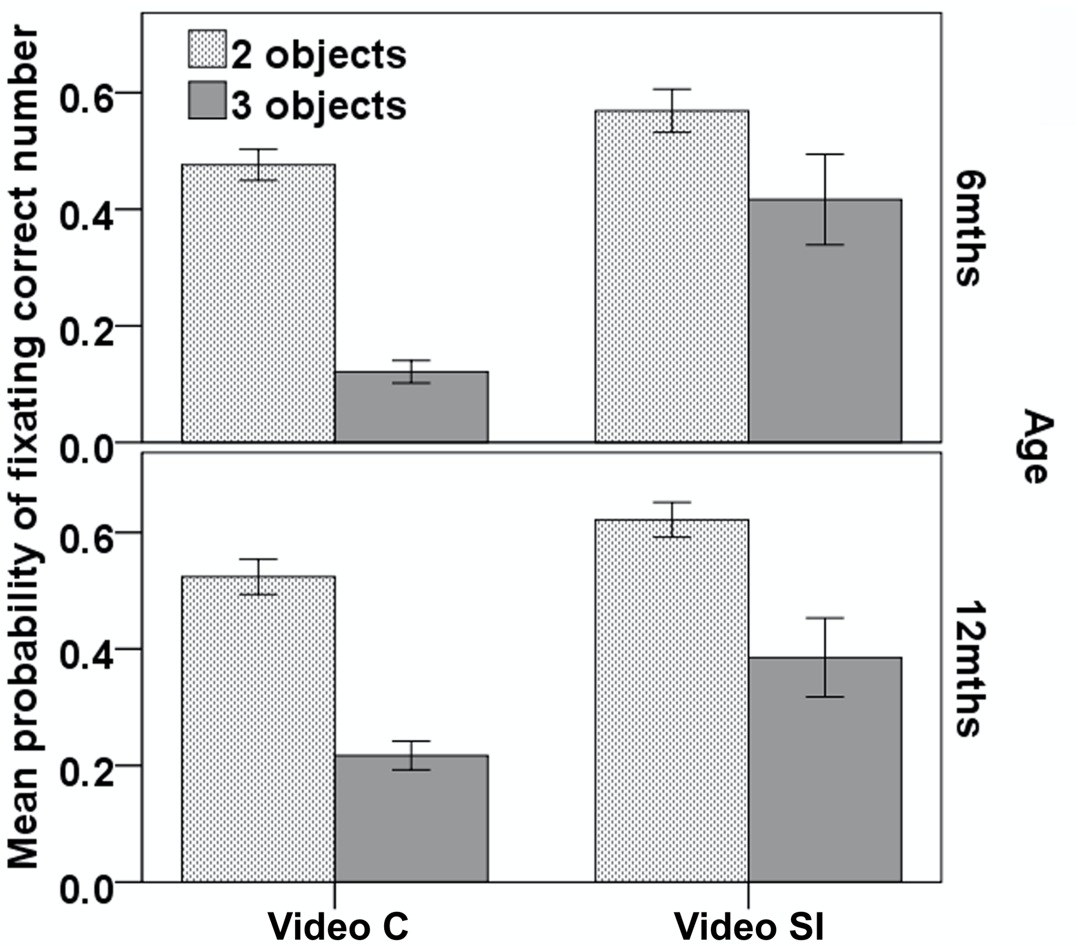

Figure 4: Probability of fixating the intended number of objects during a scene as a function of Age and Video. Error bars +/-1 SE.

Objects in the videos were coded in each frame using dynamic Regions of Interest (dROI). An object counter was incremented every time a new object was looked at within a scene. At the end of each number scene, the total count was compared to the intended number of objects. The intended number was clear to adult observers because the first half of each clip was devoted to 2-object scenes and the second half to 3-object scenes. However, a significant number of shots actually contained 4 or more objects, potentially distracting from intended content (see Figure 1). For example, in the scene of the toddler eating two cupcakes, a match would be recorded if the infant had fixated both of the cupcakes - the intended 2 objects for the scene - but not if they had fixated one cupcake and the toddler.

Figure 4 displays the average probability that the intended number of items was attended within each clip for the 2 age groups. Video SI resulted in a greater probability of fixating the correct number than Video C, $F(1,29)=21,465$, $p<.001$, confirmed at all ages and number with bonferroni-corrected t-tests. There was also a main effect of the number of objects ($F(2,29)=76.87$, $p<.001$) with the intended number of objects more likely to be attended in the 2-object than 3-object scenes, although this effect was mostly driven by Video C. The greater accuracy for 2-object scenes observed overall, may be due to the fact that the two potential targets for attention can be easily identified and shifted back-and-forth between, without competition from other salient items.

3) Gaze Similarity: The analyses above suggest that Video SI held infant attention more, and was more successful in directing gaze to the intended number of objects. These results begin to suggest that Annette's developmentally-informed design considerations may have been successful. Next, we explored how two composition elements, namely visual salience from rapid luminance changes (*flicker*), and the presence of a face in a scene, helped infants actively engage with key plot points of the videos. We tested this by measuring the effects of these two factors on infant and adult *gaze similarity*, an index of how well gaze positions correspond across viewers and scenes. We used this measure to quantify how much infant gaze patterns deviated from adult gaze patterns across each video, for scenes with faces and varying flicker. This approach follows the reasoning that while we cannot be sure which gaze pattern for infants reflects active processing of scene content, the extent to which they look at the same focal points as adults gives a close ground-truth. Gaze of younger infants is more influenced by visual salience than older children and adults, (Frank et al, 2009; Franchak et al., 2016), which may both help and hinder infants in fixating key plot points.

