How Infants Perceive Animated Films

Sermin Ildirar Kirbas & Tim J. Smith

Abstract

Today, many infants begin consistently viewing videos at 4 to 9 months of age. Due to their reduced mobility and linguistic immaturity younger infants are good watchers, spending a lot of time sitting and watching the actions and (also emotional) reactions of both real and televised people as well as animated characters. Since babies can perceive the similarity between a 2-dimensional image and the real 3-dimensional entity that is depicted, they respond to the video image of another person with smiles and increased activity, much as they would to the actual person. Furthermore, emotional reaction of a televised person can influence their behaviour. Infant attention to films as to natural scenes begins by being stimulus-driven and progresses to top-down control as the child matures cognitively and acquires general world knowledge. The producers of infant-directed animations however use low-level visual features to guide infants' attention to semantic information which might explain infants' preference for them. In this chapter, we will discuss the developmental foundations of (animated) film cognition, focusing mainly on the perception of emotional cues based on recent empirical findings.

Introduction

Due to their reduced mobility and linguistic immaturity, infants are good watchers spending a lot of time sitting and watching the actions and emotional reactions of other people. In this way they learn how to interpret and predict others' behaviours and to relate this to their own behaviour. Prior to the Twentieth century, the source of such exposure was entirely limited to observing real people performing real actions in the infant's immediate surroundings or dramatizations of human behaviours in a play. However, after the Lumière brothers premiered their moving images to the public in 1895 the physical proximity constraints on watching human behaviours disappeared and audiences of all ages could be exposed to an infinite repertoire of human behaviours projected on to the two-dimensional (2D) surface of the movie screen. Initially, infant exposure to moving images would have been infrequent but as television brought the images into our homes and then mobile technologies brought them into our hands the opportunity for learning from screens became pervasive. Using data collected during the early 1990s, Certain and Kahn (2002) reported that 17% of 0- to 1-year-olds and 48% of 1- to 2-year-olds watch television. More recently, researchers have reported that by 3 months of age, about 40% of children regularly watched television, DVDs, or videos. By 24 months, this proportion rose to 90%. The median age at which regular media exposure was introduced was 9 months. Among those who watched, the average viewing time per day rose from 1 hour per day for children younger than 12 months to more than 1.5 hours per day by 24 months (Zimmerman, Christakis and Meltzoff 2007). Prior to the advent of infant-directed media, children under 2 years paid little attention to television, beginning regular TV viewing at about 30 months. With the introduction of *Teletubbies* in 1991 and *Baby Einstein* in 1996, however, modern programs and videos now target babies who have not even spoken their first words or taken their first steps. The revenue of *Baby Einstein* grew from $1 million in 1998 to $200 million in 2005 (Bronson and Merryman 2006). BabyTV, which is a television channel targeting infants and toddlers, is launched in 2003 and is distributed in over 100 countries, broadcasting in 18 languages (as of 2013), and it is just one of the many other television channels targeting babies. And unlike adult-directed TV most of their content is animated. The precise motivation for the prominence of animation in kid's TV is not fully known but one potential key component may be that animated films give an animator absolute control over the visual content of each shot in a way that might be financially impossible in live-action. Such control allows the animator to shape the flow of visual storytelling across 'shots'. The colors, light, movement, dialogue and music can be tailored to direct immature gaze to semantic features. Characters, who are at the same age as their viewers (which would not easy in live action) can be designed in their entirety and their actions and emotions scripted to create age and educationally-appropriate stories. And most importantly, fascination, humor and entertainment can be foregrounded above a strict adherence to reality so that the young viewers enjoy the experience and come back for more.

However, this great potential for kid's TV to captivate young audiences through the power of animation fuels a major controversy. On the one hand, producers market infant-directed programs as being educationally or developmentally beneficial (Christakis and Garrison 2005); on the other hand, the American Academy of Paediatrics recommends that children younger than 18 months of age should not be exposed to electronic screens (unless for the purpose of videochat; AAP 2016). The AAP is concerned that media exposure may contribute to language delays and potential attention problems in young children citing a correlational study by Zimmerman, Christakis, and Meltzoff (2007). However, another study reanalysing the Zimmerman et al.'s (2007) data set, did not find strong inferences about a connection between exposure to media and language development in young children (Ferguson and Donnellan 2013). With debates continuing to rage as to whether screen exposure is negative or not what babies understand from what they watch and what factors play a role in this understanding are fundamental questions trying to be answered by conducting experiments on babies.

In this chapter, we will summarise the empirical studies used to test how babies perceive visual scenes and how these skills inform our understanding of film and TV perception. Discussion of the perception of live-action content will be interspersed with animation as kid's TV is unique in the flexible way in which the images are created, traditionally promoting animation above live-action (the inverse of adult-directed TV) and also intermixing a cornucopia of other techniques including cell animation, cut-outs, CGI, stop-motion, collage and puppetry. We will cover the perceptual and cognitive skills the babies need to have or develop in order to make sense of their visual environments and moving images. We will also discuss how infants perceive the similarity between a 2-dimensional image and the real 3-dimensional entity and what are the difficulties they face in understanding the representational nature of 2D images, if they can learn from videos, what strategies contemporary kids programme makers use to help their little viewers make sense of the content and how these strategies are related to the early days of cinema. But to begin with we will briefly discuss the unique challenges of running studies on perception with infants.

Testing Babies

Babies are very hard to conduct experiments on. They can't understand instructions, they can't press buttons, they get bored, they cry and they fall asleep. On the other hand, their looking is a major gateway to the their mind before language develops since their visual abilities are quite sophisticated: A baby pays attention and looks at anything that is new and interesting, which is the root of preferential looking tasks frequently used in developmental studies. *Preferential looking* tasks present two stimuli to a baby and the length of time the infant looks at each is measured. The longest amount of time can be inferred to be the one that the baby finds the most interesting. *Habituation* is a technique developed from the ideas of preferential looking. In such experiments, babies are shown a stimulus until they are bored of it, and look at it no longer. Then this habituated stimulus is presented alongside a test stimulus. The preferential looking technique is then applied. If the baby now looks at the novel stimulus more than the habituated one, it is concluded that it can understand the difference between them. *Violation of Expectation*, which is another method used to illustrate different aspects of infant cognition, means pretty much what it says: When babies' expectation is violated then infants look longer at the stimuli.

For decades, researchers have studied infant looking by relying on human observers who were coding the duration and direction of looking. Since the beginning of 2000s, there has been an explosion of research using *eye tracking* with infants (Smith and Saez De Urabain 2017). Eye tracking records the movements of the eye (typically using high-speed infrared cameras and pupil image tracking) relative to a calibrated field of view (usually a screen) allowing researchers to see exactly where the participant's centre of gaze was directed and infer from this which parts of a visual scene were preferentially processed. Researchers have examined infants' memory processes, perceptual learning, understanding of joint attention, face processing, and many other topics by using eye tracking procedures. Eye-tracking procedures have been successfully implemented in infants as young as 3 months. Since eye trackers provide detailed information about infants' point-of-gaze from moment to moment, it is possible to ask how infants' distribution of looking over the area of stimulus or over time varies by age or stimulus type, which is great for studying perception of moving images by very young children. As well as devising innovative techniques for quantifying infant behaviour, developmental scientists have also needed to find ways to make their experimental stimuli as intrinsically motivating so that babies continue watching. Interestingly, from the perspective of this book, a lot of the techniques they learned to employ have been directly borrowed from kid's TV and specifically animation. The use of bold color schemes, caricatures, simple audiovisual events (e.g., "Boing!"), puppets for live-action studies, and short/simple video sequences have all been used as proxies for studying how cognition develops in real-world scenes. As such, the studies we will review below often endeavour to further our understanding of real-world cognition through the use of animation and can therefore also provide insight into the perceptual foundations required for film cognition.

2D, or not 2D: Can babies learn from screens?

The first potential barrier between film content and infant comprehension is the two-dimensional (2D) nature of the stimulus. The absence of full 3D depth cues in the image as well as other optical and physical aberrations, such as impossible object sizes or framing (e.g., an upside down extreme long-shot showing tiny people hanging from the top of the screen) may cause problems for infant comprehension of the content. You may also consider these problems exacerbated by the prominent use of 2D animation in kid's TV, in which depth and perspective are often absent or creatively flaunted. However, studies have shown that infants are able to perceive the similarity between a 2D image and the real three-dimensional (3D) entity that is depicted. For example, 2- to 5-month-old infants respond to the video image of another person with smiles and increased activity, much as they would to the actual person (Muir, Hains, Cao and D'Entremont 1996). By 6 months, infants can recognize video images of their parents and associate them with a familiar label like 'Mama' and 'Papa' (Tincoff and Jusczyk 1999). Other research shows that infants are capable of discriminating video images of people and objects from their real counterparts. Four- to 6-month-old infants smile more at a real person than at a live video view of that person even though the person is equally responsive to the baby in both cases (Hains and Muir 1996). Infants can also acquire new information from screen media. For example, 12-to-18 month-olds play with toys that they see on television more than they do with novel toys (McCall et al. 1977); 14-month olds can duplicate actions depicted on television, even when they are presented by a stranger using an unfamiliar object and even there is a 24-hour delay between watching the action and having access to the real 3D object (Meltzoff 1988); 18- but not 14-month olds show a visual preference for a novel toy after televised model engages infants in joint reference during familiarization with another toy (Cleveland and Striano 2008). Infants can also predict televised models' actions as they do with real people. Twelve-month-olds -but not 6-month-olds- look at the target of the action in a video before the agent's hand arrives in the goal area (Falck-Ytter, Gredebaeck, and Hofsten 2006). Such action prediction doesn't require the infant to map those actions onto their own motor repertoire (a common previously held assumption; see de Klerk, Southgate, and Csibra 2016 for discussion). The actions do not even have to be performed by a human to be predicted: 6.5-month-old infants can attribute a goal to an inanimate box (Csibra 2008).

Whilst responses to 2D screen content may resemble natural 3D content in babies there is considerable evidence that they struggle with understanding the representational nature of 2D images. Despite the fact that young infants can recognize on-screen objects, even 3-year-olds still make errors in this regard, believing for instance that photographs taken in advance will change if the represented scene changes (Donnelly, Gjersoe, and Hood 2013). Similarly, 3-year-olds assume that popcorn would spill out of a televised popcorn bowl if the television was turned upside down (Flavell et al. 1990). As these results show, the development of representational insight follows a similar course for video as for still images; while 9-month old infants try to grasp objects on the screen, between 15 and 19 months of age, they will instead begin to point at the screen (Pierroutsakos and Troseth 2003). In time, toddlers progress from perceiving a 'picture as object' to 'picture as representation'. When they perceive the videos as object, before 18 months of age, they probably do not understand what kind of an object a video is. This may be why research focusing on learning from video has also found that toddlers require twice as much exposure to learn from video than from a real-life event (Barr, Muentener, and Garcia 2007; Strouse and Troseth 2008), which is referred to as the '*video deficit*' effect (Anderson and Pempek 2005). At least until 30 months of age, toddlers more often imitate target behaviors demonstrated by an in-person model than by that same model on video (Barr and Hayne 1999; Hayne, Herbert, and Simcock 2003).

