Infant Behavior and Development 66 (2022) 101661

Contents lists available at ScienceDirect

Infant Behavior and Development

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

Understanding the differential impact of children's TV on executive functions: a narrative-processing analysis

Claire Essex [a,*], Teodora Gliga [b], Maninda Singh [c], Tim J. Smith [d]

[a] Centre for Brain and Cognitive Development, Birkbeck, University of London, UK
[b] University of East Anglia, Norwich, UK
[c] School of Psychology, University of Surrey, Guildford, Surrey, UK
[d] Department of Psychological Sciences, Birkbeck, University of London, UK

ARTICLE INFO

Keywords:
Children's TV viewing
Executive Function
Cognitive Development
Narrative-Processing Analysis
Content Properties

ABSTRACT

Evidence from multiple empirical studies suggests children's Executive Functions are depleted immediately after viewing some types of TV content but not others. Correlational evidence suggests any such effects may be most problematic during the pre-school years. To establish whether "screen-time" is developmentally appropriate at this age we believe a nuanced approach must be taken to the analysis of individual pieces of media and their potential demands on viewer cognition. To this end we apply a cognitive theory of visual narrative processing, the Scene Perception and Event Comprehension Theory (SPECT; Loschky, Larson, Smith, & Magliano, 2020) to the analysis of TV shows previously used to investigate short-term effects of TV viewing. A theoretical formalisation of individual content properties, together with a quantitative content-based analysis of previously used children's content (Lillard & Peterson, 2011; Lillard et al., 2015b) is presented. This analysis found a pattern of greater stimulus saliency, increased situational change and a greater combined presence of cognitively demanding features for videos previously shown to reduce children's EF after viewing. Limitations of this pilot application of SPECT are presented and proposals for future empirical investigations of the psychological mechanisms activated by specific TV viewing content are considered.

1. Introduction

The influence of TV viewing on children's cognitive and emotional development has long been of interest to researchers (Foster & Watkins, 2010), with recent shifts in the family media environment (i.e., introduction of mobile touchscreen devices, internet-enabled smart devices such as smart TVs and streaming services) sparking renewed interest amongst parents and early-years practitioners. There is currently broad consensus across a number of international bodies (e.g., American Academy of Pediatrics, 2016; (World Health Organisation) issuing recommendations to restrict screen time for children younger than 18 months and limit daily screen time for pre-school aged children. These recommendations have been born out of concerns regarding the impact time spent engaging with screen content may have on early development. A prominent concern has been that time spent with screen media may displace other more beneficial activities (Kirkorian, Wartella, & Anderson, 2008), such as parent-child interactions (Chonchaiya & Pruksananonda,

* Correspondence to: Institute Centre for Brain and Cognitive Development, Department Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK.
E-mail address: cessex01@mail.bbk.ac.uk (C. Essex).

https://doi.org/10.1016/j.infbeh.2021.101661
Received 9 July 2021; Received in revised form 25 October 2021; Accepted 28 October 2021
Available online 14 November 2021

2008), or physical activities (Sisson, Broyles, Baker, & Katzmarzyk, 2010). Yet others have suggested that screen media may directly influence the developing attention skills (Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004; Landhuis, Poulton, Welch, & Hancox, 2007; L. E. Levine & Waite, 2000; Miller et al., 2007; Özmert, Toyran, & Yurdakök, 2002). In our view, the investigation of these direct links still lacks a clear theoretical framework. In the sections that follow we apply a cognitive theory of visual narrative processing, the Scene Perception & Event Comprehension Theory (SPECT; Loschky, Larson, Smith, & Magliano, 2020) to the analysis of TV shows previously used in two influential studies investigating the effects of TV viewing on children's Executive Functions immediately after viewing (Lillard, Drell, Richey, Boguszewski, & Smith, 2015b; Lillard & Peterson, 2011).

1.1. The Media Environment in early childhood

Amid concerns about the pervasiveness of media use among young children and the possible consequences for development, several international bodies have issued recommendations for parents to limit screen-time during early childhood. The American Academy of Pediatrics (American Academy of Pediatrics, 2016) and the World Health Organisation (WHO; 2019) both recommend screen-time is restricted for children younger than 18 months, while solo media use should be avoided for children under 2 years. For children aged between 2 and 5 years, screen-time should be kept to under 1 h p/day of high-quality programming, ideally co-viewed with an adult. Despite these recommendations survey data suggests engagement with screen media during early childhood often exceeds these limits. In the US, a nationally representative survey found children younger than 2 years engaged with screen media for approximately 1 h per day, while children aged between 2 and 4 years engaged with screen media for 2 h per day on average (Rideout, 2017). In the UK, in 2019 children aged between 3 and 4 years spent 12.7 h per week on average watching TV (Statista, 2020). These surveys also show that while there has been a shift in the way young children access screen content (i.e., through mobile devices and video-on-demand streaming services), video viewing continues to be a mainstay of children's early screen experience. This data points to a discord between family choices around screen-time and recommendations from leading health bodies; and also with the current scientific literature which suggests consequences of TV viewing may be most detrimental during this early childhood period (Barr, Danziger, Hilliard, Andolina, & Ruskis, 2010; Cheng, Maeda, Yoichi, Yamagata, & Tomiwa, 2010; Christakis et al., 2004; Landhuis et al., 2007; L. E. Levine & Waite, 2000; Miller et al., 2007; Özmert et al., 2002). In particular, a substantial body of correlational evidence suggests a negative association between early TV viewing and attention problems (Christakis et al., 2004; Landhuis et al., 2007; Levine & Waite, 2000; Miller et al., 2007; Özmert et al., 2002). These studies suggest the more time children spend viewing TV early in life the more likely they are to experience difficulties maintaining attention, suppressing unnecessary distractors, and rising to the demands placed on them in the classroom later in childhood. Yet, findings remain equivocal, with some studies failing to replicate the association (Obel et al., 2004; Stevens & Mulsow, 2006).

1.2. The relation between content differences and immediate effects of TV viewing

One explanation for the inconsistency in these findings is that studies have largely overlooked the role content differences may play in the reported affects (Barr, Lauricella, Zack, & Calvert, 2010). A number of studies have attempted to address this empirically by investigating the direct impact of different types of content on children's Executive Function, immediately after viewing (i.e., Geist & Gibson, 2000; Kostyrka-Allchorne, Cooper, Gossmann, Barber, & Simpson, 2017; Kostyrka-allchorne *et al.* 2019; Kostyrka-Allchorne, Cooper, & Simpson, 2019; Lillard et al., 2015b; Lillard & Peterson, 2011). Executive Functions (EF) are a set of cognitive control processes that enable us to deploy our attention and behaviour in a goal-directed manner (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). They include inhibition, interference control, working memory and cognitive flexibility (Diamond, 2013) and have been associated with children's success in school and later in life (i.e., Blair & Razza, 2007; Duncan et al., 2007; Gathercole, Pickering, Ambridge, & Wearing, 2004). As such, the influence of TV viewing on these developing skills is a key consideration for research (Lillard, Li, & Boguszewski, 2015a).

A number of content and production features have been examined by studies seeking to establish how content may differentially effect children's Executive Functions after viewing. Typically, these studies have used widely viewed shows which differ on a dimension of interest. For example, the initial focus was on the influence of *pacing* on children's EF. In the developmental literature pacing has been defined as the rate of onscreen audio-visual changes (Anderson, Levin, & Lorch, 1977; Cooper, Uller, Pettifer, & Stolc, 2009; Huston-Stein, Fox, Greer, Watkins, & Whitaker, 1981; McCollum and Bryant, 2003; (Watt & Welch, 1983). It is important to note that this formalisation of pace differs from the definition found in the adult literature where pace is specifically equated with the number of cuts (e.g., Lang, Bolls, Potter, & Kawahara, 1999). In contrast, the developmental construct of *pace* is a conflation of several features. These include measures of the amount or frequency of changes in low-level visual features (colour, luminance, motion, flicker) or audio changes (changes in music and speech) that may be amplified by camera movements (pans, zooms) or editorial transitions between shots (cuts, fades, wipes).

These featural changes have two effects: large changes in low-level features may re-orient attention to bottom-up processing (interrupting the processing of ongoing events, i.e., Singer & Singer, 1979) and they may also correspond to a faster pace of event content, which require additional processing resources. It has been suggested that an over-reliance on bottom-up processing coupled with difficulties processing (fast-paced) content, may reduce levels of attention control immediately after viewing (Lillard *et al.* 2015a), albeit the mechanisms through which EF may be affected remain unknown. A study by Geist and Gibson (2000) found 30-minute viewing of a fast-paced show (Mighty Morphin' Power Rangers©) resulted in increased attention switching and less time in focused attention during subsequent free-play relative to control groups (no viewing; slow-paced show, Mister Roger's Neighbourhood©). Lillard and Peterson (2011) replicated this finding with a study of a fast-paced show (SpongeBob SquarePants©) vs a slow-paced show

(Caillou©) and a control condition (free drawing). Composite scores from a battery of EF tasks completed immediately after viewing, indicated children who viewed the fast-paced cartoon performed significantly worse than the control group and marginally worse than the slow-paced group. While these studies point to a potentially negative impact of increased pacing on children's EF immediately after viewing, as discussed earlier there is a conflation of features in the operationalisation of pacing used by these studies. As such it remains unclear which aspect of pacing may account for the reported differences in EF immediately after viewing.