Measuring Gaze Similarity: To computing gaze similarity, we first removed rapid eye movements (saccades) to identify gaze positions (fixations). We then used a modification of a common technique used to measure attentional synchrony, by expressing the observers' gaze on each frame as a probability distribution (for review see Le Meur & Baccino, 2013). How our metric, *gaze similarity* is computed, is described in detail in Loschky, Larson, Magiano & Smith (2015). In brief: for adults, we used a leave-one-out approach: fixations of all but one adult within a 225ms window were replaced with 2D circular fovea-sized Gaussian distributions ($\mu = xy_{fixation}$, σ=1.2° visual angle). For pixels covered by multiple Gaussians, the multiple intensity values were summed. The result is a probabilistic gaze map for each timeframe, with high values reflecting high likelihoods of fixation, and low values reflecting low likelihoods. These distributions were then normalized relative to the mean and SD of all intensities across the entire video, resulting in spatiotemporal map of gaze similarity z-scores across the video. This map was used to compute gaze similarity for the left-out participant, by taking the z-score corresponding to their gaze position in each time window. This leave-one-out procedure was repeated for all adults in the group, until each participant had gaze similarity z-scores for each frame. These z-scores reflect 1) how well the gaze position of an individual adult fits within the group for a given scene (below-average values = poorer fit), 2) how the mean gaze similarity across all adults varies across the video: A z-score close to zero indicates the scene has average gaze similarity compared to the rest of the movie, negative values indicate less than average gaze similarity (i.e., more varied looking across observers), and positive values indicate more gaze similarity.

We next quantified infant gaze similarity relative to adult gaze similarity. For each infant gaze point, the probability that it belongs to the adult gaze distribution for the corresponding video time window is identified by sampling the value at that location from the adult gaze probability distribution (this time leave-one-out is not used as the gaze does not belong to the same distribution so cannot be sampled twice). The resulting raw probabilities are then normalized to the reference adult distribution. Of critical importance, if gaze distributions are identical across age, the average z-scored similarity for infants should overlap with the adult gaze similarity index, expressing more (positive z-score) or less (negative z-score) attentional synchrony at the same moments. However, a lower similarity score in infants does not necessarily mean lower synchrony of gaze – it only means that the infant distribution differs more from the adult distribution than itself. This could reflect weaker clustering of infant gaze around the same screen location (lower attentional synchrony), that infant gaze is tightly clustered but focused on a different screen location, or a combination of the two. Of interest are (1) whether infant gaze similarity is identical to that of adults, suggesting adult-like viewing behaviour, and (2) whether infant gaze synchrony fluctuates in the same way as adult gaze even if the overall score deviates, suggesting that the same aspects of the video drive gaze patterns across age.

Quantifying Flicker Entropy: Flicker is the luminance change from one frame to another that correlates with perceptual features such as motion and optic flow (Mital, Smith, Hill & Henderson 2011). Mital and colleagues (2011) demonstrated that flicker was highly predictive of adult gaze when gaze was highly clustered (i.e. had large synchrony) and hypothesised that this was due to *low entropy* in this feature at those moments (e.g., all points of high flicker are clustered together rather than spread randomly throughout the image) rather than *high entropy*. In other words, higher gaze similarity is predicted for scenes high in flicker but low *Flicker Entropy*, when flicker is spatially clustered. To compute Flicker Entropy, two-dimensional luminance flicker maps (i.e. ignoring the colour channels) were computed for both videos and entropy across these maps was calculated (see Mital, Smith, Hill & Henderson 2011, for details). High entropy values (measured in *bits*) indicate a fairly uniform distribution of flicker across the frame (like an old grainy TV signal), and low values indicate clear peaks/clusters of flicker, e.g. a small light switching on/off.

Identifying Faces: Previous studies have shown that the presence of semantically rich information (i.e. a face) helps guide infants attention to important semantic content in scenes of varied complexity (Frank et al., 2009). To investigate if faces were helping infants direct their gaze to important plot elements in the video, we identified scenes with faces and investigated effects of flicker entropy separately for these scenes.

Comparing Gaze Similarity as function of scene composition across both videos

First we first provide a qualitative analysis of gaze similarity, flicker entropy fluctuations, and use of scenes with faces across the two videos, to understand how and why specific video content causes adult and infant looking patterns to deviate.

Both videos are designed for young infants and share some common visual features including the use of simple bold colours, objects photographed against block colour or white backgrounds and sequentially presented objects or actions. However, the videos differ considerably in the complexity of their images and the pace at which new images are presented. Video C has a higher mean flicker entropy (4.20 bits), smaller variance of flicker entropy (1.934 bits) and faster shot rate (mean shot length is 5.26s) compared to Video SI (1.70 bits, 2.60 bits and 10.54s, respectively). These differences in complexity can be seen in the screenshots included in Figures 5 and 6 and the associated plots of Flicker entropy over time.