Research using many tasks (e.g., imitation, word-learning, self-recognition) suggests that the video deficit is most pronounced around 15–24 months of age (when they realize that video differs from reality and that is why it is not trustworthy) and then declines until about 36 months (see Anderson and Hanson 2010; Barr 2010; DeLoache et al. 2010; Troseth 2010). The video deficit in performance of more difficult tasks may even persist beyond 36 months (Dickerson et al. 2013; Roseberry et al. 2009). In a more recent study (Kirkorian et. al. 2016), recorded 24-month-olds' eye movements while they were watching an experimenter hiding a sticker behind a different shape on a felt board. Then children were given the felt board and asked to find the sticker. For half of the participants, the hiding events were in person; for the other half, the hiding events appeared on screen via closed-circuit video. Compared to those watching in-person events, children watching video spent more time looking at the target location overall, yet they had relatively poor search performance. Children who watched in-person hiding events had high success rates even if they paid relatively little visual attention to the correct location. Their findings also confirmed the video deficit hypothesis.

What are the minimum cognitive skills necessary for babies to perceive movies?

To be able to make sense of any audio-visual medium, whether cinematographic representations or real-world scenes or animations infants need i) the sensory abilities to competently see and hear the content, ii) then they need to develop some perceptual and cognitive skills to organize and interpret the sensory information in order to represent and understand the events represented, and finally iii) they need to learn to decode some cinematic techniques which do not have a real life counterpart, in other words they need to gain film literacy.

Regarding sensory skills, we know that infants do not arrive with all of their senses fully formed. Although *hearing* is the most mature sense at birth, the quietest sound a newborn responds to is about 4 times louder than the quietest sound an adult respond to. Moreover, adults usually hear in a narrow band of sound, while babies seem to be listening broadband or to all frequencies simultaneously. *Vision* on the other hand is the least mature of all the senses at birth since the foetus has very little to look at. New-borns are extremely nearsighted. Their lenses are immature. The smallest stripes to which new-borns respond are about 40 times larger than what can be resolved by adults with normal vision (Brown and Yamamoto 1986). There is at least a fivefold improvement in visual acuity -sharpness of vision- by 6 months of age, increasing slightly more than 1 octave every 3 months. Visual acuity further improves by about one-half an octave at each of 12, 24, and 36 months.

However, even at 36 months, mean acuity is still 0,75 octaves less than that of adults (Courage and Adams 1990). Visual acuity was found to be fully mature between the ages of 5 and the mid teenage years, while contrast sensitivity was found to mature fully between the ages of 8 to 19 years, which is later than previously thought (Leat, Yadav, and Irving 2009).

Another question regarding infant vision is what they look at. New-borns actively scan their surroundings, even in a completely darkened room, which may be an initial, primitive basis for looking behaviour (Haith 1980). New-borns can follow moving images with a series of saccades. The ability to smoothly track moving images develops rapidly over the first few months. It is, however, easier for them to track horizontally moving objects than vertically moving objects. Infants prefer to look at moving objects instead of non-moving objects. They also prefer to look at patterned stimuli instead of plain, non-patterned stimuli. More relevantly to our topic, infants attend preferentially to faces and face-like configurations (Farroni et al. 2005; Johnson et al. 1991) where they can receive attentional and emotional cues (Phillips, Wellman, and Spelke 2002), which we will discuss in detail in the next section.

A challenging part of visual perception is that the same object can look very different from different distances. We need to see an object as its real size despite its distance from us to the perceive stability. This feature of perception is called *size constancy* and as studies using preferential looking method reveal, it presents at birth (Granrud 1987; Slater, Mattock, and Brown 1990). This, however, does not mean that babies can transfer this ability to moving images. Even adults who have no prior exposure to moving images fail to understand that an object depicted across shots of varying sizes is actually the same object. An anecdote about first-time adult viewers in Africa suggested that naïve viewers thought that the mosquitoes shown in close-up shots were giant mosquitoes and since they do not have such big mosquitos in their villages, they did not need to worry about malaria (Forsdale and Forsdale 1966). This anecdote shows that a viewer needs to have the notion of camera and know that it can approach or move away from objects or scenes to be able to perceive the size of the depicted object constantly. Similar misinterpretations can be observed in children. Author SI's daughter thought that a dog depicted in closer shot was the mother of the dog presented in a previous long shot in her picture book before she was younger than 2 years of age. So this visual skill needs to be supported by literacy of the conventions of each medium and infants needs to decode this symbolic system used to present content.

Another perceptual skill we need to have in order to make sense of our visual environment is *perceptual completion*. In real world and in films, many visible objects are partly occluded by other objects and in films we usually see a part of an object but adults do not have difficulty to perceive them as complete. They do this by registering missing portion of the object and using available information from the visible segments, including their shape, position, orientation, motion, relative distance, luminance, color, and texture. Research on the development of perceptual completion uses the object unity paradigm, developed initially by Kellman and Spelke (1983). After habituation to a moving rod with its center occluded by a box, 4 months old infants look longer at a non -occluded presentation of a broken rod than at a complete rod, which indicates that this is not what they expect. In contrast, new-borns consistently prefer the complete rod test display, implying that the broken rod is familiar relative to the habituation display (Slater et al. 1996; Slater et al. 1994; Slater, Morison, Somers, Mattock, Brown, & Taylor, 1990). Viewers need this perceptual skill to make sense of shot sizes other than long and very long shots. Although an early film theorist Yhcam recommended directors to avoid using medium long shots as well as close-ups frequently to make their films understandable and argued against American shot size (medium-long, 'knee' film shot) saying that such shots show people on the screen like disabled, majority of film makers in film history did not listen to him and contemporary adult viewers in Western countries seem to have no problem with making sense of such shot sizes (Abel 1912). The film illiterate adults, however, interpreted the medium shot size (showing only the upper part of depicted person) as it was showing the depicted person as sitting (although this was not the case, see Ildirar and Ewing 2018).

Mental rotation is another perceptual skill we need to have to make sense of both real world and films. Studies reveal that 2-month old infants appeared to perceive the 3-D shape of rotating objects (Johnson et al. 2003). Other studies found that 4-month-old infants form dynamic mental representations that allow them to both track the movement of a 2-D object rotating in the frontal plane and anticipate the object's ultimate orientation (Rochat and Hespos 1996; Hespos and Rochat 1997). We also know that, infants who manually explore the test object before testing (Moehring and Frick 2013), and infants who are able to crawl (Schwarzer, Freitag, and Schum 2013) are both more successful in performing the mental rotation task than infants of the same age. Boys are also better than girls at mental rotation (Moore and Johnson 2008). The ability to rotate mentally (measured in terms of decline in response time) peaks in young adulthood and declines thereafter. This ability is essential to make sense of different camera angles edited together. In a study with film illiterate adults (Ildirar and Schwan, 2015) we asked participants how many animals they saw in the film clips, each showing an animal from two different camera angles edited together. Majority of the first-time adult viewers thought that there were two different animals whilst all of the same aged experienced viewers from the same culture with the same education level thought there was only one animal shown from different perspectives. Here the problem was not an inability of mental rotation but the lack of the knowledge about the film making process: possibility of recording the same scene from different camera angles and editing them together.

Perceiving continuity across cuts

To make sense of moving images, after gaining required perceptual and cognitive abilities one also needs to perceive the continuity between film shots. In the early days of cinema, most films depicted simple real-world scenes or staged narratives filmed in a single run (a shot) from a static camera. Such *tableaux* often creatively intermixed live-action with animation to create fantastical effects in these single images. For example, the sudden transformations of Georges Méliès' stop-camera shorts or the hand-painted color tinting and stop-motion animation of his *Le Voyage dans la Lune* (1902). Shortly thereafter filmmakers combined multiple shots to create more compelling visual narratives and the number of the shots they use has increased dramatically over the history of film (see Cutting and Candan 2015 for numbers). And the children's programs got their share: For instance, the editing pace of *Sesame Street* increased from 4 cuts per minute in 1977 to 8 cuts per minute in 2003 (Koolstra et al. 2004). What real-world cognitive abilities do infants need to develop before they can perceive continuity across cuts?

Matched-Exit/Entrance

One of the most primitive types of cut depicts a character moving out of shot and their motion continuing in the next shot. These *matched-exit/entrances* emerged in the earliest edited films as a direct loan from the method of leaving a scene via the wings of a theatre and gave filmmakers a method of joining together two or more tableuxs (Smith 2006). This technique is prominent in any live-action film depicting human motion and is also very common in cell-animation (see Wile E. Coyote' never-ending chases of Road Runner in the classic *Looney Tunes* cartoons). Perceiving the character as continuing to exist during their absence from view directly exploits the basic perceptual ability of *object permanence*: perception of objects as persisting in time and space even with interruptions in perceptual contact. By 4 months after birth, infants provide evidence of occlusion perception in displays that depict fully occluded objects (Johnson, Bremner, Slater, Mason, Foster, & Cheshire, 2003). Newborn infants, in contrast, have been shown consistently to perceive similar partial occlusion displays solely in terms of their visible surfaces, failing to perceive object permanence (Slater et al. 1996; Slater et al. 1994; Slater, Morison, Somers, Mattock, Brown, and Taylor, 1990). At 6 months of age, object permanency skill matures. Infants begin to have expectations of the direction of object movement during occlusion (Kochukhova and Gredebäck 2007). At 7 to 9 months of age they start to track an object with smooth pursuit prior to occlusion, crosses the occluder with a saccade and continues smooth tracking once it becomes visible again (Leigh and Zee 1999).

Object permanence is also related with an individual's working memory capacity. Infants can also accurately update their representation of a hidden two-object array when one object is subtracted from it (Moher and Feigenson 2013). Six-month-old infants are only able to recall the shape of the easier-to-recall object, while 9-month-olds can recall the shape of both objects (but not their colors, Kibbe and Leslie 2013). This raises the question, however, if film cuts are function as occluders or again as a filmic code which prevents the perception of continuity by film illiterates before learning about the notion of editing in films. A recent study (Kirkorian and Anderson 2017) compared eye movements of 12-month olds, 4-year olds and adults to see if they will anticipate the reappearance of objects in successive animated shots depicting simple cartoon characters interacting. In the stimuli, the characters or objects were moving laterally or vertically on the screen, disappearing from one edge of the screen at the end of one shot and then re-emerging from the opposite side of the screen in the next shot. The logic behind the study was if viewers comprehend these transitions and represent the video as occurring within continuous space, viewers would anticipate object and character movement across cuts by shifting their gaze to the side of the screen opposite to where the object or character was last seen (Dmytryk 1986). Thus the researchers tested the effectiveness of one of the classic editing techniques believed by filmmakers to encourage effortless integration across film cuts by all viewers but the researchers found that anticipation across cuts did not happen as much in children compared to adults. Kirkorian and Anderson (2017) suggest that children might not be able to anticipate the actions when they are presented in edited moving images. Twelve month-old infants did not anticipate the object's reappearance across a cut (instead infant eye movements were reactive to the new shot content) and 4-year-olds responded to transitions more slowly and tended to fixate the center of the screen. Kirkorian and Anderson (2017) conclude that infants cannot integrate content across shots and understand how space is represented in edited video. Film literacy may come about through exposure.