A further issue for these studies is that exemplar shows also vary on other content properties (i.e., complexity of conceptual content, target age). Lillard et al. (2015b) note that "fast-paced" shows used in the existing literature are also highly fantastical in nature. For example, SpongeBob SquarePants© is a show replete with moments of physical implausibility (e.g., SpongeBob is often seen being transported through solid objects– ceilings, walls, doors etc., or impossibly changing his appearance- colour changes, changes of form: squashed or stretched). These violations of physical laws may be problematic because they may be cognitively demanding to process; violations of expectations may require re-assessment of knowledge, which may capture attention to the detriment of other aspects of the scene (Perez & Feigenson, 2020; Stahl & Feigenson, 2017). Across three studies Lillard et al. (2015b) sought to disentangle the effects of *pace* and *fantasy* with 10 exemplar shows which varied on these dimensions. A measure of pacing was obtained for each video with an automatic shot detection tool (Scene Detector Pro, 2002), a cut was defined as a frame-to-frame change in more than 85% of pixels. A viewer annotated each occurrence of a fantastical event, defined as impossible transformations (i.e., objects or characters change shape or identity in impossible ways, exhibit impossible attributes such as violations of continuity). The authors first replicated their previous finding that EF performance was significantly lower following fast and fantastical shows relative to a control play condition (Lillard & Peterson, 2011), but here using new fast and fantastical shows (a different SpongeBob© episode; a new show, Fanboy and Chum Chum©) or an educational show (Arthur©). They then replicated the effect with a longer SpongeBob episode, thought to be more in line with children's viewing habits (20 min), and extended it to an educational show matched in length which was also fast-paced and fantastical (Martha Speaks©). In their final study, Lillard et al. (2015b) fully crossed *Pace* (fast/slow) x *Content* (fantastical/non-fantastical) conditions, by selecting one exemplar show for each combination of levels. Here a condition difference showed poorer EF performance only after viewing fantastical shows, with no effect of pace and no interaction between pace x condition. In this study working memory performance decreased from pre to post-test *only for children who viewed fantastical content*.

While these studies appear to isolate the effects of TV viewing to the presence of fantastical events several confounds remain. The use of automatic software to estimate pacing means the conflation of cuts with other low-level visual feature changes continues to be an issue for the present studies. In addition, there is a degree of ambiguity around the levels of impact on viewer cognition that remains unaddressed. For example, at what level of the processing stream do fantastical events, achieved through a combination of low-level visual changes and higher-level narrative changes, negatively impact viewer cognition? As will now be discussed, we believe an alternative theoretical account of viewer cognition during visual narrative processing (SPECT; Loschky et al., 2020) can be applied as a quantitative content-based analysis of the shows used by Lillard et al. (2011; 2015b) to identify the levels at which content differences may engage and eventually reduce EF immediately after viewing. For the purpose of the present analysis, we focus on the short-term effects on EF, as tested by Lillard et al. (2011; Lillard *et al.* 2015b).

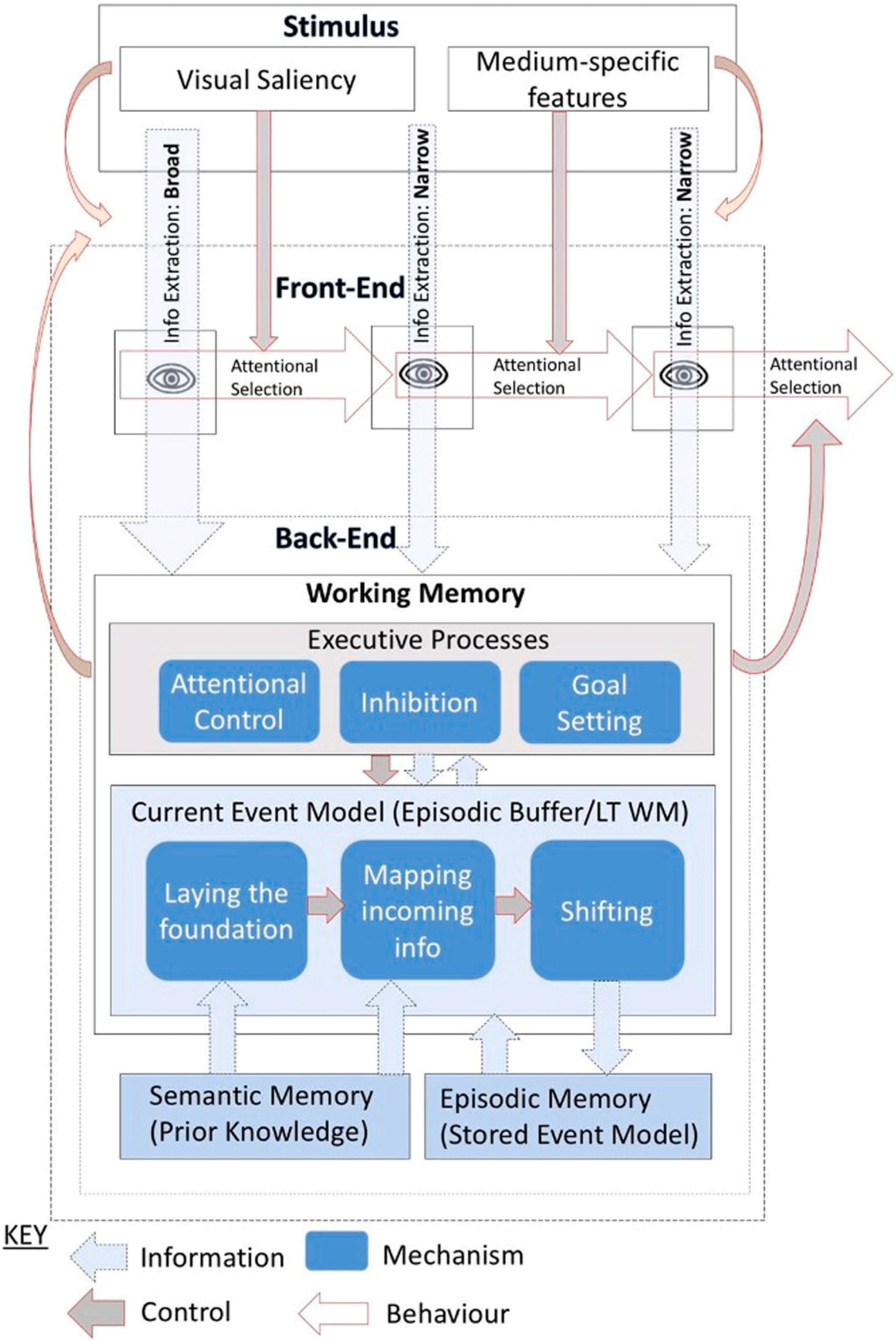
1.3. Models of TV viewing

In order to inform future educational and age-appropriate show creation as well as further our understanding of the factors of screen media influencing EF, we need a theoretical framework that can formalize the content properties of a TV show and help make clear predictions about which shows will have an impact on EF. Various approaches for analysing audio-visual content exist including computational approaches to low-level visual saliency (e.g. Carmi & Itti, 2005, 2006) analysis of discourse structures (Bateman & Wildfeuer, 2014), narrative continuity (Magliano & Zacks, 2011), and editing and cinematographic patterns (Salt, 1992). Theories for understanding how these features may impact various aspects of audience response and cognitive processing have also been proposed, suggesting that the critical processes stem from attention (Smith, 2012), information processing and physiological arousal (Lang, 2000), event comprehension (Radvansky & Zacks, 2011), or the increase in narrative comprehension with age (Pempek et al., 2010).

Lillard and colleagues sketched a detailed model of TV processing that they use to hypothesise mechanisms responsible for their EF depletion effects (see Fig. 2 in Lillard et al., 2015b; adapted from Lillard *et al.* 2015a). Their model (based on (Ohman, 1979) breaks TV processing down into stages from sensory-processing and attending to the stimuli, through encoding and processing of perceived information and ultimate storage in long-term memory. Using this model, they hypothesise two, non-mutually exclusive ways in which - primarily fantastic video content - may impact EF: 1) fantastical events are, by definition surprising and trigger involuntary attention capture limiting the opportunity for practicing top-down attentional control or 2) the problem lies further upstream as information processing resources are exhausted in the attempt to encode and store unfamiliar events for which no prior schemas exist. These hypotheses both seem plausible but are hard to test without a way to formalize the nature and amount of cognitive processing involved by fantastic elements, for example, and the ways in which these may interact.

The Scene Perception and Event Comprehension Theory (SPECT) is a theoretical framework designed to elucidate the stages of cognitive processing involved in visual narrative viewing (Loschky et al., 2020). Unlike previous approaches to theorising dynamic scene processing, SPECT aims to combine theories from multiple traditionally isolated areas, namely scene perception (Henderson & Hollingworth, 1999), event cognition (Radvansky & Zacks, 2011, 2014), and narrative comprehension (Gernsbacher, 1991); Zwaan & Radvansky, 1998). The basic architecture of SPECT distinguishes between stimulus features and front-end and back-end cognitive processes (illustrated in Fig. 1). We refer the reader to Loschky and colleagues (2020) for a full description of SPECT and instead will provide a brief overview here to motivate our subsequent analyses.