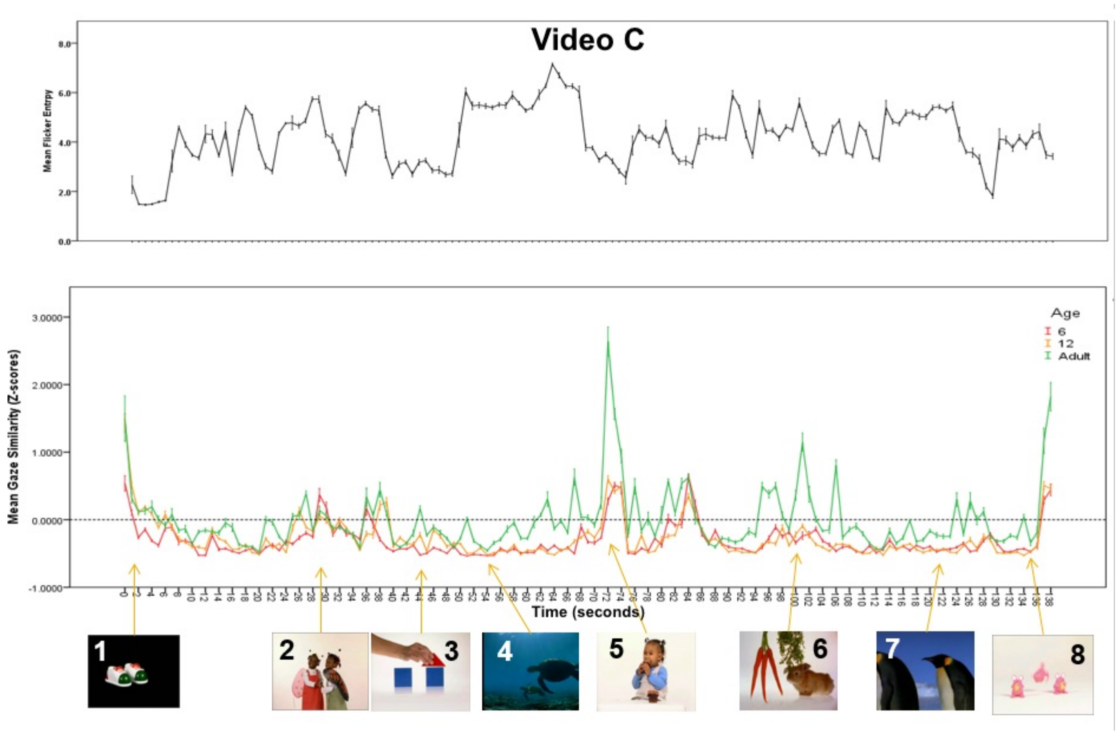

Figure 5: Top graph= Flicker entropy during every second of Video C (measured in bits). Bottom graph= Gaze Similarity (Z-scored relative to Adult gaze distribution) across the three age groups (6 mth=red line, 12 mth= orange line, Adult=green line) over time. Error bars represent +/-1 SE.

Video C uses occasional nature videos shot against a normal background that create a lot of flicker, low contrast between the objects and the background and peaks in flicker entropy (see shots 4 and 7 in Figure 5). These shots do not pose a problem to the adults who are used to parsing such scenes and can easily locate the centres of interest, typically the heads of the animals depicted. For the infants, such low contrast / high entropy images may be difficult to parse (i.e., to locate foreground objects of interest). This results in low gaze similarity as infants attend to the images in different ways. By comparison, the shots composed against a white background, usually depicting a relatively static child or animal interacting with objects (e.g. shot 5=the child eating a muffin or shot 6=the rabbit eating the carrot) therefore produce clear peaks in gaze similarity (bottom chart; figure 5). The subtle movement of the child's hands or the face of the child and animal create points of low flicker entropy relative to the static background that attract attention across all age groups. This interest in faces and hands has previously been confirmed using similar videos in infants (Frank, Vu, & Saxe, 2011). Critically, the motion is restricted to a small area of the screen allowing even the younger infants to identify it as a saccade target, move their eyes to it and fixate it before the shot has changed.