Match-on-action

Match-on-action refers to an editing technique where a subject begins an action in one shot and carries it through to completion in the next (Bordwell and Thompson 2001) The action bridge between shots distracts the viewer from noticing the cut (i.e., edit blindness; Smith and Henderson 2008; Smith and Martin-Portugues Santacreu 2016) and provides a foundation for the perception of continuity. It is also known that it enables even first-time adult viewers, who are not able to do so in the absence of continuing action through the cuts, to perceive spatiotemporal continuity between shots (Ildirar and Schwan 2014). This technique is believed to function by both cuing attentional shifts pre-cut and using motion blur post-cut (Pepperman 2004) to limit the availability of attention and perceptual discrimination ability of viewers towards the cut (Smith 2012; Smith and Martin-Portugues Santacreu 2016). Or, in a simpler explanation: puts it "so powerful is our desire to follow the action flowing across the cut we ignore the cut itself." (Bordwell and Thompson 2001: 70). Do the infants also have such a desire to follow the actions?

Falck-Ytter, Gredebaek and Hofsten (2006) found that 12-month-old babies (but not 6-month-olds) do perform goal-directed, anticipatory eye movements when observing actions performed by others. In their study, an actor placed objects in a bucket and both 12-month-old infants and adults fixated the goal of an ongoing manual action. Six-month-olds however tracked the moving hand rather than fixating the goal. Imagine a newborn, who cannot know where the bed ends and where the curtain begins since it does not know what they are. Discriminating and categorizing actions are even harder than objects as evidenced by the fact that verbs are more difficult to learn than nouns. So it cannot know where peeling the banana ends and eating the banana begins. In time, babies learn to use information about the goal or intention of the acting person to segment events into meaningful units. In one study (Baldwin et al. 2001), for example 10- to 11-month-old infants were familiarized with one of two movies depicting a woman cleaning a kitchen. Each movie depicted a salient goal-directed action (e.g., replacing a fallen dishtowel or storing an ice cream container in the freezer). After the familiarization phase, infants were presented with excerpts with one-second pauses inserted into the movie. The pauses were placed either at the moment when the woman achieved the action's goal, or several seconds before. The infants looked longer at the excerpts when the pauses were placed before the goal completions, suggesting that they found those more disruptive.

For anticipating future actions segmenting events into units is critical for both adults and infants. Infants could use these initial groupings to discover more abstract cues to event structure, such as the actor's intentions, which are known to play a role in adults' global event segmentation (e.g., Zacks 2004; Zacks and Tversky 2001). Visual sequence learning is a primary mechanism for event segmentation and research show that 8-month-old infants are sensitive to the sequential statistics of actions performed by a human agent (Roseberry et al., 2011). More interestingly, adults (Baldwin et al. 2008) as well as infants in their 1st year of life (Stahl et al. 2014) can segment a continuous action sequence based on sequential predictability alone, which suggest that before infants have top-down knowledge of intentions, they may begin to segment events based on sequential predictability. These studies do however use stimuli recorded or created on computers as one single shot. As has been shown by Kirkorian and Anderson (2017), these real-world perceptual skills may not automatically enable perception of similar actions when they are depicted across multiple shots as infants are impaired in integrating content across shots and understanding how space is represented in edited video.

Eye-line Match

Another editing technique the filmmakers found helped viewers perceive continuity between film shots is *eye-line match*, which is based on the premise that an audience will want to see what the character on-screen is seeing. A film sequence with an eye-line match begins with a character looking at something off-screen, followed by a cut of another object or person. From a developmental perspective, this refers to joint visual attention: the shared focus of two individuals on an object. This ability emerges between 6-12 months of age, and consolidates through at least the 18th month of development (Butterworth and Cochran 1980; Butterworth and Jarrett 1991; Corkum and Moore 1995). It is not until 18 months, for example, that babies begin to follow others' attention to objects that are behind them (Butterworth and Cochran 1980). A recent study (McClure et al. 2017) testing 6-24 months old babies to find out if they are capable to use joint visual attention (JVA) successfully in the video chat context by observing them video chatting with their grandparents found that the development of screen-mediated JVA is within the timeline of general JVA development. However, it is not known if the same transfer of JVA skills will occur when the eliciting screen content is presented across a series of shots joined by cuts, as is the case in eye-line match cuts.

Ellipsis, Cross-cutting and Flashbacks

Film cuts can be between shots depicting a scene from different camera angles, or an event taking place at the same time in different places or, more challengingly, the temporal sequence of a narration can be altered by using cuts. Given the fact that before 19 months of age they can not even order simple actions like poring imaginary liquid from a toy pot to a cup and (then) giving a doll drink (Fenson and Ramsay 1981), it would not be unfair to expect them to understand time gaps or flashbacks in the films. As a matter of fact, even for ten years olds flashbacks are not easy to understand. Flashbacks are not superior to jumbled editing when the children are asked to sequence the event presented (Lowe and Durkin 1999). A more recent study also found that 8 years olds (the oldest age group in the study) have difficulties coping with narrative discontinuity (Munk et al. 2012). However, it is common in children's programmes as in the early days of cinema to use special effects to alert the viewer that the action shown is a flashback; for example, the edges of the picture may be deliberately blurred, or unusual coloration or sepia tone, or monochrome may be used. So, once the children can learn the meaning of this cinematic symbol, then such effects might be useful.

Narration

Another strategy contemporary kid programmes borrow from the early days of cinema is using a narrator. Although originally live narrators were used to help viewers to understand the narration in the absence of sound during the silent era of cinema (Standish 2006), today they are used as an addition to sound to help little viewers to understand the narration (for a very successful example, see *Peppa Pig*, a British preschool animated television series).

Directing infant attention

While watching moving images, bottom-up (i.e., stimulus-driven) and top-down factors (i.e., task, preference, on-going comprehension) influence viewers' eye movements (Henderson 2007; Tatler et al. 2011). Bottom-up influences are characterized by how well the salience of low-level stimulus features based on luminance, contrast, color, orientation, and motion accounts for eye movements and studies show that adult viewers' eye movements when watching dynamic stimuli are influenced by bottom-up saliency (Borji and Itti 2013; Mital et al. 2011; Smith and Mital 2013). However, adult viewers' eye movements do not correlate with the most salient location in the image while watching Hollywood movies (Shepherd et al. 2010). To test the role of top-down factors on viewer's eye movements, some researchers gave the viewers some tasks and found that changing observers' task affects eye movements when viewing static images (Yarbus 1967), dynamic stimuli (Smith and Mital 2013), and when performing natural actions (Franchak and Adolph 2010; Hayhoe et al. 2003; Land, Mennie, and Rusted 1999). Even in the absence of an explicit task, top-down factors influence free viewing by prioritizing semantically-relevant stimuli such as objects and faces. Faces attract observers' gaze when viewing static images (Cerf et al. 2007; Yarbus, 1967) or dynamic movies (Foulsham et al. 2010; Klin et al. 2002). The tendency to look at faces, which starts from very early days of infancy, contribute to eye movement consistency among observers (Frank, Vul, and Johnson 2009).

Whether influenced by bottom-up or top-down factor, adults' eye movements are highly consistent when freely viewing dynamic stimuli – observers tend to look at the same location at the same time (Dorr et al. 2010; Hasson et al. 2008a, Mital et al. 2011; Shepherd et al. 2010; Smith and Mital 2013; Wang et al. 2012). Hollywood movies, however, evoke greater consistency in eye movements compared to homemade, "naturalistic" movies (Dorr et al. 2010; Hasson et al. 2008a, Hassan et al. 2008b). Do the eye movements of the little viewers also consistent when they are watching moving images?

Studies comparing eye movements of adults and infants suggest that similarity of gaze location increase with age (Frank et al. 2009; Kirkorian et al. 2012, Franchak et al. 2016). A study shows increasing eye movement consistency across age in 3-, 6- and 9-month-olds and adults watching short clips from *A Charlie Brown Christmas* (Frank et al. 2009). A more recent study found the same results for 1-year-olds, 4-year-olds, and adults viewing an episode of Sesame Street mixing live-action with puppetry and animation (Kirkorian et al. 2012).

Bottom-up and top-down factors influence eye movements differently in infants compared to adults. Although young infants prefer to look at faces in static image arrays over other types of stimuli (Gliga et al. 2009; Gluckman and Johnson 2013; Libertus and Needham 2011), the proportion of time spent fixating faces in static images (Amso et al. 2014) and dynamic displays (Frank et al. 2009) starts at a modest level before increasing gradually over development. Bottom-up and top-down influences are not independent, and thus are difficult to disentangle.

Watch Like Mother: Accommodating Infant Cognition in Tots TV Design

Infant brains are noisier information processing systems; this means that messages intended for infants have to allow clearer differentiation between those aspects that are required for the processing message (the 'signal') and those that are not (the 'noise'). This need for simplification of the sensory environment gives rise to exaggeration and accentuation of speech by caregivers (i.e., Motherese; Ferguson 1964) and the use of primary colors and bold shapes in the design of children's toys. The same problem of cutting through the sensory noise is faced by producers of infant-directed TV (*Tots TV*) and the formal differences between Tots TV and adult-directed TV (including the increased use of animation over live-action) seems to suggest that designers have intuited the differing demands of infant cognition.