SPECT begins with the stimulus. Stimulus features include medium-agnostic features such as the visual features (e.g., luminance



(caption on next page)

Fig. 1. Illustration of the Scene Perception & Event Comprehension Theory (SPECT) theoretical framework of visual narrative processing. The eye icon represents the position of viewer gaze on the stimulus at a particular moment. Further walk through provided in the text.
Figure reproduced with permission from Loschky et al. (2020).

contrast, or motion) that constitute the raw "signal" from which all subsequent sensory and perceptual information is derived. The spatiotemporal distribution of these stimulus features is also important as it is known that regions of high salience (i.e., perceptual qualities which are distinct from surrounding properties due to their colour/motion contrasts) are more likely to attract attention (Itti & Koch, 2001) especially in the absence of a clearly defined viewing task (e.g., Smith & Mital, 2013). Alongside these generic stimulus features, SPECT also considers *medium-specific* features, properties of how a visual narrative is communicated in a particular medium that may constrain/shape the stimulus and its subsequent processing. For example, panels, layout, and action lines in comics (Cohn, 2013), the level of visual abstraction/caricature used in cartoons (Ildirar, Sermin, & Smith, 2018), or shot composition, lighting, framing, depth-of-field, camera movements and editing in film and video (Smith, 2013; Bordwell, Staiger, & Thompson, 2003).

To get from the stimulus to the first stages of narrative comprehension, the viewer must select where to attend in the stimulus and which information to extract. In SPECT, these processes are considered the *Front-End*, occurring during a single fixation/attention period and distinct from the *Back-End* processes involved in the construction of an event model, taking place over multiple fixations and extended durations. *Attention selection* is the prioritisation of particular information over others either in terms of allocating processing resources or overt attention via eye movements. In depth information extraction occurs once an attention target has been selected. The distribution of these targets can be *broad*, capturing gist-like scene information across the visual field or *narrow*, spatially focussed on key details such as object identities, agent behaviours etc., with the shift from broad to narrow typically occurring over the first few seconds after new scene onset (Pannasch, Helmert, Roth, Herbold, & Walter, 2008). As illustrated in Fig. 1, control of both attention selection and information extraction can be *exogenous*, driven by the stimulus such as sudden onsets and motion or *endogenous*, driven by the back-end; the trade-off between the two is known as the *executive processes*. Within SPECT, attention control, inhibition and goal setting are considered executive processes that manage the flow of information between the front and back-ends via working memory and attempt to focus front-end processing on information pertinent to the current event models. We will return to these executive processes for more detailed discussion later given their clear relevance for interpreting Lillard and colleagues (2011; 2015b) executive depletion findings.

SPECT's back-end processes accumulate information across multiple eye fixations and extended time periods. A key back-end process is the construction of a coherent current event model in Working Memory (WM: e.g. the temporary retention of small pieces of information within a limited capacity system), to later be stored in episodic Long Term Memory (LTM: i.e., the storage of episodic information over an extended period; Magliano & Zacks, 2011). SPECT describes three key back-end processes involved in constructing the current event model: laying the foundation for a new event model, mapping incoming information to the current event model, and shifting to create a new event model (Gernsbacher, 1991). Each event model is constructed both from situational information including the time and place in which the events unfold (the spatio-temporal framework), the entities in the event (people, animals, objects), the properties of those entities (e.g., colours, sizes, emotions, goals), the actions of the agents, the unintentional events that occur (e.g., acts of nature), and relational information (spatial, temporal, causal, ownership, kinship, social, etc; Zwaan, Magliano, & Graesser, 1995; Gernsbacher, 1991; Zwaan & Radvansky, 1998) and inferred information absent from the stimuli but sufficiently implied that prior experience and schema knowledge from LTM can be used to construct bridging inferences (Magliano, Zwaan, & Graesser, 1999). For example, a shot of SpongeBob swinging a punch at Squidward followed by a shot of Squidward flying through the air would induce the viewer to draw the inference that the punch landed even without actually seeing this event. New information will continue to be mapped into the current event model as long as it is coherent and allows incremental elaboration (Kurby & Zacks, 2012). When the viewer perceives a discontinuity in the critical event indices *shifting* to a new event model occurs, creating an event segmentation, and storing the old event in LTM (Kurby & Zacks, 2008). This segmentation is critical for understanding complex naturalistic scenes and condensing episodic memory storage (Radvansky & Zacks, 2011). Triggers to the existence of an event boundary can be perceptual, such as sudden changes in visual motion (e.g., a character's postural change or reach for a new object; (Zacks et al., 2006b) or conceptual (e.g., perceived discontinuities in time, space, action, the introduction of new characters, or changes in characters' goal-plans; Kopatich, Feller, Kurby, & Magliano, 2019; Zacks, Speer, & Reynolds, 2009; Zwaan & Radvansky, 1998). Thus, according to SPECT successful visual narrative processing involves a complex co-ordination of perceptual and event model construction processes in WM and LTM, mediated by Executive Processes.

While the need to account for both perceptual and event-based content is largely comparable between SPECT and the Lillard et al. (2015b) model of TV viewing, only SPECT specifies the levels at which stimulus properties interact with the various stages of processing, thereby facilitating the quantification of individual content properties and offering predictions about their respective demands on cognitive resources. To elucidate the levels at which content differences in the videos used by Lillard et al. (2011; 2015b) may have reduced EF immediately after viewing we now apply the SPECT framework (Loschky et al., 2020) as a quantitative content-based analysis.

1.4. Application of the SPECT framework to the study of content properties

Before applying this framework to the investigation of specific content properties in children's TV, it is essential to also consider the developmental context for the mechanisms implicated in SPECT. These mechanisms undergo large functional changes during the preschool years and are not considered adult-like until at least adolescence. As such a characterisation of these skills during early childhood is needed.

Common across developmental studies of EF is the view that foundational components of the EF construct develop during infancy and the pre-school years, and that development of a central attention system underlies the emergence of three core EF components e.g., Working Memory, Inhibition, and Shifting (for summary review's see Garon, Bryson, & Smith, 2008; and also Hendry, Jones, & Charman, 2016). The trajectory of these changes is beyond the scope of this review; however, we briefly highlight key changes to contextualise the challenges posed by the specific content properties operationalised below. While children can selectively attend to features in their environment from infancy, the ability to do so in a goal-directed manner develops rapidly over the pre-school years. These changes in selective attention are due to the development of two attention subsystems e.g., the orienting system and the anterior attention subsystem (Rothbart & Posner, 2001; Ruff & Rothbart, 2001), and are thought to underlie improvements in children's performance across a range of EF tasks (Garon et al., 2008). Initially the control of attention is largely governed by the orienting system which allows young children to orient to features in the external environment and to shift attention. The anterior attention system develops later in infancy with significant improvements occurring between 2 and 6 years of age. By inhibiting and facilitating the orienting attention system, this system selects and enhances processing according to internal representations (Ruff & Rothbart, 2001). As a result, children are increasingly able to selectively attend and focus their attention, making them less susceptible to distractions. This improvement in the control of attention has been seen under naturalistic situations e.g., free-play, and more structured tasks e.g., the continuous performance task. Although, the length of time children are able to maintain focused attention continues to vary significantly during the pre-school years (i.e., Akshoomoff, 2002; Ruff, Capozzoli, & Weissberg, 1998). Closely related to improvements in selective attention is the emergence of the first EF component Working Memory. WM is critical for making sense of events which unfold over time (Garon et al., 2008) and the executive attention system is central to the ability to hold prior events in mind and relate these to what comes later. While infants can hold representations in WM before 6 months of age (Dempster, 1985; Diamond, 1995), the ability to maintain and manipulate representations in memory does not begin to emerge until 2 years and shows protracted development across the pre-school period (Alloway, Gathercole, Willis, & Adams, 2004; Gathercole, 1998). Critical to these age-related improvements in WM is an improved ability to inhibit interference (Hale, Bronik, & Fry, 1997), with co-ordination between the two essential for successful 'shifting'. This final EF component shows the most protracted development, with children struggling to successfully shift from one mental set to another beyond the pre-school years (Davidson, Amso, Anderson, & Diamond, 2006). Thus, over the course of the pre-school years there is an integration of skills within a more organised higher order system. While the organisation of the EF system can be characterised by partially dissociable components at the end of the pre-school years (Garon et al., 2008), there is a heavy cost associated with the complex co-ordination of these components which continues throughout childhood (i.e., Davidson et al., 2006).

Lillard et al. (2015a) lay out two, non-mutually exclusive, hypotheses for the EF depletion effects found in their studies. The first of these pertains to the role of stimulus features in driving attention in a bottom-up fashion. However, a formalisation of the specific features was not specified. As such we specify three stimulus level properties which are understood to drive attention in a bottom-up fashion. By doing so we can make predictions about specific content properties which goes beyond what has been suggested previously and allows these to be tested. One such visual feature is *flicker* (i.e., luminance difference over time). Capturing attention in a bottom-up fashion, *flicker* appears to be one of the strongest exogenous drivers of attention in a scene, with its independent contribution to gaze location higher than the cumulative contribution of all features in models of visual saliency (Itti, 2005; Mital, Smith, Hill, & Henderson, 2011). A further stimulus level property which may constrain subsequent processing is *Edge Density*. The density of edge information, due to heightened texture of depth-of-field, may slow processing by increasing the time it takes to identify meaningful targets within shots (Henderson, Chanceaux, & Smith, 2009). Another feature which will influence the degree to which attention is driven in a bottom-up fashion is *pace* (i.e., frequency of cuts). As well as determining the rate at which new information is presented, cuts re-orient attention to bottom-up properties of the scene such as *flicker or edge density* because they introduce breaks to ongoing event model maintenance. As such, it is a key mechanism through which differences in the structural properties of a video will combine with differences in low-level features to differentially impact the cognitive resources of the viewer.