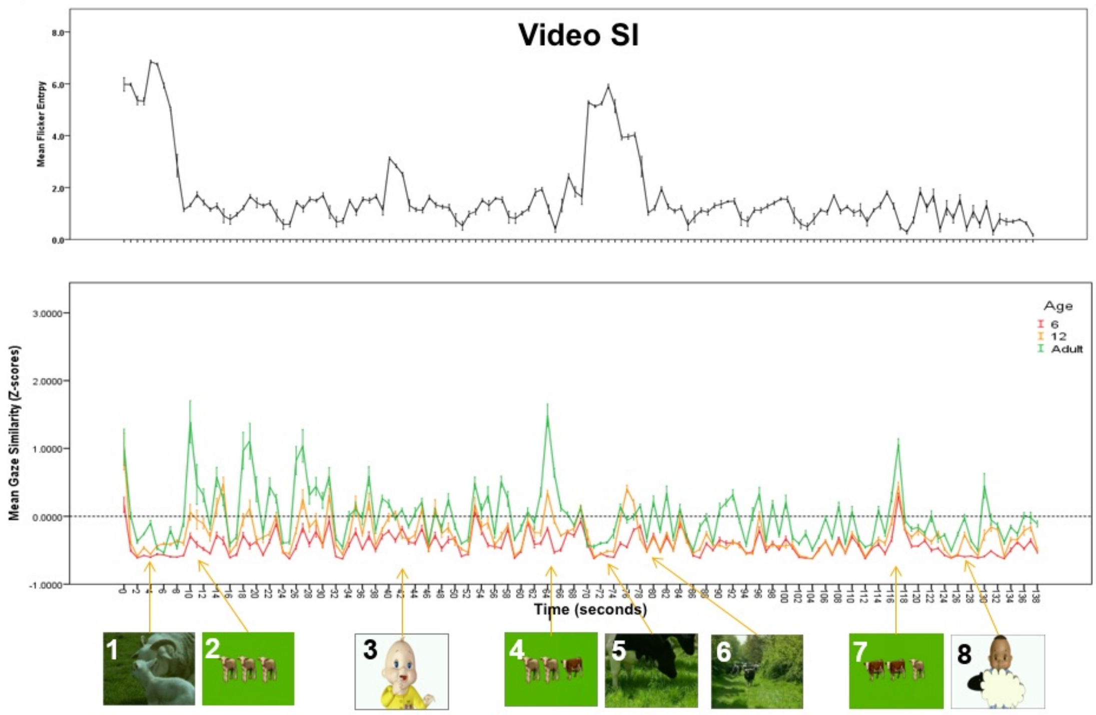

Figure 6: Top graph= Flicker entropy during every second of Video SI (measured in bits). Bottom graph= Gaze Similarity (Z-scored relative to Adult gaze distribution) across the three age groups (6 mth=red line, 12 mth= orange line, Adult=green line) over time. Error bars represent +/-1 SE.

Video SI utilises some similar nature scenes to video C (shots 1,5 and 6 in Figure 6) and these produce similar peaks in Flicker Entropy and low gaze similarity. However, the majority of Video SI's shots involve either a single object or a series of objects presented against a plain background. Instead of using editing to change the centre of interest as in Video C, Video SI uses animations to present objects sequentially by having them either fly on to the screen or suddenly grow from nothing. These presentations match the audio narration: “Look, a baby sheep! A baby sheep is called a lamb [shot 1]. Lamb. Lamb. Lamb [shot 2]. Baa. Baa. Baa. [first repetition of shot 2] One. Two. Three. [second repetition of shot 2]”. Each repetition of “Lamb”, “Baa” or the number corresponds with the appearance of a lamb flying on to the screen (shot 2, Figure 6). These audiovisual correspondences are intended to attract the infant’s eyes to the object being named and reinforce the naming and subsequent counting through repetition. The sequential presentation results in peaks and troughs of gaze clustering at all ages (e.g. Figure 6, 10 to 32 seconds) as the sudden appearance of an object on screen attracts all attention then viewers return to exploring the frame once the object has stopped. However, the reduced gaze similarity trace indicates that the 6mth and 12mth olds do not respond as quickly to the sudden appearance of the object or pursue it as accurately as the adults as it moves across the screen.

Gaze Similarity Differences Across Age

The descriptive age differences visible in gaze similarity in Figures 5 and 6 are confirmed by a repeated-measures ANOVA (Age: F(1,45)=43.158, p=.001, $\eta_p^2{}_p$ =.657). There is no main effect of Video (F(1,45)=1.674, p=.202, $\eta_p^2\eta_p^2$ =.036) or interaction (F(2,45)=1.443, p=.247, $\eta_p^2\eta_p^2$ =.06), suggesting, on average both videos create the same variance in attentional synchrony. Collapsing across Videos, the main effect of Age can be attributed to significantly lower gaze similarity at 6 months (mean=-0.363, sd=.06) than for Adults (mean=.00, sd=.155; p=.001) and a marginal difference relative to 12 months (mean=-0.273, sd=.110). The 12-month-old gaze also showed significantly less similarity to the adults (p<.001).

Gaze Similarity x Flicker Entropy and Faces.

To test for age differencesis in how flicker entropy and the presence of the face in the scene affects gaze, video frames with and without face were binned by Flicker Entropy (0bits-low to 6bits-high entropy) for subsequent analyses. The two videos were analysed separately as the simpler composition of Video SI meant that we could not identify shots containing a face and high flicker entropy like in Video C.