Analysis of infant gaze behaviour while watching TV has indicated that the fixation locations of young infants (3-6 months) is more predicted by visual salience than by semantic features of the scene such as faces (Frank, Vul, and Johnson 2009). However, visual salience can also be used to guide attention to the most important area of an image, typically the face of the speaking character. Franchak et al. (2016) compared eye movements between 6- to 24-month-old infants and adults during free viewing of one-minute-long clip from *Sesame Street* and found that whilst adults fixated the human actor's face more frequently than did infants, infants spent more time fixating the highly salient Muppet faces. This potential for well-designed animated or puppet faces to attract immature gaze was also demonstrated in a developmental study from our lab. Smith Wass, Dekker, Mital, Saez De Urabain and Karmiloff-Smith (2014) compared the gaze behaviour of 6-month, 12-month and adult participants viewing excerpts from baby DVDs including live-action clips of animals, children interacting and 2D animations of cartoon shapes moving across the screen. Infants had less *attentional synchrony* (i.e., spatiotemporal clustering of gaze across participants; Smith and Mital 2013) than adults (Frank et al. 2009; Kirkorian et al. 2012) but these differences were mostly due to scenes with high feature entropy, e.g., no clear peaks of high flicker, luminance or color contrast. Such scenes tended to be live-action depictions of complex visual scenes in which the foreground objects (e.g. cows) were hard to discern from the background (e.g. fields and bushes). During scenes with low entropy (e.g. simple animations of cartoon animals, or real objects shot against a white backdrop) or the presence of a face, infant gaze behaviour was indistinguishable from adult gaze patterns. These results suggest that the designers had optimised these shots to accommodate the limited top-down control of infant gaze and guide attention to the most informative part of the image, a face. Further evidence that designers of Tots TV optimise 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) comes from a computational corpus analysis performed in our lab. Wass and Smith (2015) compared the distribution of visual features (i.e., luminance, colors, edges, flicker, and motion) in high-quality Tots and Adult TV and found that Tots TV has a better signal-to-noise ratio than Adult TV with peaks in low-level visual features predicting the location of the speaking face (the *signal*) more often than in Adult TV. The editing rate of Tots TV was slower and the average shot size larger (i.e., more Long Shots) which may give the slower attentional system of young children longer to find the focal object within a frame. Camera movements were less frequent and in combination all of these design decisions allowed the composition to point directly to the speaking face in most shots. By comparison, Adult TV was frenetic with rapid editing, highly-mobile cameras and muted palette which allocate as much saliency to peripheral features as speaking faces making the task of finding character faces a more effortful/cognitive task. As we and others have demonstrated, this can result in greater variance in how infants distribute their gaze across the frame and a lower likelihood of finding the speaking face (Frank et al. 2009; Kirkorian et al. 2012; Smith et al. 2014; Franchak et al. 2016). Interestingly, these formal differences between TotsTV and AdultTV were not entirely due to the TotsTV being animated. The TotsTV programs we analysed were a combination of live-action (e.g., *Teletubbies*, *In The Night Garden*), 2D and 3D animation (*Charlie & Lola*, *Tree Fu Tom*, *Octonauts*) and mixtures (*Baby Jake, Abadas*). We also included a selection of adult-directed animated (2DTV), puppetry (*Mongrels*) and stop-motion animated programs (*Rex The Runt*), and whilst these shared similar brightness and color profiles to the TotsTV, they were shot and edited similar to the AdultTV, suggesting that their creators understood the need to tailor the flow of AV information to their older audiences.

Of course, ensuring an infant attends to the part of the image you want does not guarantee they perceive the content in the intended way. Pempek and colleagues (2010) showed 6, 12, 18 and 24-month-old babies normal and distorted version (its shots were randomly ordered and the dialogues it included were reversed) of the Teletubbies, a television program designed for very young children. They found that the youngest infants (6, 12 and to some degree 18-month-olds) looked at the normal and distorted versions for the same amount of time, suggesting they could not discern any perceptual difference in the two types of video. Only 24-month-olds distinguished between normal and distorted video by looking for longer durations towards the normal version (Pempek et al. 2010). You can lead a horse to water, but you can't make it drink!

How Babies Perceive Faces and the Emotions They Express

The studies above demonstrate the significant role faces play in structuring film and TV content and guiding viewer gaze. But what evidence is there that infants respond to the depicted faces once they attend to them? Children, older than 3-month-olds show a clear face preference in either dynamic displays or static stimulus arrays (see Frank, Amso, and Johnson

2014 for a review), which might explain why the trains, cars, flowers etc. have faces in the animated films. This preference varies with individual differences in attentional control. Infants watched two different videos (*Charlie Brown*, and a live-action clip from *Sesame Street*) and the scholars tested not only their looking at faces but also their attentional abilities (Frank, Amso, and Johnson 2014). Replicating previous findings, they found that looking at faces increased with age. In addition, they found that infants who showed weaker attentional abilities also looked less at faces and more at salient regions. This relation was primarily seen in the youngest infants (the 3-month-old group), and was stronger than the relation between chronological age and face looking (both in that group and in the 6- and 9-month-olds groups).

There are also studies about how infants scan individual components of a face. There is evidence that 1-month-olds fixate primarily on the outer contour of the face but by 2-month-olds focus on internal elements, mostly the eyes and mouth (Hainline 1978; Maurer and Salapatek 1976). When scanning of the internal features of static faces, young infants spend more time attending to the eyes than the mouth (Haith, Bergman, and Moore 1977; Hunnius, de Wit, Vrins, and von Hofsten 2011). Most faces infants see, however, are not static but dynamic and viewed within a social context. A study (Wilcox et al. 2013) exploring infants' scanning of dynamic faces, in which an adult female speaking and acting in a positive and engaging way, found that by at 3 to 4 month of age infants perceive both eyes and mouth as important sources of information and scan faces accordingly. In contrast, the 9-month-olds spent a significantly greater proportion of time looking at the eyes than the mouth of the face. These scholars conclude that with time and experience infants identify that a great deal of socially relevant information can be quickly and effectively gathered from the eyes, making mouth scanning redundant and unnecessary.

When looking into eyes, infants prefer direct gaze: Two- to 5-day-old new-borns prefer to look at direct gaze when presented along side an averted gaze of the same person (Farroni et al. 2002) and contemporary infant-directed television programmes seem to use this information as well by breaking the fourth wall (see *Teletubbies* for an example). The position of a face relative to the observer has an important role on emotion detection and recognition due to availability of information presented in the face. A study (Goren and Wilson 2006) found that happiness is the least affected by peripheral presentation.

Which expressions can be discriminated in the first few days after birth and whether there is a particular expression that is preferred over others are the initial questions tried to be answered by the scholars working on the emotion perception in infancy. Given the results of adult cognitive neuroscience studies found that fearful faces may maximally engage rapid and subcortical processing (Adolphs and Tranel 2003) and the fearful expression contains wide eyes and an open mouth, one might think that new-borns must be sensitive to faces that display fear. However, new-borns cannot discriminate between the neutral and fearful faces (Farroni, Rigato and Johnson 2007, experiment 2) but they do discriminate happy and fearful faces and prefer to look at happy faces not fearful ones (Farroni, Rigato and Johnson 2007, experiment 3). Studies have also demonstrated that infants display a wide range of emotional expressions as early as 3 months of age, including interest, enjoyment, surprise, sadness/distress, anger, and discomfort/pain (Haviland and Lelwica 1987; Malatesta and Haviland 1982).

Many mother–infant studies using live interaction paradigms suggest that infants recognize the emotional expressions of their own caregivers and respond to them meaningfully as early as 2 to 3 months of age (e.g., Beebe and Gerstman 1980; Cohn and Ellmore 1988; Field 1977). At approximately 3 months of age, infants can discriminate among facial expressions of happiness, anger, fear, surprise, anger, and disgust (e.g., Barrera and Maurer 1981; Kuchuck, Vibbert, and Bornstein 1986; Serrano, Iglesias, and Loeches 1992). Again by 3 months of age, infants have 'expectations' about their mother's behaviour during social interactions and respond to violations of those expectations with meaningful affective changes (Gusella et al. 1988; Izard et al. 1995). In contrast to interaction studies, findings from experimental investigations of infants' recognition of emotional expressions suggest that only 7-month-olds are capable of discriminating among happy, interested, angry and sad expressions of strangers (Soken and Pick 1999). By 7 months, infants who hear a happy vocalization look longer at a 'happy face' than at a 'sad face' (e.g., Kahana-Kalman and Walker-Andrews 2001; Soken and Pick 1992). As babies gain mobility and begin to explore the world, they instinctively return to the caregiver periodically for emotional cues and respond to the emotional signal conveyed (Sigman and Capps 1997). By 12 months, infants are thought to use this early understanding of facial expressions to guide their own behavior through social referencing (e.g., Hertenstein and Campos 2004). It is now known that for testing young infants' emotion recognition, videos are better stimuli than still images since the crucial affective information is conveyed in the dynamic motion is lost in static photographs (Ekman, Friesen, and Ellsworth 1972). It is also known that multimodal displays are better. Experiments conducted on infants' intermodal perception of emotions demonstrate that 5- to 7-month old infants look preferentially to a dynamic facial expression accompanied by its characteristic vocal expression even when many of the relations are distorted or eliminated (Soken and Pick 1992; Walker-Andrews 1986). For example, Walker-Andrew (1986) presented infants with two films portraying a woman with two different expressions along with a single vocal expression corresponding to one of the facial expressions. In a series of experiments, both 5- and 7-month old infants looked more to all facial expressions (happy, sad, angry, and neutral) when they were sound specified than when they were not. Another factor effecting the difference that may contribute to the reported age difference is that in most studies the emotional expressions are portrayed by an unfamiliar actress, whereas in interaction studies they are typically portrayed by a familiar person (i.e., mothers). Indeed, Barrera and Maurer (1981) showed that 3- month-olds find it easier to discriminate among facial expressions portrayed by their own mothers than among the same expressions portrayed by a stranger. Kahana-Kalman and Walker-Andrews (2001) also found that person familiarity plays an important role on infants' early recognition of emotional expression. In contrast to an earlier study (Walker 1982), which used strangers to test the between the face and voice and found that not 5 month-olds but 7 month-olds prefer the affectively concordant display, Kahana-Kalman and Walker-Andrews (2001) found that 3.5 month-old infants detect and respond to the affective correspondences in their own mothers' facial and vocal expressions even when synchrony relations between the face and voice were disrupted.

In the above-mentioned experiments, the stimuli are typically presented in a very accessible format: at eyelevel, large enough so that all details can be appreciated. To be able to understand how infants actually see the faces in real life, Franchak et al. (2010) used head-mounted eye-tracker and found that 14-month-olds rarely fixated their mother's face, even when she spoke to them directly. They looked instead at her hands or other parts of her body. The authors interpreted that this result might have been due to the mother's location, usually high above the child. Frank (2012) recorded 2-3 hours of the visual experience of a single child at ages 3, 8, and 12 months with a head-mounted camera and found systematic changes in the visibility of faces across the first year and showed the postural shifts (lying, sitting, standing, crawling, or being held) as possible reason for that.

However, in contrast to mobile mothers, screens can be placed at eyelevel of the babies regardless of their postures, maximising the probability that the baby looks at the screen. But how do we ensure they find the face or perceive the expressions once their gaze lands there? As discussed previously (Wass and Smith 2014), creators of TotsTV see to have intuited the use of animation as a way to simplify this task for their immature viewers. By minimising background features and limiting camera movements and editing, TotsTV make the faces of speaking characters more salient relative to the background, enabling viewer gaze to find them faster. Also, we have shown that visual features are actually more clustered in the faces in TotsTV compared to AdultTV, with specific areas of high contrast around the eyes, the center of emotional expression (Wass and Smith 2014). These high-contrast eyes (think the classically enlarged and baby-like ‘Disney eyes’) help guide immature gaze to the most emotionally-rich element in each frame, potentially aiding narrative comprehension and understanding of character intentions. Empirical studies of the impact of caricature on adult face recognition have confirmed cartoonists’ intuitions that by caricaturing a person they are making recognition faster and easier than using a realistic likeness (Rhodes, Brennan, and Carey 1987). Ironically, simplification of the facial features in TotsTV might inadvertently lead to simplified potential for a range of emotional expressions (often this is exacerbated by limited range of movements in puppeted heads or time constraints on drawing complex facial action units). However, the remaining extremes of emotion that are possible may mirror the simplification of the narratives and characterisation of TotsTV typically used. As audiences mature, so does the complexity of such features as well as migration towards the nuanced performances of live-action actors (or more life-like animation; see Japanese Anime).