According to the SPECT framework (Loschky et al., 2020) these properties of the content will constrain front-end processes as children seek to selectively attend to informative locations in the scene and extract pertinent information for narrative comprehension. An increased presence of any, or all of them, will serve to reduce the efficiency of these front-end processes, slowing the early stages of processing and increasing demands on cognitive resources. We therefore predict greater levels of *Flicker*, *Edge Density*, and *Cuts* will be found in videos previously shown to deplete EF immediately after viewing. To investigate the impact of cuts frequency we extracted the duration of each shot which is then averaged over all shots to give the Average Shot Length (ASL) for every video. The shorter the shot durations the faster the delivery of new information. Previous measures of *Pace* have conflated *flicker* and *cuts*. If ASL, rather than low-level stimulus properties, contributed to the reported effects of "pacing", we would expect the "fast-paced" videos identified by Lillard and colleagues (2011; 2015b) to have shorter ASLs than the "slow-paced" shows. Alternatively, if the pacing effects were driven by low-level visual properties, we may expect to find the "fast-paced" videos do not differ from the "slow-paced" videos on ASL but contain greater levels of *flicker* and *edge density*.

The second hypotheses proposed by Lillard and colleagues (2015b) posits that EF depletion effects may be the result of an overwhelmed EF system which has been unable to successfully align novel screen events with existing representations in memory. As discussed, within the SPECT framework a set of back-end processes are responsible for the successful creation and maintenance of these representations. The allocation of resources to these processes are understood to differ depending on the rate of situation change in the content (Zacks, Speer, Swallow, Braver, & Reynolds, 2007a). Discontinuities across key event indices introduce unpredictability to ongoing perception, known as prediction errors, which eventually trigger a shift in the event model . During periods of low prediction error, pathways from sensory inputs are thought to be inactive and stable, thereby conserving resources, while more intensive processing occurs during periods of increased prediction errors (Zacks et al., 2007a). This view has been supported by a number of studies

Table 1
Summary of exemplar videos used by Lillard & Peterson (2011) and Lillard et al. (2015b) to be included in analysis. Durations of videos obtained for the present analysis are presented in minutes and decimal seconds, with the durations reported by Lillard et al. *(2011; 2015b) presented bracketed in bold and italicised. Differences in durations between our own analysis and the original studies are due to unspecified shortening of episodes by Lillard* et al. *(2011; 2015b). As such entire episodes have been included in this analysis.*

Episode	Paper	Study	Condition	Video Duration	Frame Rate (/second)	Significant Effect compared to control condition at p < .05
Caillou- April Fools	1	–	Slow	7.83 min ***(9.00)***	29.97	No
SpongeBob SquarePants-Fools in April	1	–	Fast	10.60 min ***(9.00)***	25	Yes
Arthur and DW Clean Up	2	1	Slow	11.42 min ***(11.02)***	30	No
Arthur and DW Library Card	2	1	Slow	11.40 min ***(11.02)***	30	No
Fanboy and Chum Chum-Fan-Boy in a Plastic Bubble	2	1	Fast Fantastical	12.00 min ***(12.08)***	30	Yes
SpongeBob SquarePants-Doing Time	2	1	Fast Fantastical	11.22 min ***(11.08)***	30	Yes
Martha Speaks-Return of the bookbots	2	2	Fast Fantastical-Educational	21.47 min ***(22.48)***	30	Yes
SpongeBob SquarePants-Pest of the West	2	2	Fast Fantastical	22.47 min ***(22.50)***	30	Yes
Little Einsteins-Flight of the instrument fairies	2	3	Slow Fantastical	18.62 min ***(8.57)***	30	Yes
Phineas and Ferb- Flop Stars	2	3	Fast Non-Fantastical	7.90 min ***(7.68)***	30	No
SpongeBob SquarePants- Bad Guy Club	2	3	Fast Fantastical	8.28 min ***(8.02)***	30	Yes

using the Event Indexing Model (Magliano, Miller, & Zwaan, 2001) to assess narrative understanding. For example, Zacks and colleagues (2009) have shown changes in situational properties are highly predictive of perceiving event boundaries, and periods of high prediction error are associated with increased processing, as indicated by slowed reaction times during sentence reading.

Moreover, perceiving event boundaries where they are not (e.g. when high *flicker* does not actually correspond to an event transition) may trigger erroneous event segmentation leading to premature closure of previous events and interfering with event processing. Consequently, unfamiliar or immature viewers are overwhelmed with the flow of unstructured, unrepresented events and their comprehension fails. For example, studies of event perception and memory have shown successful event segmentation to be key for subsequent recall and recognition of events (Swallow, Zacks, & Abrams, 2009). An effect which holds true even after controlling individual differences in a number of key cognitive domains (i.e., processing speed, working memory, crystallised knowledge and laboratory episodic memory; Sargent et al., 2013). Furthermore, the relationship between the two appears to be causal, with targeted interventions of event segmentation leading to improved event memory (Flores, Bailey, Eisenberg, & Zacks, 2017). Thus, if the young viewer has been unable to successfully build event representations and shift these to LTM, they will be lacking in the key predictive knowledge they would need to support maintenance of ongoing event models as viewing progresses.

A secondary implication of monitoring situational change pertains to the memory processes responsible for storing ongoing event information in WM. According to Event Segmentation Theory (EST; Zacks, Speer, Swallow, Braver, & Reynolds, 2007b), the amount of information contained within active event models will likely exceed what can be maintained within a limited capacity system. To overcome this, capacity within WM may be augmented by prior knowledge stored in 'event schemata' in LTM (e.g., representations of locations already visited, tracking of characters [including their spatial positions], common character behaviours etc), which effectively expands the capacity by storing predictive information about the future relevance of aspects of the events. However, while adults are assumed to have amassed sufficient experience with a wide range of events to facilitate such a mechanism, children may lack the ability to build such event schemata.

This is a critical consideration for the role of fantastical elements. Assuming this level can be isolated from low-level saliency/pacing differences, these events will challenge event model construction and maintenance as prior knowledge and expectation is critical to allow efficient event representation (Bläsing, 2014; D. Levine, Hirsh-Pasek, Pace, & Golinkoff, 2017; Sargent et al., 2013; Zacks, Speer, Vettel, & Jacoby, 2006a). The novelty of each scene/action will make it difficult for the new information to be mapped into the existing event model, triggering a shift and laying the foundations for a new event model. Even if we are sufficiently familiar with the events presented, we still need to track a variety of critical situational details (changes in the actions agents carry out, or in the scene angle, or in the presence/absence of objects) to identify when one event ends and a new one begins. It seems justified to assume that for fantastical events there is a premature updating of an event model either through the erroneous detection of event boundaries due to low-level visual changes, because unrecognisable events will give rise to shifting of event models, or through a combination of the two. The "pacing" tracked by Lillard et al. (2015b) with automated change detection cannot tease apart low-level visual changes that signal or not changes in these situational details. Thus, we examine the degree to which the videos used by Lillard et al., (2011; 2015b) may differ in their rate of situational change and predict a greater degree of situational change for those shown to reduce EF performance immediately after viewing (Lillard et al., 2011; 2015b).

Finally, we consider the cumulative impact of these properties on cognitive resources. Our discussion thus far has treated specific content properties as dissociable properties of the content which can be assessed independently. However, clearly these properties co-

Table 2
Summary results for the measures of interest are presented together with the original measures of Pace and Fantasy reported by Lillard and colleagues (2011; 2015b) highlighted in grey. Columns highlighted in blue indicate EF depleting shows as identified by Lillard and colleagues (2011;2015b).