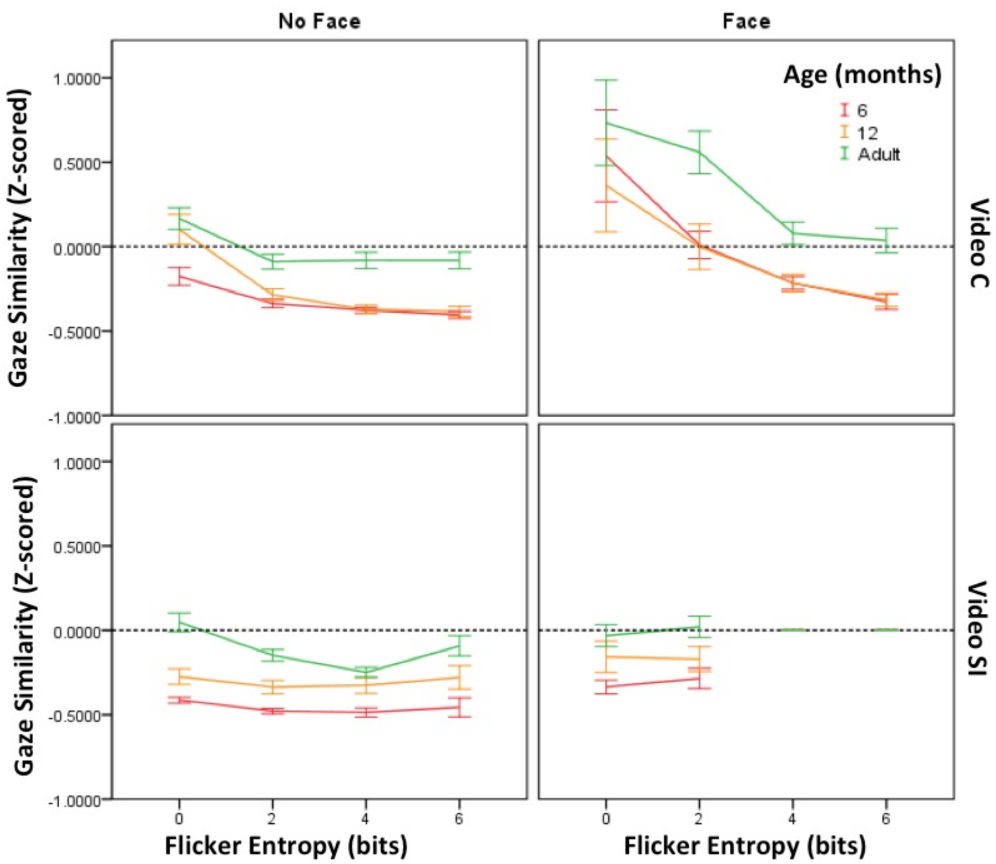

Figure 7: Gaze similarity (expressed as a Z-score relative to the within-video adult distribution) across the three age groups (6 mth=red line, 12 mth= orange line, Adult=green line) and mean flicker entropy for a frame (binned into 2 bit bins; x-axis) split into shots containing at least one human face ('Face') and shots not containing faces ('no face'). Error bars represent +/-2 SE.

Video C: A repeated-measures ANOVA on gaze similarity within each age group (6mth, 12mth, & Adults) with factors Flicker Entropy (0, 2, 4, 6 bit bins) and whether the shot contains at least one human face (face vs. no face) within Video C revealed a main effect of Flicker Entropy (F(3,135)=37.782, p=.001, $\eta_p^2\eta_p^2$=.456), a main effect of Face (F(1,45)=41.645, p=.001, $\eta_p^2\eta_p^2$=.481), an effect of Age (F(2,45)=9.362, p=.001, $\eta_p^2\eta_p^2$=.294) and Face x Entropy interaction (F(3,135)=6.245, p=.001, $\eta_p^2\eta_p^2$=.122). The effect of Flicker Entropy can clearly be seen in Figure 7. As Flicker Entropy decreases (i.e. flicker becomes more concentrated) gaze similarity increases from -.246 (SE=.024) when entropy is large (6 bits) to +.287 (0 bits; SE=.081;

p=.001). This pattern is similar across all three age groups suggesting a universal influence of image features on gaze behaviour.

Splitting the data into shots with and without a human face, the main effects of Age (F(1,45)=19.089, p=.001, $\eta_p^2\eta_p^2$=.459) and Flicker Entropy (F(1.44,64.975)=58.096, p=.001, $\eta_p^2\eta_p^2$=.564; Greenhouse Geisser corrected) are very large within the No Face shots suggesting that both age and flicker entropy strongly influence gaze. There is also an interaction between Age and Flicker Entropy (F(2.88,64.975)=4.262, p=.009, $\eta_p^2\eta_p^2$=.159) due to the difference in gaze similarity between 12mths and Adults disappearing in the lowest Flicker Entropy bin (0 bits) and 12mths displaying greater gaze similarity than the 6mths (t(30)=-2.694, p=.011). This indicates that by 12 months, infants can display adult-like gaze behaviour if non-social video sequences are designed to guide their eyes to the centre of interest using low-level visual features (e.g. low flicker entropy; Wass & Smith, 2015).

When human faces are present in Video C, the main effects of Age (F(2,45)=5.182, p=.009, $\eta_p^2\eta_p^2$=.187) and Flicker Entropy (F(1.26, 56.73)=18.914, p=.001, $\eta_p^2\eta_p^2$=.296; Greenhouse Geisser corrected) remain but the interaction disappears (F<1) as all ages show the same increase in gaze similarity for the lowest Flicker Entropy bin (see Figure 7, top right, 0 bits). In the presence of a face, the difference between 6 and 12mth Gaze Similarity disappears for all entropy bins and in the lowest bin (0 bits) even the difference between 6mths, 12mths and Adults disappears (all *t*s<1). This suggests that when social scenes are composed with only one area of visual change at a time even the youngest infants are able to exhibit adult-like viewing behaviour.

Video SI: For confirmation of the influence of Flicker Entropy and Faces on gaze we can now turn to Video SI. Given the simpler composition of Video SI there were no shots containing faces with a high degree of Flicker Entropy so the analyses focussed on only the lowest two entropy bins (0 and 2 bits). A repeated-measures ANOVA reveals main effects of Age (F(2,45)=15.764, p=.001, $\eta_p^2\eta_p^2$=.412), Faces (F(1,45)=17.276, p=.001, $\eta_p^2\eta_p^2$=.277), and Flicker Entropy (F(1,45)=4.184, p=.047, $\eta_p^2\eta_p^2$=.085). The effect of Face is similar to Video C in that shots containing faces produce greater gaze similarity than without faces.