How Babies Perceive Emotions Portrayed by Fictitious Characters?

One example of clear social learning from TV is learning an emotional response from an actor. For example, 12-month-olds avoid a novel toy after watching a televised model that showed negative reaction toward it (Mumme and Fernald 2003). In their study, Mumme and Fernald (2003) showed infants a televised scenario of an actress reacting with neutral, positive, or negative emotion to one of two interesting novel objects in front of her and gave them an opportunity to interact with the real-world versions of the two objects to find out how the emotional reactions of the actress influenced the infant's own reactions. After witnessing the televised adult react negatively towards one object, 12-month-old infants avoided that object once it was within reach. In the positive condition, however, there was no significant change in infants’ tendency to touch the target object. In contrast to the differential responding of the 12- month-old infants, the 10-month-old infants did not vary their behaviour toward the objects in response to positive and negative emotional reaction. Learning appropriate emotional reactions through simulation of others is a key developmental process. Research on infants’ emotional life suggests that dramatic and rapid changes occur in the first 2 years of life with respect to infants’ abilities to perceive and respond to the emotions of the others (see Walker-Andrews 1997, for a review). Such developmental changes contribute to infants’ ability to become active participants in social interactions

(Malatesta et al. 1989) and may also influence how infants emotionally respond to screen-based content, although considerably more research is required on this developmental question.

Although differentiating emotions in stories is more difficult than in pictures, studies have shown that children can label the emotion conveyed by brief stories describing causes and consequences (e.g., Reichenbach and Masters 1983; Widen and Russell 2002). Even 3-year-olds can differentiate situations that elicit positive emotions from those that lead to negative emotions (Stein and Levine 1989), however it is more difficult to establish the causes of negative emotions. Furthermore, it is not clear that young children are interpreting what they perceive in terms of discrete separate emotions. Children use different emotion labels with different frequencies even when presented with an equal number of facial expressions for each emotion; the order from highest to lowest is typically happiness, sadness, anger, fear, surprise, and disgust (Gosselin and Simard 1999; Izard 1994; Widen and Russell 2003). This pattern had been observed for children's 'correct' responses, but the same order was also found for children's 'incorrect' uses (Widen and Russell 2003), suggesting that differential use of emotion labels reflects children's developing category system.

Two years old children can only divide emotions into two broad categories – feeling good or feeling bad. The older they get, the more precise their ability to make a distinction becomes (Widen and Russell 2010). Widen and Russell (2010) asked to label the emotion conveyed by each of five facial expressions and, separately, by stories about the corresponding emotion's antecedent cause and behavioral consequence. They did not require children to choose from a prespecified list of labels but asked how the character feels. They categorized the labelling levels of children into four groups: The children at labelling level 1 (mean age 33 months) used happiness to label not only the smiling 'happy face' and the happy story but, indeed, also all faces shown to them and all stories told to them. At labelling level 2 (mean age 37.9 months), some children used happiness and sadness, others used happiness and anger. In either case, children used these labels liberally. Happiness was used more narrowly than in labelling level 1, and the negative term was used broadly for most negative faces and stories. At labelling level 3 (mean age 41.4 months), children used three labels (happiness, sadness, anger) with fewer events falling into each category, but still more events than would be seen with adults. At labelling level 4 (mean age 48.1 months), children used four labels (happiness, sadness, anger, fear) to cover the 10 emotional stimuli. For example, even though they did not use the label disgust, they were not silent when faced with a disgust face or story. They assimilated these stimuli to their four categories. By the age of 4 or 5, all basic emotions (Ekman and Friesen 1975) are recognized (Camras and Allison 1985). By the age of 5, children can distinguish fear and sadness (Harris et al. 1989), whereas the discrimination between sadness and anger is difficult even for 6- to 7-year-olds (Levine 1995).

As a matter of fact, emotions portrayed by fictious characters is important part of all kind if narrative media and requires inference generation and building situation models accordingly. The ability to build emotional inferences depends on several factors, such as the sufficiency of background information (Molinari et al. 2009), the reader's knowledge about emotions (Gernsbacher and Robertson 1992) and age (Diergarten and Nieding 2015). By constructing various kinds of inferences, we combine information given in the narration with our own world knowledge to form a sophisticated representation of the state of affairs described or just implied. Whereas most of the research on emotional inferences concentrates on written text, films are capable of generating perceptually enriched situation models given that perceptible information enhances emotional inferences in comparison with nonperceptible information. For example, Gillioz, Gygax, and Tapiero (2012) found that behavioral information about emotion led to greater differences in reading times than in emotional labels (e.g., "She danced all night" vs. "Suzanne was feeling happy", p. 240). This is also supported by research on the modality effect, which shows that a dual presentation (audio and visual) is superior to a visual presentation (Leahy and Sweller 2011). However, every kind of medium is based on a certain set of symbol systems and understanding these systems is essential for understanding the content. So, film specific media literacy is also required to build the situation models and generate emotional inferences. By the age of 3 years, children are able to comprehend simple edited video stories as well as they comprehend unedited stories (presented as one continuous shot), and by the age of 4 years, they are able to substantially comprehend shot sequences conveying spatial relations, such as those implied by deleted actions, simultaneity of action, and character point of view (Smith, Anderson, and Fischer 1985). However, comprehension of video sequences during early childhood is limited. It continues to improve throughout middle childhood (Calvert 1988; Smith et al. 1985).

Potter (1998, 2013) distinguishes between rudimentary and advanced skills of media literacy. Rudimentary skills include the fundamental abilities to recognize the symbols used by media, recognize patterns composed by these symbols, and ascribe meaning to them, and they are developed between 3 and 5 years of age. Nieding and Ohler (2008) call this "media sign literacy" ("Mediale Zeichenkompetenz"; p. 382) and together with their colleagues they suggest that the ages between 4 and 8 years are crucial for the development of rudimentary media literacy skills. Advanced media literacy skills develop during school years and adulthood (Munk et al. 2012).

Conclusion

Babies' media exposure and developmental processes can no longer be studied in isolation from one another due to the fact that many infants begin consistently viewing moving images at 3 months of age become regular viewers when they are only two years old. In spite of the usage of the same perceptual and cognitive skills used to perceive the moving images and real visual environment, the differences between moving images and real life should not be underestimated. Films systematically deviate from the course of natural perception. For example, often the film anticipates certain events, cutting to a place immediately before something important will happen there (Bordwell 2005). As demonstrated by recent studies (Smith et al. 2014; Wass and Smith 2015) creators of kid's media seem to have intuited how to optimise the AV content in order to simplify the viewing process for their audience's immature brains by relying heavily on the potential of animation, puppetry and cartoon-like live-action scenes (think clowns and Telletubbies). Also, even simple dialogue scenes contain abrupt changes of viewing points, which are impossible in real-world situations. By utilizing formal features, films induce and shape predictive inferences in a manner that is different from real-world cognition (Magliano, Dijkstra, and Zwaan 1996). Hence, while films may make use of principles of natural perception (Anderson 1996; Smith 2012), they also contain numerous deviations from real-world conditions, possibly requiring viewers to possess sufficient knowledge of cinematic conventions to be able to comprehend the films content. Even up to 11 years of age, children struggle to distinguish reliably between reality and fiction (Mares and Sivakumar 2014; Woolley and Ghossainy 2013) and need to be educated in media literacy, preferably from as early as preschool (Diergarten et al. 2017). However, our scientific understanding of how media literacy develops and, especially its origins in typical infant cognitive development is still very poorly understood. As we have attempted to lay out in this review chapter, collaborations between developmental psychologists, media researchers and content producers (e.g., directors, producers, animators, etc.) is sorely needed to further our understanding. Until this gap in knowledge is filled, parents, teachers and content producers should be aware of the difficulties children face when decoding the audio-visual codes used in films. The immaturity of their viewers audiovisual perceptual systems should be respected in the design of content and, where possible caregivers should be encouraged to co-view content with their children to help maximise learning and the socio-emotional richness of the experience.

Acknowledgment

The research for this chapter has received funding from the European Union's Horizon 2020 research and innovation programme under the MArue Sklodowska-Curie grant agreement No. 656779

References

AAP Council On Communications and Media. "Media and Young Minds". *Pediatrics*. 2016; 138(5): e20162591.

Abel, Richard. "Yhcam discoursing on cinema: France, 1912." *Framework*, no. 32-33, 1986: 150-159.

Adolphs, Ralph and Daniel Tranel. "Amygdala damage impairs emotion recognition from scenes only when they contain facial expressions." *Neuropsychologia* 41, no. 10, 2003: 1281-1289.

Amso, Dima, Sara Haas, Elena Tenenbaum, Julie Markant, and Stephen J. Sheinkopf. "Bottom-up attention orienting in young children with autism." *Journal of autism and developmental disorders* 44, no. 3, 2014: 664-673.

Anderson, Daniel R. and Katherine G. Hanson. "From blooming, buzzing confusion to media literacy: The early development of television viewing." *Developmental Review*30, no. 2, 2010: 239-255.

Anderson, Daniel R. and Tiffany A. Pempek. "Television and very young children." *American Behavioral Scientist* 48, no. 5, 2005: 505-522.

Baldwin, Dare A., Jodie A. Baird, Megan M. Saylor, and M. Angela Clark. "Infants parse dynamic action." *Child development* 72, no. 3, 2001: 708-717.

Baldwin, Dare, Annika Andersson, Jenny Saffran, and Meredith Meyer. "Segmenting dynamic human action via statistical structure." *Cognition* 106, no. 3, 2008: 1382-1407.

Barr, Rachel and Harlene Hayne. "Developmental changes in imitation from television during infancy." *Child development*70, no. 5, 1999: 1067-1081.

Barr, Rachel, Paul Muentener, Amaya Garcia, Melissa Fujimoto, and Verónica Chávez. "The effect of repetition on imitation from television during infancy." *Developmental Psychobiology* 49, no. 2, 2007: 196-207.

Barr, Rachel. “Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice.” *Developmental Review* 30, no. 2, 2010: 128-154.

Barrera, Maria E. and Daphne Maurer. “Recognition of mother's photographed face by the three-month-old infant.” *Child development,* 1981: 714-716.

Beebe, Beatrice and Louis J. Gerstman. “The ‘packaging’ of maternal stimulation in relation to infant facial-visual engagement: A case study at four months.” *Merrill-Palmer Quarterly of Behavior and Development* 26, no. 4, 1980: 321-339.

Bordwell, David and Kristin Thompson. *Film Art: An Introduction*. New York: McGraw-Hill, 2001.

Bordwell, David. *Narration in the fiction film*. Routledge, 2005

Borji, Ali and Laurent Itti. “State-of-the-art in visual attention modeling.” *IEEE transactions on pattern analysis and machine intelligence* 35, no. 1, 2013: 185-207.

Bronson, Po and Ashley Merryman. “Baby Einstein vs. Barbie [Electronic Version]. Time. Retrieved February 21, 2008.” 2006: http://content.time.com/time/nation/article/0,8599,1538507,00.html (14.11.2017).