	Lillard & Peterson (2011)		Lillard et al., (2015)								
	Caillou-April Fool	Sponge Bob- Fools in April	Arthur-CleanUp	Arthur-Library Card	Fanboy and Chum Chum-In a Plastic Bottle	Sponge Bob-Doing Time	Martha Speaks-Return of the Bookbots	Sponge Bob-Pest of the West	Little Einsteins-Flight of the instruments	Phineas and Ferb-Flop Starz	Sponge Bob- Bad Guy Club for Villains
Condition	Slow	Fast Fantastical	Slow Slow		Fast Fantastical	Fast Fantastical	Fast Fantastical	Fast Fantastical	Slow Fantastical	Fast Non-Fantastical	Fast Fantastical
Effect	No Effect	Negative No	Effect No Effect		Negative	Negative	Negative	Negative	Negative No Effect		Negative
L&P Pacing (Scenes per minute)* **	1.76	5.45	22	22	38	35	41	42	11	41	45
L&P Fantastical Rate (events per minute)	NA	NA	0	0	3.31	3.25	2.36	1.56	1.28	0.13	3.99
Average Shot Length in MS (SD)	3773.414 (1765.98)	3493.187 (2227.95)	3323.461 (1820.32)	3240.137 (1953.50)	5002.66 (4371.27)	4578.592 (3487.91)	4323.46 (3836.52)	4122.514 (3368.811)	7300.654 (7457.59)	3001.418 (3256.95)	3087.578 (3554.90)
Discontinuous Edits (As % of Total Edits)	25.45	37.57 31.22	30.00		42.66	45.21	45.12	43.56	38.82 51.59		48.75
Mean Flicker With Cuts *(SD)*	63376.00 *(229650.00)*	109910.00 *(282970.00)*	70979.00 *(192820.00)*	61492.00 *(180540.00)*	74844.00 (224270.00)	91367.00 (209520.00)	136290.00 (297800.00)	70179.00 (189180.00)	68433.00 (174460.00)	87574.00 (304240.00)	92242.00 (265730.00)
Mean Flicker Without Cuts *(SD)*	53174.77 (212860.60)	104503.2 (272857.40)	71812.89 (193494.40)	62567.38 (181484.30)	74787.85 (224417.90)	91315.52 (206936.30)	134462.7 (282156.70)	70299.16 (187877.70)	68591.62 (174921.60)	79554.99 (240668.70)	90982.68 (240074.30)
Mean Edge Density *(SD)*	5496.10 (1118.90)	7983.20 (1487.10)	6377.00 (903.64)	6035.00 (797.69)	5856.20 (878.75)	5675.00 (1028.30)	7289.90 (1398.10)	5688.70 (971.50)	5864.90 (1352.60)	5540.10 (1157.20)	5445.30 (913.68)
Z Score ASL (Reverse)	0.27	0.50 0.64 0.70			-0.72	-0.38	-0.17	-0.01	-2.57 0.90		0.83
Z Score Discontinuous Edits	-1.76	-0.29 -1.06	-1.21		0.32	0.63	0.62	0.43	-0.14 1.41		1.06
Z Score Flicker (With Cuts)	-0.92	1.13 -0.58	-1.00		-0.41	0.31	2.29	-0.62	-0.70 0.15		0.35
Z Score Flicker (Without Cuts)	-1.27	0.99 -0.45	-0.86		-0.32	0.41	2.32	-0.52	-0.59 -0.11		0.40
Z Score Edge Density	-0.76	2.30 0.32	-0.10		-0.32	-0.54	1.45	-0.52	-0.31 -0.71		-0.82

* **NB Measures of pacing differ between Lillard & Peterson (2011) and Lillard et al., (2015b). In Lillard & Peterson (2011) pacing was hand coded based on the number of complete scene changes that occurred, while in Lillard et al., (2015b) pacing was reported as number of scenes per minute as estimated by an automatic shot detection tool. For the purposes of this table pacing is presented as scenes per minute for both Lillard & Peterson (2011) and Lillard et al., (2015b)

occur in any given video and are therefore likely to assert a combined influence on the processes which may underly deficits in EF after viewing. As such, for our final analysis we assess a composite feature score containing our individual measures (i.e., *Average Shot Length, Situational Change, Flicker, Edge Density*). This composite score represents a hypothesis driven accumulation of the features thought to impact EF, with a positive cumulative Z score indicating a greater prevalence of these features. We expect to find higher scores for videos previously shown to deplete EF after viewing, compared to shows which were not shown to deplete EF after viewing.

The methodology for quantifying each of the properties outlined above is now presented, followed by analysis results and discussion.

2. Method and Materials

2.1. Exemplar Videos

For the present analysis exemplar videos used by Lillard et al. (2011; 2015b) were sourced from online channels in mp4 format, except for Little Bill- All Tied Up© which was not readily available. This episode was therefore not included in the analysis. In line with details provided by Lillard et al. (2015b) one the exemplar videos (Phineas and Ferb©) was edited to remove a fantastical subplot using QuickTime Player (v. 10.5). Aside from this edit specified by Lillard et al. (2015b) all videos were used in full for the present analysis. As indicated in Table 1 some episodes differ in duration between our analysis and the original durations provided by Lillard et al. (2011; 2015b). This is due to unspecified shortening of episodes in some of the studies. Videos under investigation are presented in Table 1 below.

2.2. Measures

Four measures categorising the content properties of interest are used in the present analysis: *Flicker*, *Edge Density*, *Average Shot Length*, and *Situational Change*. *Flicker*, representing the change in luminance over time, was computed from the normalised resolution (256 ×256) and the full range of luminance change at each pixel between corresponding pixels of adjacent frames. For the purpose of the present analysis luminance difference values were averaged across all frames to give *Mean Flicker* for the entire video. This is the closest match to the Lillard et al. (2015b) measure of pacing. However, due to a possible conflation of *flicker* with cuts (i.e., editorial transitions from one shot to another) in existing dynamic scene viewing studies (Mital, Smith, Hill, & Henderson, 2011), we also computed a measure of *flicker* excluding cuts by removing the luminance difference values from two frames preceding each cut to one frame post cut (*Mean Flicker-without cuts*).

The second stimulus feature investigated in the present analysis was *Edge Density*. Edge Density was extracted by calculating the percentage of pixels that were edge pixels for each frame of the video and then averaging them over all frames. Both measures were abstracted using the computer vision toolbox in Matlab (v.9.5). All videos were downsampled to 256 × 256 pixels to ensure videos were compared on a standard resolution. Videos were first converted into CIELab color space separating luminance from colour before calculating *Flicker* and *Edge Density*.

To abstract *Average Shot Length* and rate of *Situational Change* we first manually identified cuts using Datavyu. For *ASL* we averaged individual shot durations across all shots for each video. Secondly, by applying the Event Indexing Model (Magliano et al., 2001) to these editorial transitions we were able to quantify the degree of situational change in each video. All videos were coded from the end of the opening credits to the beginning of the end credits, with the shot preceding and following each cut used to determine whether the cut was continuous/discontinuous on the following dimensions. ***Action***- whether the action preceding the cut is continuous/discontinuous after the cut; ***Spatial Region***- whether the spatial region preceding a cut is continuous/discontinuous after the cut (i.e., when the shot following a cut introduced a new region or reintroduced a previous spatial region but is a new region from the previous shot, this would be considered discontinuous in spatial region); ***Spatial Movement***- whether the spatial movement preceding a cut is continuous/ discontinuous after the cut (i.e., when there is a change in the location of a main character within their world); ***Narrative Time***- whether the narrative timing of a shot is continuous/discontinuous following a cut (i.e., when there is a chunk of time missing from the story world between the shot preceding a cut and the shot following a cut, this would be discontinuous in narrative time). We refer the reader to Magliano and Zacks (2011) for a more detailed description of each of the dimensions. *Situational Change* was calculated as the rate of edits which were discontinuous in Action, Space, or Narrative time as a proportion of all edits (number of discontinuous edits/total number of edits). For both measures a second coder coded > 25% of the exemplar videos (3x Videos: Arthur- Library Card ; Phineas and Ferb- Flop Starz; SpongeBob SquarePants- Bad Guy Club for Villains). Coder reliability was assessed with Intraclass Correlations (ICC). ICC estimates and their 95% confidence intervals were calculated using SPSS statistical package (v.24) based on absolute agreement, in a 2-way random effects model. These Intraclass Correlations showed excellent reliability for *ASL* (0.97), and good reliability for *situational change* (0.83).

Finally, we assessed the cumulative presence of these features with a composite feature score. Derived by summing individual Z scores for *ASL*, *Mean Flicker (with cuts)*, *Mean Edge Density* and *Situational Change* for each video. As longer shot durations are considered to be better than shorter shot durations, the *ASL* values were reverse scored. Summary results are presented in Table 2.

3. Results

For the purpose of the analysis each video is treated as an individual observation from which the measures of interest are abstracted and inspected at the group level (i.e., 'EF depleting' vs 'Non EF depleting' videos). Due to the limited number of sample videos and the mismatch in number of videos for the two categories under investigation differences cannot be tested statistically. As such reported results are purely descriptive. Results are presented in Fig. 2 & 3 below.

3.1. Stimulus Features and Front-End Processes

Flicker was examined as an average across all shots for each of the videos. We found greater average luminance difference for shows previously found to deplete EF immediately after viewing (*Mean:* 91895.00, *SD:* 234847.14), than for 'Non EF depleting' shows (*Mean:* 70855.25, *SD:* 226812.50) as shown in Fig. 2a. The video with the lowest level of flicker was Arthur- Library Card (*Mean:*70979.00, *SD:*192820.00) while the video with the greatest level of flicker was Martha Speaks- Return of the Bookbots (*Mean:*136290.00, *SD:*297800.00). This pattern held true when cuts were excluded, with a higher average luminance difference for 'EF depleting' shows (*Mean:* 90706.10, *SD:* 23324.99) than for 'Non EF depleting' shows (*Mean:* 66777.51, *SD:* 11421.89). We next assessed the image complexity of each frame with *Edge Density* by calculating the percentage of pixels that were edge pixels, averaged over all frames (Fig. 1b). Here we found *Edge Density* values were higher for 'EF depleting' shows (*Mean:* 6257.60, *SD:* 1147.15) than for 'Non EF depleting' shows (*Mean:* 5862.05, *SD:* 994.36) as shown in Fig. 2b.