The effect of Flicker Entropy is considerably weaker than in Video C probably due to the effect being measured over fewer levels of entropy. There is also an interaction between Face and Flicker Entropy (F(1,45)=12.143, p=.001, $\eta_p^2\eta_p^2$=.212) due to the Flicker Entropy effect being absent in the Face shots (F<1) but present in the No Face shots (the other two levels of entropy can be added to this analysis; F(1.72,77.27)=10.566, p=.001, $\eta_p^2\eta_p^2$=.190). Within the No Face shots there is also an interaction between Flicker Entropy and Age (F(3.43,77.269)=3.021, p=.029, $\eta_p^2\eta_p^2$=.118) but unlike in Video C this is not because the Age difference at 12 months old disappears at the lowest flicker bin. Instead, the interaction is caused by the Flicker Entropy effect which is only present in Adults, F(1.95,29.23)=13.905, p=.001, $\eta_p^2\eta_p^2$=.481) and not 6 mths or 12 mths (both Fs<=1). This may be because, unlike in Video C which has much more varied shot content there are only two types of non-face scene, all the high entropy shots are natural scenes which the infants are unable to parse (e.g. shots 1, 5, and 6, Figure 6) and all of the low entropy scenes depict a series of animal images flying into shot sequentially (e.g. shots 2, 4, and 7, Figure 6). The infant gaze is unable to exactly match the location of the adult gaze during these shots as their anticipation and pursuit of moving targets is sluggish. However, as was evident in the number ROI analysis (Figure 4), this sequential presentation helps the infants arrive at the correct number of objects even if they may get there at a later time than the adults (leading to low gaze similarity scores).

In sum, our gaze similarity analysis reveals that infants visually engage with infant-directed video in different ways than adults do, and highlights the importance of using salience in the scenes in a manner that helps infants process meaningful content. Specifically, in visually complex scenes, a large amount of image change (i.e. Flicker Entropy) caused by object or camera motion may prevent infant gaze from locating key plot elements in the scene (e.g., animals moving though naturalistic scenes). The inclusion of faces, which are highly salient to infants and adults alike, can help mitigate this. Differences in looking patterns can also occur in low-entropy scenes with clear points of saliency, when the more sluggish developing visual system is engaged in tracking dynamic content.

DISCUSSION

The empirical case study presented here suggests that by carefully designing I-DTV to simplify the process of parsing the visual *signal* -whether it be a particular number of objects or a human face- from the background visual *noise,* the gaze behavior of even 6 month-old infants can resemble that of adults. By comparing a commercially available I-DTV clip, Baby Einstein (Video C) to a clip depicting the same concept but using scientifically-informed design, Baby Bright (Video SI), Annette aimed to demonstrate the importance of empirical tests in understanding how a scientifically-informed videos may can lead to better engagement, more accurate attending to the number of objects and a simpler active viewing process (indexed by the designers. The results above show that whilst the first two predictions are supported by our results, the greater variation in shot content and compositional style used in Baby Einstein lead to mixed results. On the one hand, scenes with high flicker entropy overwhelmed infant gaze, reducing their ability to identify key content. On the other hand, it created moments during which low flicker entropy and the presence of faces within complex scenes guided infant gaze to the same point as adult gaze.

These findings confirm our earlier predictions derived from feature-analysis of I-DTV (Wass & Smith, 2015) and prior eye tracking studies (Kirkorian, et. al, 2012; Franchak et al., 2015, Frank et al, 2009). By comparison, the much simpler sequential presentations used in Baby Bright were successful in representing the concept of number, encouraging infants to actively seek out objects and maintain their interest in the screen. The infants looked at the correct number of objects and matched the gaze pattern of the adults (see the matching peaks and troughs of gaze similarity in Figure 6). However, the high speed of the moving objects caused the sluggish infant gaze to lag behind the objects and the adult gaze, giving rise to lower gaze synchrony overall.

This highlights the importance of striking a balance between Annette's recommendation 3 (avoid passive fixation) and 7 (slow down!) where movement across the screen might be stimulating for the developing infant's visual system, but moving too fast may prevent infants from processing key plot elements. The most notable dissociation between infant and adult viewing was for nature shots of animals included in both videos. Anecdotally, during testing parents often commented on how much they liked the natural scenes and believed they would be educational for their infants. However, the reduced looking time overall and the lower gaze similarity results indicate that the high flicker entropy in these scenes meant that infants found them very hard to watch (Figures 5&6). This is a clear example of how adult designers must inhibit their own preferences when designing for a developmentally immature visual system[1].

These results confirm the importance of *content* and stimulus design for understanding how the developing mind responds to the stimulus and how it may encourage learning. These results also demonstrate the importance of appreciating *how* infants engage with screen media when trying to understand the long-term consequences of regular screen time on infant neurocognitive and behavioural development. In order to combat propaganda against screen time and non-evidence-based guidelines detailed longitudinal studies of associations between different types of screen use and infant development are required. This work fits in with the emerging perspective in the developmental cognitive neuroscience of screen use, that the *content* and *context* of screen media use is important for understanding how we can maximise the benefits of screen time for supporting learning, education, and social connection in childhood and adolescence and avoiding negative impacts around displacement of other activities such as sleep and physical exercise (Courage, 2017).