Brown, Angela M. and Misao Yamamoto. “Visual acuity in newborn and preterm infants measured with grating acuity cards.” *American journal of ophthalmology* 102, no. 2, 1986: 245-253.

Butterworth, George and Edward Cochran. “Towards a mechanism of joint visual attention in human infancy.” *International Journal of Behavioral Development* 3, no. 3, 1980: 253-272.

Butterworth, George and Nicholas Jarrett. “What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy.” *British journal of developmental psychology* 9, no. 1, 1991: 55-72.

Calvert, Sandra L. “Television production feature effects on children's comprehension of time.” *Journal of Applied Developmental Psychology* 9, no. 3, 1988: 263-273.

Camras, Linda A. and Kevin Allison. “Children's understanding of emotional facial expressions and verbal labels.” *Journal of nonverbal Behavior* 9, no. 2, 1985: 84-94.

Cerf, Moran, Jonathan Harel, Wolfgang Einhäuser, and Christof Koch. “Predicting human gaze using low-level saliency combined with face detection.” *Advances in neural information processing systems*, 2008: 241-248.

Certain, Laura K. and Robert S. Kahn. “Prevalence, correlates, and trajectory of television viewing among infants and toddlers.” *Pediatrics* 109, no. 4, 2002: 634-642.

Christakis, Dimitri A. and Michelle M. Garrison. "Teacher in the Living Room? Educational Media for Babies, Toddlers, and Preschoolers, A." Menlo Park, Calif.: Kaiser Family Foundation, 2005.

Cleveland, Allison and Tricia Striano. "Televised social interaction and object learning in 14- and 18-month-old infants." *Infant Behavior and Development* 31, no. 2, 2008: 326-331.

Cohn, Jeffrey F. and Marquita Elmore. "Effect of contingent changes in mothers' affective expression on the organization of behavior in 3-month-old infants." *Infant Behavior and Development* 11, no. 4, 1988: 493-505.

Corkum, Valerie and Chris Moore. "Development of joint visual attention in infants." (1995).

Corkum, Valerie, and Moore, Chris. The origins of joint visual attention in infants. *Developmental Psychology, 34*(1), 1998: 28-38.

Courage, Mary L. and Russell J. Adams. "Visual acuity assessment from birth to three years using the acuity card procedure: cross-sectional and longitudinal samples." *Optometry & Vision Science* 67, no. 9, 1990: 713-718.

Csibra, Gergely. "Goal attribution to inanimate agents by 6.5-month-old infants." *Cognition* 107, no. 2, 2008: 705-717.

Cutting, James E. and Ayse Candan. "Shot durations, shot classes, and the increased pace of popular movies." *Projections.* Volume 9, Issue 2, 2015: 40-62.

de Klerk, Carina CJM, Victoria Southgate, and Gergely Csibra. "Predictive action tracking without motor experience in 8-month-old infants." *Brain and cognition* 109, 2016: 131-139.

DeLoache, Judy S., Cynthia Chiong, Kathleen Sherman, Nadia Islam, Mieke Vanderborght, Georgene L. Troseth, Gabrielle A. Strouse, and Katherine O'Doherty. "Do babies learn from baby media?" *Psychological Science* 21, no. 11, 2010: 1570-1574.

Dickerson, Kelly, Peter Gerhardstein, Elizabeth Zack, and Rachel Barr. "Age-related changes in learning across early childhood: A new imitation task." *Developmental Psychobiology* 55, no. 7, 2013: 719-732.

Diergarten, Anna Katharina and Gerhild Nieding. "Children's and Adults' Ability to Build Online Emotional Inferences During Comprehension of Audiovisual and Auditory Texts." *Journal of Cognition and Development* 16, no. 2, 2015: 381-406.

Diergarten, Anna Katharina, Thomas Möckel, Gerhild Nieding, and Peter Ohler. "The impact of media literacy on children's learning from films and hypermedia." *Journal of Applied Developmental Psychology* 48, 2017: 33-41.

Dmytryk, Edward. *On Filmmaking*. Woburn, Mass.: Focal Press, 1986.

Donnelly, Katherine E., Nathalia L. Gjersoe, and Bruce Hood. “When pictures lie: Children’s misunderstanding of photographs.” *Cognition* 129, no. 1, 2013: 51-62.

Dorr, Michael, Thomas Martinetz, Karl R. Gegenfurtner, and Erhardt Barth. “Variability of eye movements when viewing dynamic natural scenes.” *Journal of vision* 10, no. 10, 2010: 28-28.

Ekman, Paul and Wallace V. Friesen. *Unmasking the face: A guide to recognizing emotions from facial cues.* Oxford: Prentice Hall, 1975.

Ekman, Paul, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the Human Face: Guidelines for Research and an Integration of Findings: Guidelines for Research and an Integration of Findings*. Braunschweig [u.a.]: Pergamon, 1972.

Falck-Ytter, Terje, Gustaf Gredebäck, and Claes von Hofsten. “Infants predict other people's action goals.” *Nature neuroscience* 9, no. 7, 2006: 878-879.

Farroni, Teresa, Enrica Menon, Silvia Rigato, and Mark H. Johnson. “The perception of facial expressions in newborns.” *European Journal of Developmental Psychology* 4, no. 1, 2007: 2-13.

Farroni, Teresa, Gergely Csibra, Francesca Simion, and Mark H. Johnson. “Eye contact detection in humans from birth.” *Proceedings of the National Academy of Sciences* 99, no. 14, 2002: 9602-9605.

Farroni, Teresa, Mark H. Johnson, Enrica Menon, Luisa Zulian, Dino Faraguna, and Gergely Csibra. “Newborns' preference for face-relevant stimuli: Effects of contrast polarity.” *Proceedings of the National Academy of Sciences of the United States of America* 102, no. 47, 2005: 17245-17250.

Fenson, Larry and Douglas S. Ramsay. “Effects of modeling action sequences on the play of twelve-, fifteen-, and nineteen-month-old children.” *Child Development*, 1981: 1028-1036.

Ferguson, Charles A. “Baby talk in six languages.” *American anthropologist* 66, no. 6_Part2, 1964: 103-114.

Ferguson, Christopher J. and M. Brent Donnellan. “Is the association between children’s baby video viewing and poor language development robust? A reanalysis of Zimmerman, Christakis, and Meltzoff (2007).” *Developmental psychology*50, no. 1, 2014: 129.

Field, Tiffany Martini. “Effects of early separation, interactive deficits, and experimental manipulations on infant-mother face-to-face interaction.” *Child development*, 1977: 763-771.

Flavell, John H., Eleanor R. Flavell, Frances L. Green, and Jon E. Korfmacher. "Do young children think of television images as pictures or real objects?" *Journal of Broadcasting & Electronic Media* 34, no. 4, 1990: 399-419.

Forsdale, Joan Rosengren and Louis Forsdale. "Film literacy." *Journal of the University Film Producers Association* 18, no. 3, 1966: 9-27.

Foulsham, Tom, Joey T. Cheng, Jessica L. Tracy, Joseph Henrich, and Alan Kingstone. "Gaze allocation in a dynamic situation: Effects of social status and speaking." *Cognition*117, no. 3, 2010: 319-331.

Franchak, John M. and Karen E. Adolph. "Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults." *Vision research* 50, no. 24, 2010: 2766-2774.

Franchak, John M., David J. Heeger, Uri Hasson, and Karen E. Adolph. "Free viewing gaze behavior in infants and adults." *Infancy* 21, no. 3, 2016: 262-287.

Frank, Michael C., Dima Amso, and Scott P. Johnson. "Visual search and attention to faces during early infancy." *Journal of experimental child psychology* 118, 2014: 13-26.

Frank, Michael C., Edward Vul, and Rebecca Saxe. "Measuring the Development of Social Attention Using Free-Viewing." *Infancy* 17, no. 4, 2012: 355-375.

Frank, Michael C., Edward Vul, and Scott P. Johnson. "Development of infants' attention to faces during the first year." *Cognition* 110, no. 2, 2009: 160-170.

Gernsbacher, Morton Ann, H. Hill Goldsmith, and Rachel R.W. Robertson. "Do readers mentally represent characters' emotional states?" *Cognition & Emotion* 6, no. 2, 1992: 89-111.

Gillioz, Christelle, Pascal Gygax, and Isabelle Tapiero. "Individual differences and emotional inferences during reading comprehension." *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*66, no. 4, 2012: 239.

Gliga, Teodora, Mayada Elsabbagh, Athina Andravizou, and Mark Johnson. "Faces attract infants' attention in complex displays." *Infancy* 14, no. 5, 2009: 550-562.

Gluckman, Maxie, and Scott P. Johnson. "Attentional capture by social stimuli in young infants." *Frontiers in Psychology* 4, 2013.

Goren, Deborah and Hugh R. Wilson. "Quantifying facial expression recognition across viewing conditions." *Vision research* 46, no. 8, 2006: 1253-1262.

Gosselin, Pierre and Janik Simard. "Children's knowledge of facial expressions of emotions: Distinguishing fear and surprise." *The Journal of genetic psychology* 160, no. 2, 1999: 181-193.

Granrud, C. E. “Size constancy in newborn human infants.” *Investigative Ophthalmology and Visual Science* 28, no. 5, 1987.

Gusella, Joanne L., Darwin Muir, and Edward Z. Tronick. “The effect of manipulating maternal behavior during an interaction on three-and six-month-olds' affect and attention.” *Child Development*, 1988: 1111-1124.

Hainline, Louise. “Developmental changes in visual scanning of face and nonface patterns by infants.” *Journal of experimental child psychology* 25, no. 1, 1978: 90-115.

Hains, Sylvia M.J. and Darwin W. Muir. “Infant sensitivity to adult eye direction.” *Child development* 67, no. 5, 1996: 1940-1951.

Haith, Marshall M. *Rules that babies look by: The organization of newborn visual activity*. Hillsdale N.J.: Lawrence Erlbaum Associates, 1980.

Haith, Marshall M., Terry Bergman, and Michael J. Moore. “Eye contact and face scanning in early infancy.” *Science* 198, no. 4319, 1977: 853-855.

Harris, Paul L., Carl N. Johnson, Deborah Hutton, Giles Andrews, and Tim Cooke. “Young children's theory of mind and emotion.” *Cognition & Emotion* 3, no. 4, 1989: 379-400.

Hasson, Uri, Eunice Yang, Ignacio Vallines, David J. Heeger, and Nava Rubin. “A hierarchy of temporal receptive windows in human cortex.” *Journal of Neuroscience* 28, no. 10, 2008a: 2539-2550.

Hasson, Uri, Ohad Landesman, Barbara Knappmeyer, Ignacio Vallines, Nava Rubin, and David J. Heeger. “Neurocinematics: The neuroscience of film.” *Projections* 2, no. 1, 2008b: 1-26.

Haviland, Jeannette M. and Mary Lelwica. “The induced affect response: 10-week-old infants' responses to three emotion expressions.” *Developmental Psychology* 23, no. 1, 1987: 97.

Hayhoe, Mary M., Anurag Shrivastava, Ryan Mruczek, and Jeff B. Pelz. “Visual memory and motor planning in a natural task.” *Journal of vision* 3, no. 1, 2003: 6.