Next, to address the possible conflation of these low-level features with pacing we quantified pace for each video by calculating Average Shot Length (Fig. 2c) for each of the videos. Counter to what we had expected from the pacing values provided by Lillard et al., (2011; 2015b) we found longer shot durations on average for 'EF depleting' shows (*Mean:*4558.38 ms, *SD:*1562.50) than for 'Non EF depleting' shows (*Mean:*3334.61 ms, *SD:*781.18). This pattern also held true when the apparent outlier 'Little Einsteins' was excluded (Mean: 4101.33 ms, SD: 705.24 ms). This discrepancy may arise from differing methodology between our analysis and the studies conducted by Lillard and colleagues (2011; 2015b). We will return to this discrepancy in the discussion. *Mean Flicker* values with and without cuts, *Mean Edge Density* values and *ASL* values for each video are summarised in Table 2.

3.2. Back-End Comprehension Processes – Situational Change

We now consider the influence of content properties on the SPECT (Loschky et al., 2020) defined ***back-end processes***. The relative frequency of shifts in key event indices were calculated as a proportion of all edits for each video, averaged across the two types of videos ('EF depleting' vs 'Non EF depleting'). As shown in Fig. 2d we found a greater proportion of edits containing discontinuous shifts for 'EF depleting' shows (*Mean*: 43.10, *SD*: 3.87) than for 'Non EF depleting' shows (*Mean*: 34.69, *SD*: 11.56). Thus, videos previously shown to reduce EF performance after viewing were found to contain more situational change.

3.3. Composite Feature Scores

Finally, we computed a composite feature score for each of the videos using the variables *Flicker*, *Edge Density*, *ASL*, and *Situational Change*. The results of this analysis are presented in Fig. 3. We found a general pattern of positive composite scores for 'EF depleting' videos (*Mean*:.53, *SD*:2.78), and negative composite scores for 'Non EF depleting' videos (*Mean*: −0.93, *SD*:2.06) see Fig. 3a. This pattern held true when the composite score was calculated using *flicker* without cuts (*Mean*:.58, *SD*: 2.72 for 'EF depleting' videos; *Mean*: −0.26, *SD*: 2.08 for 'Non EF depleting' videos). The direction of the summed z scores for the two types of videos ('EF Depleting' vs 'Non EF depleting') indicate a greater prevalence of features thought to tax cognitive resources for the 'EF depleting' videos than for 'Non EF' depleting videos.

4. Discussion

There is a widely held view in the developmental literature that a causal link exists between TV viewing and differences in children's developing EF skills. Yet a theoretical framework which can formalise specific psychological mechanisms activated by specific TV viewing content has been lacking. It was therefore the aim of this analysis to develop a quantitative content-based approach to examining individual content properties in children's TV shows, from the perspective of event comprehension theory which could be applied to videos previously used to investigate the effects of TV viewing on children's EF.

This analysis found a pattern of greater stimulus saliency, increased situational change, and a greater combined prevalence of features considered to be most demanding in terms of cognitive resources, for videos previously found to reduce EF after viewing (Lillard & Peterson, 2011; Lillard et al., 2015b). We fully acknowledge that, given the small sample of videos, these are tentative differences which have not been tested statistically.

In line with our first prediction, videos previously shown to reduce EF after viewing contained greater levels of the visual feature *flicker*. This finding appears consistent with the proposal that differential effects of TV viewing on children's EF may, in part, be due to the presence of stimulus features which act as exogenous drivers of attention. By increasing the saliency in the scene, higher levels of *flicker* will bias attention to bottom-up processing. Further, the power of *flicker* to drive attention in this way may be amplified by the presence of *cuts* which re-orient attention toward salient properties of a shot. As such, *flicker* and *cuts* can be considered highly related in terms of their impact on processing. Within the framework presented here the constraints these features may place on processing

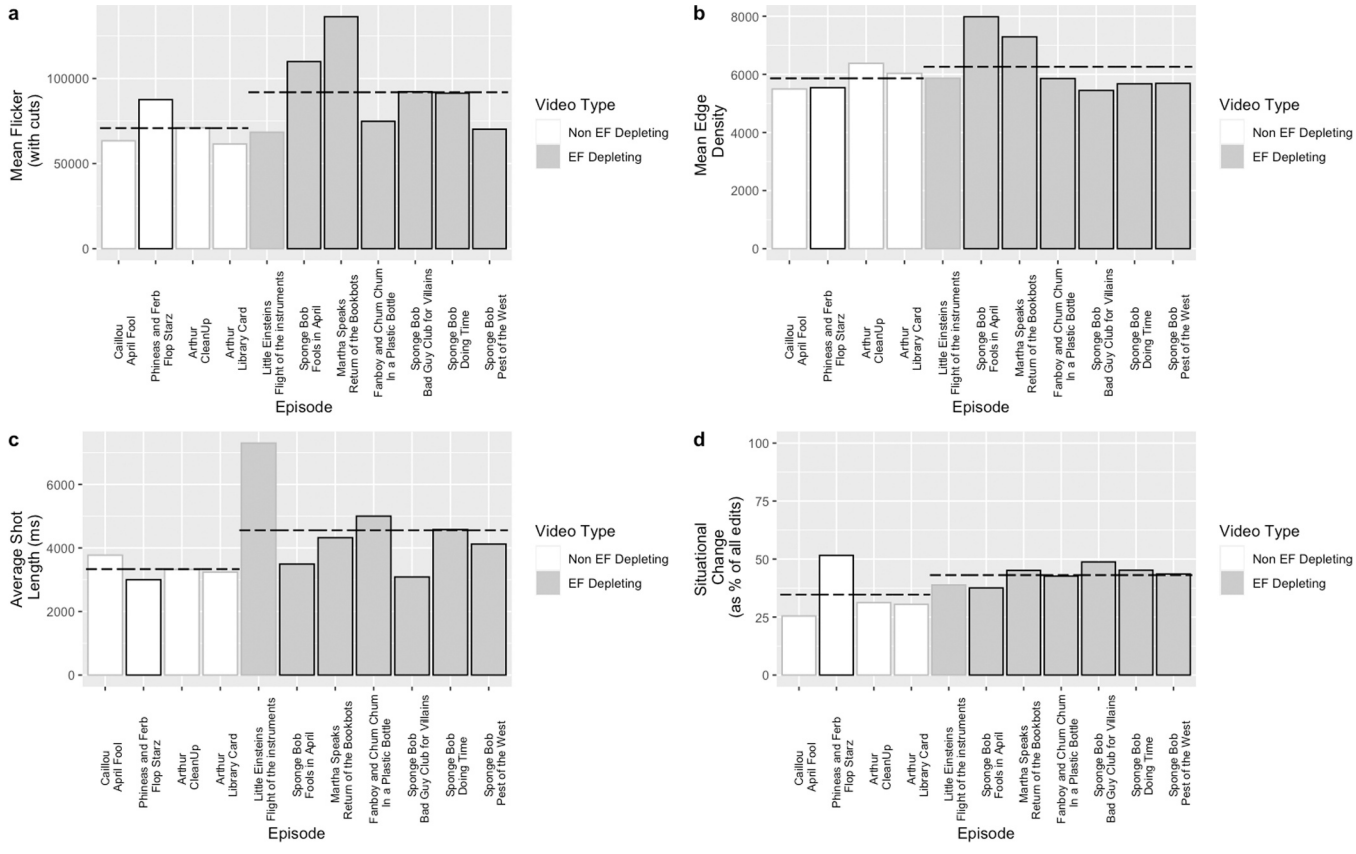


Fig. 2. The results of the content property analyses are presented in Plots a-d. Plot a) shows the Mean Flicker (with cuts) for each video; Plot b) shows the Mean Edge Density for each video; Plot c) shows the Average Shot Length for each video; Plot d) shows the rate of situational change for each video as a proportion of all edits. Dashed horizontal lines represent the group average according to video type (i.e., EF depleting vs Non EF depleting shows). The pacing for each of the videos, as specified by Lillard et al. (2011; 2015b), is indicated by the bar outline- 'fast-paced' videos are shown with a black outline , 'slow-paced' videos are shown with a grey outline.