CONCLUSIONS AND LEGACY

Annette believed that that "the dynamics of screen exposure may be an important supplementary stimulation to the young infant's visuo-cognitive system, alongside images in books and mobiles rotating in central vision, provided that the content and design of videos are based on infant scientific research and encourage active screen exploration, rather than passive central fixation" (Annette Karmiloff-Smith unpublished manuscript, 23rd January, 2012). As such, she cautioned against using screen time as baby-sitter to replace social interaction, or displace other developmentally critical activities such as physical exploration, sleep, and feeding. Our findings endorse Annette Karmiloff-Smith's view that rather than the "mindless" passive viewing assumed by some critics of I-DTV (e.g. Greenfield, 2015), screen exposure can be designed to encourage saccadic exploration, rapid visuo-spatial

[1] Annette would wish us to document that she actually advised against using such natural video scenes in Baby Bright precisely because she predicted these difficulties!

orienting, anticipation, figure/ground separation, and attention to number of intended targets making the infant a highly active viewer.

Annette's pioneering perspective on how screens may best be used to support infant and child development was invaluable when she co-founded the Toddler Attentional Behaviours and Learning with Touchscreens (TABLET) Project with author TS and Dr. Rachael Bedford (King's College London). TABLET is on-going project aiming to test whether infant exposure to touchscreen devices at 12 months-of-age is associated with long-term developmental differences at 18 and 42 months-of-age in domains including attention, temperament, developmental milestones, language, executive function and sleep. The intention is to triangulate effects using parent-report questionnaires in a large online sample (N>700), and behavioural, questionnaire, EEG and eye tracking measures in an intensive lab sample (N=60). Annette's contribution to the TABLET project were critical for guiding its future direction (see her APS interview on the project[2]) and she contributed to the first findings establishing associations between infant touchscreen use and sleep problems (Cheung et al., 2017) and earlier fine-motor development (Bedford et al, 2016).

The TABLET project is testament to how Annette's rigorous experimental approach to studying developmental cognitive neuroscience has broad implications for many areas of development and their societal implications such as the pressing issue of childhood screen time, and the need to understand the mechanisms of visual processing to understand visual impairments and how to best treat them.

Annette many not have had the opportunity to see her empirical contributions on this topic published during her lifetime but her influence will be long-lasting for those of us who had the opportunity to collaborate with and be inspired by her.

In closing, we can think of no better way to end our exploration of Annette's work on screen time than with her own words:

"*In sum, screen exposure for young children should be far more than a display of coloured patterns and music to mesmerize babies; they should be a scientifically*

---

2 https://www.psychologicalscience.org/publications/observer/obsonline/toddlers-and-touchscreens-a-science-in-development.html

designed effort to stimulate babies. We live in a media-saturated world. It is far better that parents know how to choose the right television or DVD programmes for their children than to make them ashamed at even thinking of ever using screen exposure." (Karmiloff-Smith, 2012)

REFERENCES

Anderson, Daniel R. and Katherine G. Hanson. (2010) From Blooming, Buzzing Confusion to Media Literacy: The Early Development of Television Viewing. *Developmental Review*, 30(2), 2010: 239–255.

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, *48*(5), 505-522.

Baker, T. J., Tse, J., Gerhardstein, P., & Adler, S. A. (2008). Contour integration by 6-month-old infants: Discrimination of distinct contour shapes. *Vision Research.* 48(1), 136-148.

Barr, R., Muentener, P., & Garcia, A. (2007). Age-related changes in deferred imitation from television by 6-to 18-month-olds. *Developmental Science*, *10*(6), 910-921.

Bedford, R., Saez de Urabain, I. R., Cheung, C. H., Karmiloff-Smith, A. & Smith, T. J. (2016) Toddlers' Fine Motor Milestone Achievement Is Associated with Early Touchscreen Scrolling. *Front Psychol* 7, 1108, doi:10.3389/fpsyg.2016.01108.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision ResearchR*, *51*(13), 1588-1609.

Bremner, J. G., & Wachs, T. D. (Eds.). (2010). *The Wiley-Blackwell Handbook of Infant Development, Basic Research* (Vol. 1). John Wiley & Sons.

Cheung, Celeste and Bedford, R. and Saez De Urabain, Irati R. and Karmiloff-Smith, Annette and Smith, Tim J. (2017) Daily touchscreen use in infants and toddlers is associated with reduced sleep and delayed sleep onset. *Scientific Reports* 7 , p. 46104. ISSN 2045-2322.

Christakis, D. A., & Zimmerman, F. J. (2006). Media as a public health issue. *Archives of Pediatrics & Adolescent MedicinePAM*, *160*(4), 445-446.

Brown, A. (2011) Media Use by Children Younger Than 2 Years: Policy Statement of the AAP, *Pediatrics;* 128(5): 1-6.

Courage, M. L. (2017). Screen media and the youngest viewers: Implications for attention and learning. In *Cognitive Development in Digital Contexts* (pp. 3-28). Academic Press.

Courage, M. L., & Howe, M. L. (2010). To watch or not to watch: Infants and toddlers in a brave new electronic world. *Developmental Review*, *30*(2), 101-115.

Cristia, A. & Seidl, A. Parental Reports on Touch Screen Use in Early Childhood. *PLoS One* **10**, e0128338, doi:10.1371/journal.pone.0128338 (2015).

Davies S.C., Atherton F., Calderwood C., McBride M. (2019) United Kingdom Chief Medical Officers' commentary on 'Screen-based activities and children and young people's mental health and psychosocial wellbeing: a systematic map of reviews'. Department of Health and Social Care

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child DevelopmentD*, *62*(4), 736-752.

Dickson K, Richardson M,, Kwan I , MacDowall W, Burchett H , Stansfield C, Brunton G, Sutcliffe K1 Thomas J (2018*) Screen-based activities and children and young people's mental health: A Systematic Map of Reviews*, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Ferguson, Charles A. (1964) Baby Talk in Six Languages. *American Anthropologist*, 66(2/6),: 103–114.