Hayne, Harlene, Jane Herbert, and Gabrielle Simcock. “Imitation from television by 24-and 30-month-olds.” *Developmental Science* 6, no. 3, 2003: 254-261.

Henderson, John M. “Regarding scenes.” *Current directions in psychological science* 16, no. 4, 2007: 219-222.

Hertenstein, Matthew J. and Joseph J. Campos. “The retention effects of an adult's emotional displays on infant behavior.” *Child development* 75, no. 2, 2004: 595-613.

Hespos, Susan J. and Philippe Rochat. “Dynamic mental representation in infancy.” *Cognition* 64, no. 2, 1997: 153-188.

Hunnius, Sabine, Tessa CJ de Wit, Sven Vrins, and Claes von Hofsten. “Facing threat: Infants' and adults' visual scanning of faces with neutral, happy, sad, angry, and fearful emotional expressions.” *Cognition and Emotion* 25, no. 2, 2011: 193-205.

Ildirar, Sermin and Louise Ewing. “Revisiting the Kuleshov effect with first time viewers*”*. *Projections*. ISSN 1934-9688 (2018) 32, 1986: 150.

Ildirar, Sermin and Stephan Schwan. “First-time viewers' comprehension of films: Bridging shot transitions.” *British Journal of Psychology* 106, no. 1, 2015: 133-151.

Izard, Carroll E. “Innate and universal facial expressions: evidence from developmental and cross-cultural research.” *Psychological Bulletin115* (2), 1994: 288-99.

Izard, Carroll E., Christina A. Fantauzzo, Janine M. Castle, O. Maurice Haynes, Maria F. Rayias, and Priscilla H. Putnam. “The ontogeny and significance of infants' facial expressions in the first 9 months of life.” *Developmental Psychology* 31, no. 6, 1995: 997.

Johnson, Mark H., Suzanne Dziurawiec, Hadyn Ellis, and John Morton. “Newborns' preferential tracking of face-like stimuli and its subsequent decline.” *Cognition* 40, no. 1, 1991: 1-19.

Johnson, Scott P., J. Gavin Bremner, Alan Slater, Uschi Mason, Kirsty Foster, and Andrea Cheshire. “Infants’ perception of object trajectories.” *Child development* 74, no. 1, 2003: 94-108.

Johnson, Scott P., Leslie B. Cohen, Kathryn H. Marks, and Kerri L. Johnson. “Young infants' perception of object unity in rotation displays.” *Infancy* 4, no. 2, 2003: 285-295.

Kahana-Kalman, Ronit and Arlene S. Walker-Andrews. “The role of person familiarity in young infants' perception of emotional expressions.” *Child development* 72, no. 2, 2001: 352-369.

Kellman, Philip J. and Elizabeth S. Spelke. “Perception of partly occluded objects in infancy.” *Cognitive psychology* 15, no. 4, 1983: 483-524.

Kibbe, Melissa M. and Alan M. Leslie. “What’s the object of object working memory in infancy? Unraveling ‘what’and ‘how many’.” *Cognitive psychology* 66, no. 4, 2013: 380-404.

Kirkorian, Heather L. and Daniel R. Anderson. “Anticipatory eye movements while watching continuous action across shots in video sequences: A developmental study.” *Child development* 88, no. 4, 2017: 1284-1301.

Kirkorian, Heather L., Daniel R. Anderson, and Rachel Keen. “Age differences in online processing of video: An eye movement study.” *Child development* 83, no. 2, 2012: 497-507.

Kirkorian, Heather L., Heather J. Lavigne, Katherine G. Hanson, Georgene L. Troseth, Lindsay B. Demers, and Daniel R. Anderson. “Video deficit in toddlers’ object retrieval: What eye movements reveal about online cognition.” *Infancy* 21, no. 1, 2016: 37-64.

Klin, Ami, Warren Jones, Robert Schultz, Fred Volkmar, and Donald Cohen. “Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism.” *Archives of general psychiatry* 59, no. 9, 2002: 809-816.

Kochukhova, Olga and Gustaf Gredebäck. “Learning about occlusion: Initial assumptions and rapid adjustments.” *Cognition* 105, no. 1, 2007: 26-46.

Koolstra, Cees M., Juliette van Zanten, Nicole Lucassen, and Nazreen Ishaak. “The formal pace of Sesame Street over 26 years.” *Perceptual and motor skills* 99, no. 1, 2004: 354-360.

Kuchuk, April, Martha Vibbert, and Marc H. Bornstein. “The perception of smiling and its experiential correlates in three-month-old infants.” *Child Development*, 1986: 1054-1061.

Land, Michael, Neil Mennie, and Jennifer Rusted. “The roles of vision and eye movements in the control of activities of daily living.” *Perception* 28, no. 11, 1999: 1311-1328.

Leahy, Wayne and John Sweller. “Cognitive load theory, modality of presentation and the transient information effect.” *Applied Cognitive Psychology* 25, no. 6, 2011: 943-951.

Leat, Susan J., Naveen K. Yadav, and Elizabeth L. Irving. “Development of visual acuity and contrast sensitivity in children.” *Journal of Optometry* 2, no. 1, 2009: 19-26.

Leigh, John R. and David S. Zee. “The vestibular-optokinetic system.” *The neurology of eye movements* 19, 1999: 89.

Levine, Linda J. “Young children's understanding of the causes of anger and sadness. ” *Child development* 66, no. 3, 1995: 697-709.

Libertus, Klaus and Amy Needham. “Reaching experience increases face preference in 3-month-old infants.” *Developmental science* 14, no. 6, 2011: 1355-1364.

Lowe, Phillip J. and Kevin Durkin. “The effect of flashback on children's understanding of television crime content.” *Journal of Broadcasting & Electronic Media* 43, no. 1, 1999: 83-97.

Magliano, Joseph P., Katinka Dijkstra, and Rolf A. Zwaan. “Generating predictive inferences while viewing a movie.” *Discourse Processes* 22, no. 3, 1996: 199-224.

Malatesta, Carol Z., Clayton Culver, Johanna Rich Tesman, Beth Shepard, Alan Fogel, Mark Reimers, and Gail Zivin. “The development of emotion expression during the first two years of life.” *Monographs of the society for research in child development*, 1989: i-136.

Malatesta, Carol Zander and Jeannette M. Haviland. “Learning display rules: The socialization of emotion expression in infancy.” *Child development*, 1982: 991-1003.

Mares, Marie-Louise, and Gayathri Sivakumar. “‘Vámonos means go, but that’s made up for the show’: Reality confusions and learning from educational TV.” *Developmental psychology* 50, no. 11, 2014: 2498.

Maurer, Daphne and Philip Salapatek. “Developmental changes in the scanning of faces by young infants.” *Child development*, 1976: 523-527.

McCall, Robert B., Ross D. Parke, Robert D. Kavanaugh, Robert Engstrom, Judy Russell, and Elizabeth Wycoff. “Imitation of live and televised models by children one to three years of age.” *Monographs of the society for research in Child Development*, 1977: 1-94.

McClure, Elisabeth R., Yulia E. Chentsova-Dutton, Steven J. Holochwost, W. G. Parrott, and Rachel Barr. “Look At That! Video Chat and Joint Visual Attention Development Among Babies and Toddlers.” *Child Development*, 2017.

Meltzoff, Andrew N. “Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli.” *Developmental psychology* 24, no. 4, 1988: 470.

Mital, Parag K., Tim J Smith, Robin L. Hill, and John M. Henderson. “Clustering of gaze during dynamic scene viewing is predicted by motion.” *Cognitive Computation* 3, no. 1, 2011: 5-24.

Moher, Mariko and Lisa Feigenson. “Factors influencing infants’ ability to update object representations in memory.” *Cognitive development* 28, no. 3, 2013: 272-289.

Möhring, Wenke and Andrea Frick. “Touching up mental rotation: Effects of manual experience on 6-month-old infants’ mental object rotation.” *Child Development*, 84, 2013: 1554-1565.

Molinari, Carlos, Débora Burin, Gastón Saux, Juan Pablo Barreyro, Natalia Irrazabal, María Susana Bechis, Aníbal Duarte, and Verónica Ramenzoni. “Fictional characters' emotional state representation: What is its degree of specificity?” *Psicothema* 21, no. 1, 2009: 9-14.

Moore, David S. and Scott P. Johnson. “Mental rotation in human infants: A sex difference.” *Psychological Science* 19, no. 11, 2008: 1063-1066.

Muir, Darwin W., Sylvia M.J. Hains, Y. Cao, and Barbara D'Entremont. “3-to 6-month-olds' sensitivity to adult intentionality: The role of adult contingency and eye direction in dyadic interactions.” *Infant Behavior and Development* 19, 1996: 199.

Mumme, Donna L. and Anne Fernald. “The infant as onlooker: Learning from emotional reactions observed in a television scenario.” *Child development* 74, no. 1, 2003: 221-237.

Munk, Carmen, Günter Daniel Rey, Anna Katharina Diergarten, Gerhild Nieding, Wolfgang Schneider, and Peter Ohler. “Cognitive Processing of Film Cuts Among 4-to 8-Year-Old Children.” *European Psychologist*, *17*(4), 2012: 257-265.

Nieding, Gerhild and Peter Ohler. “Mediennutzung und Medienwirkung bei Kindern und Jugendlichen.” In: Batinic, Bernad and Markus Appel (Eds.) *Medienpsychologie*. Berlin/Heidelberg: Springer, 2008: 379-400.

Pempek, Tiffany A., Heather L. Kirkorian, John E. Richards, Daniel R. Anderson, Anne F. Lund, and Michael Stevens. “Video comprehensibility and attention in very young children.” *Developmental Psychology* 46, no. 5, 2010: 1283.

Pepperman, Richard. *The Eye is Quicker: Film Editing; Making A Good Film Better*. Los Angeles, USA: Michael Wiese Productions, 2004.

Phillips, Ann T., Henry M. Wellman, and Elizabeth S. Spelke. “Infants' ability to connect gaze and emotional expression to intentional action.” *Cognition* 85, no. 1, 2002: 53-78.

Pierroutsakos, Sophia L. and Georgene L. Troseth. “Video verite: Infants’ manual investigation of objects on video.” *Infant Behavior and Development* 26, no. 2, 2003: 183-199.

Potter, W. James. *Media literacy.* [1st Edition]. Thousand Oaks, CA: Sage, 1998. Publications.

Potter, W. James. *Media literacy.* [6th Edition]. Thousand Oaks, CA: Sage, 2013. Publications

Reichenbach, Lisa and John C. Masters. “Children's use of expressive and contextual cues in judgments of emotion.” *Child Development*, 1983: 993-1004.

Rhodes, Gillian, Susan Brennan, and Susan Carey. “Identification and ratings of caricatures: Implications for mental representations of faces.” *Cognitive Psychology,* 19 (4), 1987: 473-497.

Rochat, Philippe and Susan J. Hespos. “Tracking and anticipation of invisible spatial transformations by 4-to 8-month-old infants.” *Cognitive Development* 11, no. 1, 1996: 3-17.