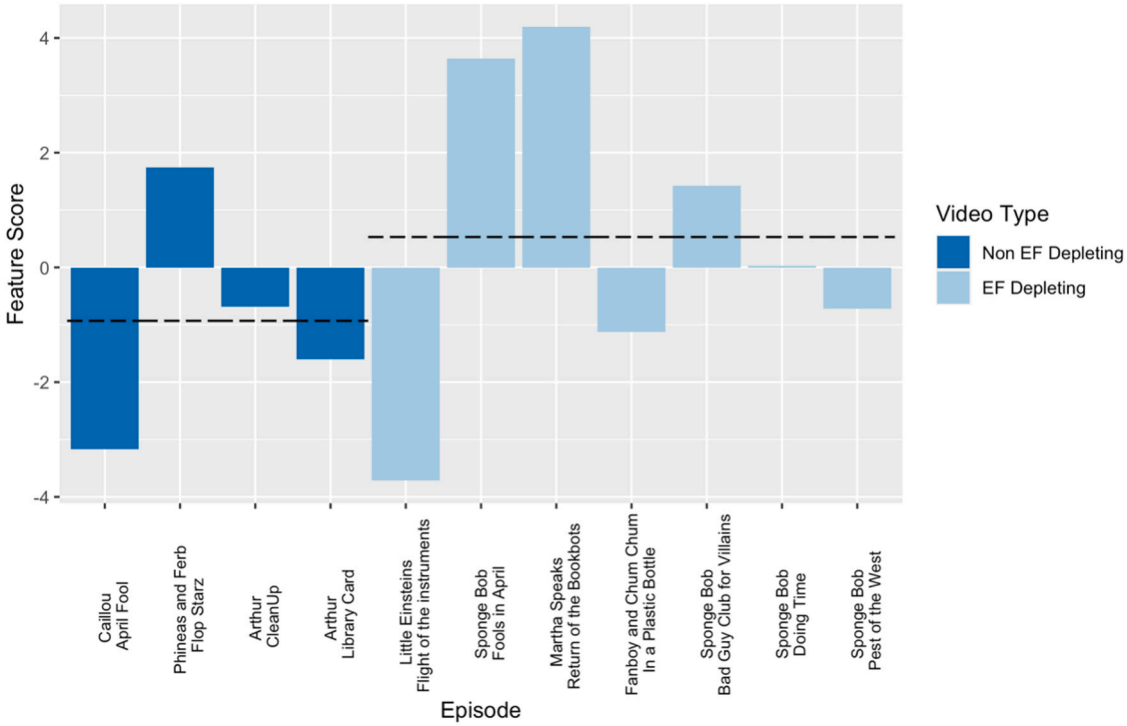


Fig. 3. Results of the composite feature score analyses are presented. Each bar represents the summed z score (ASL, Mean Flicker (with cuts), Mean Edge Density, and Rate of Situational Change) for that video. Bars are shaded according to the type of effect found (i.e., Non EF depleting vs EF depleting).

capacity were also specified. By signalling the need to update/shift an event model the levels of both features will directly influence the demand placed on mechanisms within working memory responsible for the maintenance and creation of these models. A higher presence of either will increase the frequency with which event models need to be updated or shifted to new models. Thus, for 'EF depleting' videos a greater prevalence of *flicker* in combination with *cuts* would not only bias the viewer to bottom-up processing but also place a greater burden on these higher-level working memory processes.

That said, the absence of an effect on children's EF for the 'non EF depleting' videos could be taken as evidence that there may be a level at which these features do not challenge processing to this extent. While lower levels of *flicker* and/or *cuts* will lessen the need to update/shift event models, their presence will still necessitate maintenance of event models. A study by Wass and Smith (2015) suggests it may be important to also consider how *flicker* manifests in shots. In a comparison study of children's and adult's TV content, *flicker* was found to be more localised to elements of the shot which are key for narrative comprehension (i.e., speaking character) in children's content. While in adult content it was more widely distributed across shots. In our own analysis a higher average *flicker* is indicative of a greater dispersal of flicker across the whole frame, i.e., less focus. As such the higher levels of flicker for 'EF depleting' videos may be suggestive of a more adult like pattern, akin to that seen in the study by Wass and Smith (2015), supporting our earlier supposition that with higher levels of *flicker* abstraction of key elements for narrative comprehension will be more challenging and will increase the likelihood that working memory mechanisms will detect the need to shift to a new event model. Thus, it will be key for future studies to assess both level of *flicker* and localisation of *flicker*.

A secondary aim of the feature level analysis was to establish the degree to which low level features may have contributed to measures of pacing in previous studies. We had hypothesised that if speed of content delivery (measured here with Average Shot Length), rather than low-level stimulus properties, contributed to the effects reported by Lillard and colleagues (2011; 2015b) videos identified as "fast paced" would have shorter ASLs than those identified as "slow-paced". Alternatively, if the pacing effects were driven by low-level visual features, we predicted greater levels of our two low-level visual properties (*flicker* and *edge density*) would be found for "fast-paced" videos, while ASLs would not differ. Our results were largely consistent with the latter of the two. We found higher levels of *flicker*, and *edge density*, for videos previously identified as "fast-paced". This is consistent with the view that commercial cut detection tools incorrectly detect shot boundaries at points where there are sudden changes in luminance signal (De Santo, Percannella, Sansone, Santoro, & Vento, 2002) and supports the suggestion that measures of pacing used in the developmental literature are in fact a conflation of low-level visual features and editorial actions (Kostyrka-Allchorne, Cooper, & Simpson, 2017). However, it is important to also note that we found a difference in ASL between "fast-paced" and "slow-paced" videos which appears inconsistent with our prediction and with the existing literature. The results showed shorter shot durations, typically equated with a faster pace of content delivery for "slow-paced" videos, and longer shot durations which are normally indicative of a slower pace of content delivery for "fast-paced" shows. Although our chosen methodology (hand coding) differs from the existing measures of pacing

used by Lillard et al. (2011; 2015b), ASL has been widely used in the existing adult literature as a measure of pace. We therefore believe this is an accurate quantification of the pace of delivery for the videos analysed. The absence of an effect for videos with shorter shot durations may suggest "pace" is less detrimental for children's EF than previously thought. With the combination of longer shot durations and greater levels of *flicker* for the 'EF depleting' videos pointing to *flicker* (or its correlate, motion) as the more powerful influence on processing resources. However, the imbalance in the number of videos in each of the categories in the present analysis prevents us from drawing any strong conclusions. It will be crucial for future work to ascertain whether this finding extends to other content choices and if so, what the relative impact of "pacing" may be when featural properties such as *flicker* are taken into account. Further, while the current analysis has concentrated on the power of these features during video viewing, this analysis would also extend to the study of content choices which are on in the periphery e.g., background TV. Studies which suggest background TV disrupts focused attention in infants (Setliff & Courage, 2011) and may be particularly detrimental for self-regulation, language and literacy skills (e.g., Ribner, Barr, & Nichols, 2021) currently lack systematic measurement of the features within the content. With the present framework it would be possible to also determine whether low-level visual features such as flicker and cuts assert an influence on the reported effects found for background TV.

Next, the degree to which videos differed in their rate of *situational change* was assessed. In line with our prediction, we found videos previously shown to reduce EF after viewing contained a higher rate of *situational change*. This finding not only lends support to the hypothesis proposed by Lillard et al. (2015b) which posits that EF depletion effects may be the result of an EF system which has been overwhelmed by the task of processing unfamiliar novel screen events. But also, offers the field a framework with which it will be possible to test their impact empirically.

A formalisation of the properties which would underly novel screen events has been lacking in existing studies of short-term effects of TV viewing on children's EF. With the present framework we posit that tracking key situational information is essential for the successful alignment of novel screen events with representations in LTM. We theorised several important ways in which tracking *situational change* could impact processing resources. Periods of high *situational change* (i.e., changes across time and space) which signal event boundaries give rise to an increased allocation of processing resources. Detecting these boundaries is key for segmenting the flow of action into meaningful units which can be shifted to LTM as new event models are created. However, where there is sufficient presence of signal change from a featural property (i.e., *flicker*), an event boundary may be erroneously triggered. Consequently, unstructured and unrepresented events, which lack the predictive knowledge necessary to support the maintenance of event models in WM, may be shifted to LTM. With only disrupted representations to draw on in LTM, maintenance of ongoing event models will become compromised and ultimately comprehension will fail.

It seems plausible that 'fantastical events' - which may be achieved through a combination of *situational change* and changes in low level featural properties such as *flicker*- may be particularly intensive in terms of cognitive resources. However, as fantastical events were not treated independently by this analysis future work is needed to establish whether this pattern is amplified for fantastical events given their semantic complexity or common across all rapid event sequences. If this pattern is confirmed, several questions should be addressed empirically. Firstly, do these fantastical events represent a specific class of events in children's content, as posited by Lillard et al. (2015b), which can entirely account for the negative effects of these videos. The frequencies of fantastical events reported by Lillard et al. (2015b) suggests these are still relatively infrequent within episodes and therefore their presence may not be sufficient to account for the detrimental effects in EF after viewing. Secondly, if the issue with these events concerns the integration of current event information held in WM with predictive knowledge about future events in LTM, as our framework would predict. Why do children not update representations to incorporate fantastical events already seen? For example, SpongeBob's sponge-like body is often impossibly moulded into the shapes of objects he collides with. Why would children not update their representations of SpongeBob to incorporate these frequently seen transformations? Irrespective of the nature of novel screen events – fantastical or otherwise - the results of this analysis suggests the rate of *situational change* in combination with levels of *flicker* may be critical for understanding the impact on children's EF after viewing.