Fiser, J. & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822-15826.

Franchak, John M., David J. Heeger, Uri Hasson, and Karen E. Adolph. (2016) Free Viewing Gaze Behavior in Infants and Adults. *Infancy*, 21(3), 262–287.

Frank, Michael C., Edward Vul, and Rebecca Saxe. (2012) Measuring the Development of Social Attention Using Free ‐ Viewing. *Infancy*, 17(4), 355–375.

Greenfield, S. (2015). *Mind Change: How digital technologies are leaving their mark on our brains*. Random House.

Hill, J., Inder T, Neil J, Dierker D, Harwell J, Van Essen D (2010). Similar patterns of cortical expansion during human development and evolution. *Proc Natl Acad Sci U S A*. 107 (29): 13135-40.

Karmiloff-Smith, Annette and House, R. (2015) ICT - for & against: Toddlers, TV and touchscreens. *Nursery World*, ISSN 0029-6422.

Karmiloff, K. and Karmiloff-Smith, Annette (2011) A unique child: cognitive development: tv times. *Nursery World* , ISSN 0029-6422.

Karmiloff, K. & Karmiloff-Smith, A. (2011-12). Series of 6 articles on Infant TV and DVDs; Growing Up Multilingual; Gender Differences; Special Talents in Early Development; Handedness in Humans, Apes & Prehistoric Man; Infant Predictors of Reading Abilities. *Nursery World*.

Karmiloff-Smith, A. (1994). *Baby It's You: A unique insight into the first three years of the developing baby.* London: Ebury Press, Random House.

Karmiloff-Smith, A. (2012). 'TV is bad for children' – less emotion, more science please!, In Philip Adey and Justin Dillon, *Bad Education,* Open University Press, 231-244

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*(40), 9867-9874.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research MethodsRM*, *45*(1), 251-266.

Loschky, L.C. and Larson, A.M. and Magliano, J.P. and Smith, Tim J. (2015) What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS One* 10 (11), e0142474. ISSN 1932-6203.

Mital, P.K. and Smith, Tim J. and Hill, R.L. and Henderson, J.M. (2011) Clustering of gaze during dynamic scene viewing is predicted by motion.*Cognitive Computation* 3 (1), pp. 5-24. ISSN 1866-9956.

Ofcom (2018, February 21) *Children and Parents: Media Use and Attitudes Report.* Retrieved from https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2018

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child DevelopmentD*, 80(3), 674–685.

Pempek, Tiffany A., Heather L. Kirkorian, John E. Richards, Daniel R. Anderson, Anne F. Lund, and Michael Stevens. (2010) Video Comprehensibility and Attention in Very Young Children. *Developmental Psychology*, 46(5), 2010: 1283.

Radesky, J., Christakis, D., Hill, D., Ameenuddin, N., Chassiakos, Y., Cross, C., ... & Moreno, M. A. (2016). Media and young minds. *Pediatrics*, *138*(5).

Saez De Urabain, I. R. (2015) *Investigating the mechanisms underlying fixation durations during the first year of life: a computational account* , Unpublished Doctoral Thesis, Birkbeck, University of London. http://bbktheses.da.ulcc.ac.uk/148/1/cp_thesis_irati.pdf

Saez de Urabain, Irati R. and Nuthmann, A. and Johnson, Mark H. and Smith, Tim J. (2017) Disentangling the mechanisms underlying infant fixation durations in scene perception: a computational account. *Vision Research* 134 , pp. 43-59. ISSN 0042-6989.

Sigman, A. (2012). Time for a view on screen time. *Archives of disease in childhood*.

Smith, Tim J. (2013) Watching you watch movies: using eye tracking to inform film theory. In: Shimamura, A (ed.) *Psychocinematics: Exploring Cognition at the Movies.* New York, U.S.: Oxford University Press, pp. 165-191. ISBN 9780199862139.

Stiglic N, Viner RM. (2019) The effects of screentime on the health and wellbeing of children and adolescents: a systematic review of reviews. *BMJ Open 2019*.

Strasburger, V. C. H., M. J. Children, (2013) Adolescents, and the Media: Policy Statement of the AAP. *Pediatrics* **132**, 958-961.

Troseth, G. L. (2003). TV guide: Two-year-old children learn to use video as a source of information. *Developmental PsychologyP*, *39*(1), 140.

Valkenburg, P. M., & Vroone, M. (2004). Developmental changes in infants' and toddlers' attention to television entertainment. *Communication Research*, 31, 288-31

Viner, R., Davie, M & Firth, A. (2019) The health impacts of screen time: a guide for clinicians and parent, *Royal College of Pediatrics and Child Health*

Wartella, E., Richert, R.A. & Robb, M.B. (2010). Babies, television and videos: How did we get here? *Developmental Review,* **30***,* 116-127.

Wass, S., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current BiologyB*, *21*(18), 1543-1547.

Wass, S.V. and Smith, Tim J. (2015) Visual motherese? Signal-to-noise ratios in toddler-directed television. *Developmental Science,* 18(1), 24-37.

WHO guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.

Zimmerman, Frederick J., Dimitri A. Christakis, and Andrew N. Meltzoff. "Television and DVD/Video Viewing in Children Younger than 2 Years." *Archives of Pediatrics & Adolescent Medicine*, 161(5), 2007: 473–479.