Roseberry, Sarah, Kathy Hirsh-Pasek, Julia Parish-Morris, and Roberta M. Golinkoff. “Live action: Can young children learn verbs from video?” *Child development* 80, no. 5, 2009: 1360-1375.

Roseberry, Sarah, Russell Richie, Kathy Hirsh-Pasek, Roberta Michnick Golinkoff, and Thomas F. Shipley. “Babies catch a break: 7-to 9-month-olds track statistical probabilities in continuous dynamic events.” *Psychological Science* 22, no. 11, 2011: 1422-1424.

Schwarzer, Gudrun, Claudia Freitag, and Nina Schum. “How crawling and manual object exploration are related to the mental rotation abilities of 9-month-old infants.” *Frontiers in psychology* 4, 2013.

Serrano, Juan M., Jaime Iglesias, and Angela Loeches. “Visual discrimination and recognition of facial expressions of anger, fear, and surprise in 4-to 6-month-old infants.” *Developmental Psychobiology* 25, no. 6, 1992: 411-425.

Shepherd, Stephen V., Shawn A. Steckenfinger, Uri Hasson, and Asif A. Ghazanfar. “Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing.” *Current Biology* 20, no. 7, 2010: 649-656.

Sigman, Marian and Lisa Capps. *Children with autism: A developmental perspective*. Vol. 34. Harvard University Press, 1997.

Slater, Alan, Anne Mattock, and Elizabeth Brown. “Size constancy at birth: Newborn infants' responses to retinal and real size.” *Journal of experimental child psychology* 49, no. 2, 1990: 314-322.

Slater, Alan, Scott P. Johnson, Elizabeth Brown, and Marion Badenoch. “Newborn infant's perception of partly occluded objects.” *Infant Behavior and Development* 19, no. 1, 1996: 145-148.

Slater, Alan, Scott P. Johnson, Philip J. Kellman, and Elizabeth S. Spelke. "The role of three-dimensional depth cues in infants' perception of partly occluded objects." *Infant and Child Development* 3, no. 3, 1994: 187-191.

Slater, Alan, Victoria Morison, Marcia Somers, Anne Mattock, Elizabeth Brown, and David Taylor. "Newborn and older infants' perception of partly occluded objects." *Infant Behavior and Development* 13, no. 1, 1990: 33-49.

Smith, Robin, Daniel R. Anderson, and Catherine Fischer. "Young children's comprehension of montage." *Child development*, 1985: 962-971.

Smith, Tim J. "An attentional theory of continuity editing." Doctoral thesis, University of Edinburgh, 2006: http://dx.doi.org/10.3167/proj.2012.060102.

Smith, Tim J. "The attentional theory of cinematic continuity." *Projections* 6, no. 1, 2012: 1-27.

Smith, Tim J. and Saez de Urabain, I.R. „Eye tracking." In: Hopkins, Brian and Elena Geangu, and Sally Linkenauger (Eds.). *Cambridge Encyclopedia of Child Development.* Cambridge: Cambridge University Press, 2017: 97-101. https://doi.org/10.1017/9781316216491.

Smith, Tim J. and Janet Yvonne Martin-Portugues Santacreu. "Match-action: The role of motion and audio in creating global change blindness in film." *Media Psychology* 20, no. 2, 2017: 317-348.

Smith, Tim J. and John M. Henderson. "Edit Blindness: The relationship between attention and global change blindness in dynamic scenes." *Journal of Eye Movement Research* 2, no. 2, 2008: 1-17.

Smith, Tim J. and Parag K. Mital. "Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes." *Journal of Vision* 13, no. 8, 2013: 16-16.

Smith, Tim J., Sam Wass, Tessa Dekker, Parag K. Mital, Irati Rodriguez, Annette Karmiloff-Smith. "Optimising signal-to-noise ratios in Tots TV can create adult-like viewing behaviour in infants." *2014 International Conference on Infant Studies*, Berlin, Germany, July 3-5 2014.

Soken, Nelson H. and Anne D. Pick. "Infants' perception of dynamic affective expressions: do infants distinguish specific expressions?" *Child development* 70, no. 6, 1999: 1275-1282.

Soken, Nelson H. and Anne D. Pick. "Intermodal Perception of Happy and Angry Expressive Behaviors by Seven-Month-Old Infants." *Child development* 63, no. 4, 1992: 787-795.

Stahl, Aimee E., Alexa R. Romberg, Sarah Roseberry, Roberta Michnick Golinkoff, and Kathryn Hirsh-Pasek. "Infants segment continuous events using transitional probabilities." *Child development* 85, no. 5, 2014: 1821-1826.

Standish, Isolde. *A new history of Japanese cinema*. Bloomsbury Publishing, 2006.

Stein, Nancy L. and Linda J. Levine. "The causal organisation of emotional knowledge: A developmental study. " *Cognition & Emotion* 3, no. 4, 1989: 343-378.

Strouse, Gabrielle A. and Georgene L. Troseth. "'Don't try this at home': Toddlers' imitation of new skills from people on video." *Journal of experimental child psychology* 101, no. 4, 2008: 262-280.

Tatler, Benjamin W., Mary M. Hayhoe, Michael F. Land, and Dana H. Ballard. "Eye guidance in natural vision: Reinterpreting salience." *Journal of vision* 11, no. 5, 2011: 5.

Tincoff, Ruth and Peter W. Jusczyk. "Some beginnings of word comprehension in 6-month-olds." *Psychological Science*10, no. 2, 1999: 172-175.

Troseth, Georgene L. "Is it life or is it Memorex? Video as a representation of reality." *Developmental Review* 30, no. 2, 2010: 155-175.

Walker, Arlene S. "Intermodal perception of expressive behaviors by human infants." *Journal of Experimental Child Psychology* 33, no. 3, 1982: 514-535.

Walker-Andrews, Arlene S. "Infants' perception of expressive behaviors: differentiation of multimodal information." *Psychological bulletin* 121, no. 3, 1997: 437.

Walker-Andrews, Arlene S. "Intermodal perception of expressive behaviors: Relation of eye and voice?" *Developmental psychology* 22, no. 3, 1986: 373.

Wang, Helena X., Jeremy Freeman, Elisha P. Merriam, Uri Hasson, and David J. Heeger. "Temporal eye movement strategies during naturalistic viewing." *Journal of vision* 12, no. 1, 2012: 16.

Wass, Sam V. and Tim J. Smith. "Visual motherese? Signal to noise ratios in toddler directed television." *Developmental science* 18, no. 1, 2015: 24-37.

Widen, Sherri C. and James A. Russell. "A closer look at preschoolers' freely produced labels for facial expressions." *Developmental psychology* 39, no. 1, 2003: 114.

Widen, Sherri C. and James A. Russell. "Differentiation in preschooler's categories of emotion." *Emotion* 10, no. 5, 2010: 651.

Widen, Sherri C. and James A. Russell. “Gender and preschoolers' perception of emotion.” *Merrill-Palmer Quarterly*48, no. 3, 2002: 248-262.

Wilcox, Teresa, Jessica A. Stubbs, Lesley Wheeler, and Gerianne M. Alexander. “Infants’ scanning of dynamic faces during the first year.” *Infant Behavior and Development* 36, no. 4, 2013: 513-516.

Woolley, Jacqueline D. and Maliki E. Ghossainy. “Revisiting the fantasy–reality distinction: Children as naïve skeptics.” *Child development* 84, no. 5, 2013: 1496-1510.

Yarbus, Alfred L. “Eye movements during perception of complex objects.” In: Yarbus, Alfred L. *Eye movements and vision*. Springer US, 1967: 171-211

Zacks, Jeffrey M. “Using movement and intentions to understand simple events.” *Cognitive Science* 28, no. 6, 2004: 979-1008.

Zacks, Jeffrey M. and Barbara Tversky. “Event structure in perception and conception.” *Psychological bulletin* 127, no. 1, 2001: 3.

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, no. 5, 2007: 473-479.

Films & TV series

A Charlie Brown Christmas. Directed by Bill Melendez. 1965; USA: Lee Mendelson Films.

Baby Jake: UK 2011-2012, CBeebies, created by Darrall Macqueen Ltd.

Charlie & Lola: UK 2005-2008, CBeebies, created by Lauren Child.

In The Night Garden: UK 2006-2009, BBC, created by Andrew Davenport.

Le Voyage Dans la Lune (A Trip to the Moon). Directed by Georges Méliès. 1902; France: Star Film Company.

Looney Tunes: USA 1930–1969 (original; 1987-present (revival)), Cartoon Network, Nickelodeon, ABC, created by Warner Brothers.

Mongrels: UK 2010-2011, BBC Three, created by Adam Miller.

Octonauts: UK 2010-present, CBeebies, created by Vicki Wong and Michael C. Murphy.

Peppa Pig: UK 2004-present, Channel 5 and Nick Jr., created by Andrew Davenport.

Rex The Runt: UK 1998-2001, BBC Two, created by Richard Goleszowski.

Sesame Street: USA 1969-present, NET, PBS, HBO, created by Joan Ganz Cooney and Lloyd Morrisett.

Teletubbies: UK 1997-2001, BBC, created by Anne Wood and Andrew Davenport.

Tree Fu Tom: UK 2012-2016, CBeebies, created by Daniel Bays and Dan Hodgkins.

Sermin Ildirar Kirbas

Postdoctoral Research Fellow at Centre for Brain and Cognitive Development, Birkbeck, University of London, UK

Sermin Ildirar is a Marie S. Curie postdoctoral research fellow at the Centre for Brain and Cognitive Development, Birkbeck, University of London. She studied Film and Media studies at Istanbul University and University of Viennaa and works on perceptual and cognitive processes during viewing by adults and infants. She is also director of several short movies and co-scriptwriter of a feature movie and a computer game.

Her publications include *Watching Film for the First Time How Adult Viewers Interpret Perceptual Discontinuities* (Film. Psychological Science 2010), *First-time viewers' comprehension of films: Bridging shot transitions* (British Journal of Psychology 2015) and *Indexing the Events of an Art Film by Audiences with Different Viewing Backgrounds. Making Sense of Cinema: Empirical Studies Into Film Spectators and Spectatorship* (Bloomsbury 2016).

Tim J. Smith

Lecturer at the Department of Psychological Sciences, Birkbeck, University of London, UK

Tim J. Smith is a lecturer in the Department of Psychological Sciences, Birkbeck, University of London. His research covers all aspects of Visual Cognition with special focus on the active perception of naturalistic scenes via eye movements. His personal and professional passion for film has led to his applying his empirical methods to the issue of how we perceive film. How do we attend to, perceive and remember static, dynamic and real-world scenes? How do visual features of a scene command our attention and how do higher-order cognitive factors compete with these factors? How do we program eye movements around a complex visual scene and process information at fixation and in the periphery? How does our understanding of naturalistic Visual Cognition allow us to understand how we perceive visual media such as Film, TV, and Digital Media? He investigates these issue using a combination of behavioural experiments, eyetracking, cognitive modelling and electrophysiology.

His publications include numerous articles in international encyclopedias, prominent anthologies and renowned periodicals.