It is clear, however from the results discussed thus far that while individual content properties will influence cognitive resources at specific stages of processing, they will also combine to influence cognitive resources throughout the processing stream. The final step of the present analysis was to assess the accumulation of features, considered by this framework to be most demanding in terms of cognitive resources (i.e., *Flicker, Edge Density, ASL, & Situational Change*). As expected, this analysis found a greater prevalence of these features, as indicated by a positive composite feature score, for most 'EF depleting' videos. A number of observations warrant consideration. Firstly, our feature-level analysis showed ASL differed between the two types of shows counter to what had been predicted. Despite this difference we still found a pattern of greater combined presence of cognitively demanding features for 'EF depleting' videos. This lends further support to the possibility that pace on its own may be less problematic than initially thought and suggests a composite feature score may be a useful metric for assessing relative influences of individual properties. It is also clear from this analysis that not all videos follow the trend for their respective effect types. Phineas and Ferb© for example had a positive feature score although the study by Lillard et al. (2015b) found no effect on EF for this show. Similarly, Little Einsteins© was shown to be detrimental for children's EF by Lillard et al. (2015b) but has a negative feature score. While this could call in to question the validity of a cumulative feature score, we suggest the use of this measure may help to identify content choices which differ from a general pattern. To confirm the utility of a composite feature score future work should ascertain whether features make independent contributions or tend to co-vary across different content choices. In the event independent contributions are identified this score could potentially shed light on unexplored properties which may be most detrimental for young viewers and those which may offer a buffer from increased processing demands. Building an understanding of individual content properties in this way gives an opportunity to directly inform developmentally appropriate content creation whilst also furthering our theoretical understanding of the psychological mechanisms activated by specific properties of the content.

As outlined, translating the narrative world shown on screen into meaningful units is a complex process involving the detection of stimulus cues and integrating these with existing knowledge stored in memory. With experience we become accustomed to this unique scenic language and for adults the successful parsing of narrative information occurs largely outside of awareness. However, for the inexperienced viewer this complex process presents many challenges for abstracting meaning from what they are seeing. A theoretical framework for understanding and directly testing this impact has been lacking in the developmental literature. SPECT (Loschky et al., 2020), as applied by this analysis, has provided such a framework. Quantifying the featural and conceptual levels of the videos has allowed us to present a theoretical account of how specific content properties may account for short-term differential effects of TV viewing on children's EF. However, we acknowledge that while this framework improves on the current theoretical understanding of individual content properties and their impact on children's cognitive resources during viewing, our present work can only serve as a pilot demonstration of the potential efficacy of this multi-level approach to content analysis. We have been limited by the number of sample videos which has prevented us from testing the patterns statistically. Future work is needed to establish whether these are meaningful differences and whether they can reliably account for differences in EF after viewing. With a larger sample of videos and their impact on EF we would be able to construct statistical models of the independent variance in EF outcome predicted by each feature and the correlations across feature levels. Such analyses would require a massive data gathering exercise far exceeding the traditional developmental approach of using single or a limited set of exemplar TV shows and time-consuming, in person behavioural assessments of EF.

A further limitation concerns the level at which individual content properties can be quantified. With this framework we have successfully quantified low-level featural properties (*flicker, cuts, edge density*) and mid-to-high level properties (*situational change*). This has been achieved through a combination of automated and human coded processes. However, there remains a host of more complex properties which have not been tracked by this analysis which may prove more difficult to assess. It may be the case that complex semantic aspects of the content (i.e., character intentions) cannot be tracked automatically and will require manual coding to abstract them. In its current form the time cost of this may limit the frameworks utility for informing future educational and age-appropriate show creation, by preventing the scaling up needed to assess greater volumes of content. It may also be the case that empirically testing more complex properties of the content would be difficult with young children. For example, *situational change* which was tracked by this analysis should predict event segmentation (Zacks et al., 2007a). Investigating the link between the two will be key for establishing the degree to which segmentation may be disrupted in children when the degree of change in situational properties and/or stimulus level properties is high. However, subjective measures of event segmentation which are widely used in the adult literature would not be suitable for young children as they cannot be expected to follow the instructions necessary to actively mark event boundaries or comprehend the same events. As such, adaptation of current methodology will be necessary to successfully apply the present framework to the field of developmental research.

In conclusion, the current analysis presented a theoretical framework in which individual content properties were formalised and quantified to further our understanding of the factors of screen media which impact children's EF during viewing. Three content properties were found to be more prevalent in videos previously shown to reduce EF after viewing (*flicker*, *edge density*, *situational change*). In addition, a cumulative feature score showed EF depleting videos typically contained a greater combined presence of cognitively demanding properties. This research makes several key contributions to the field. Firstly, it offers a framework with which it will now be possible to make predictions and test empirically the impact of specific content properties. Secondly, it expands the current theoretical understanding of the properties which are most influential in terms of cognitive resources. Future research should seek to apply this framework to a wider range of content choices, testing the relationships between properties empirically to identify the mechanisms which sustain effects of TV viewing in the short term, while striving to establish whether a causal link exists between short-term effects and longer-term effects on children's EF.

CRediT authorship contribution statement

Claire Essex: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft, Visualisation **Teodora Gliga:** Conceptualization, Supervision, Writing – Reviewing and Editing, **Maninda Singh:** Investigation, Writing – Review & Editing, **Tim Smith:** Conceptualization, Supervision, Writing – Reviewing and Editing.

Acknowledgements

This work was made possible by an Economic and Social Research Council (ESRC) Industrial Challenges Studentship to Claire Essex (Grant Number: ES/P000592/1). The authors would like to thank Ben Ewing for his help with data coding. The authors would also like to thank the reviewers involved in the peer-review of this paper for their support in preparing this manuscript for publication.

Declaration of interests

The authors report no declarations of interest.

References

Akshoomoff, N. (2002). Selective attention and active engagement in young children. *Developmental Neuropsychology, 22*(3), 625–642. https://doi.org/10.1207/S15326942DN2203_4

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology, 87*(2), 85–106. https://doi.org/10.1016/j.jecp.2003.10.002

American Academy of Pediatrics. (2016). Media and young minds. *Pediatrics, 138*(5), Article e20162591.

Anderson, D. R., Levin, S. R., & Lorch, E. P. (1977). The effects of TV program pacing on the behavior of preschool children. *Audio-Visual Communication Review, 25*(2), 159–166. https://doi.org/10.1007/BF02769779

Barr, R., Danziger, C., Hilliard, M. E., Andolina, C., & Ruskis, J. (2010). Amount, content and context of infant media exposure: a parental questionnaire and diary analysis. *International Journal of Early Years Education, 18*(2), 107–122. https://doi.org/10.1080/09669760.2010.494431

Barr, R., Lauricella, A., Zack, E., & Calvert, S. L. (2010). Infant and early childhood exposure to adult-directed and child-directed television programming: Relations with cognitive skills at age four. *Merrill-Palmer Quarterly, 56*(1), 21–48.

Bateman, J A, & Wildfeuer, J (2014). A multimodal discourse theory of visual narrative. *Journal of Pragmatics, 74*, 180–208. https://doi.org/10.1016/j.pragma.2014.10.001

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Bläsing, B. E. (2014). Segmentation of dance movement: Effects of expertise, visual familiarity, motor experience and music. *Frontiers in Psychology, 5*(OCT), 1–11. https://doi.org/10.3389/fpsyg.2014.01500

Bordwell, D., Staiger, J., & Thompson, K. (2003). Story causality and motivation. *The Classical Hollywood Cinema* (pp. 30–41). Routledge,.

Carmi, R., & Itti, L. (2005). Causal saliency effects during natural vision. Eye Tracking Research and Applications Symposium (ETRA), 2005, 11–18. https://doi.org/10.1145/1117309.1117313.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46*(26), 4333–4345. https://doi.org/10.1016/j.visres.2006.08.019

Cheng, S., Maeda, T., Yoichi, S., Yamagata, Z., & Tomiwa, K. (2010). Early television exposure and children's behavioral and social outcomes at age 30 months. *Journal of Epidemiology, 20*((SUPPL.2), 1–2. https://doi.org/10.2188/jea.JE20090179

Chonchaiya, Weerasak, & Pruksananonda, Chandhita (2008). Television viewing associated with delayed language development. *Acta Paediatrica, 97*(7), 977–982. https://doi.org/10.1111/j.1651-2227.2008.00831.x

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., & McCarty, C. A. (2004). Early Television Exposure and Subsequent Attentional Problems in Children. *PEDIATRICS, 113*(4), 708–713. https://doi.org/10.1542/peds.113.4.708

Cohn, N (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. Bloomsbury Academic.

Cooper, N. R., Uller, C., Pettifer, J., & Stolc, F. C. (2009). Conditioning attentional skills: Examining the effects of the pace of television editing on children's attention. *Acta Paediatrica, International Journal of Paediatrics, 98*(10), 1651–1655. https://doi.org/10.1111/j.1651-2227.2009.01377.x

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Dempster, F.N. (1985). Short-Term Memory Development in Childhood and Adolescence. https://doi.org/10.1007/978–1-4613–9541-6_5.

Diamond, A. (1995). Evidence of robust recognition memory early in life even when assessed by reaching behavior. *Journal of Experimental Child Psychology, 59*(3), 419–456. https://doi.org/10.1006/jecp.1995.1020

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Flores, S., Bailey, H. R., Eisenberg, M. L., & Zacks, J. M. (2017). Event segmentation improves event memory up to one month later. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(8), 1183–1202. https://doi.org/10.1037/xlm0000367

Foster, E. M., & Watkins, S. (2010). The Value of Reanalysis: TV Viewing and Attention Problems. *Child Development, 81*(1), 368–375. https://doi.org/10.1111/j.1467-8624.2009.01400.x

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive Function in Preschoolers: A Review Using an Integrative Framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(1), 3–27. https://doi.org/10.1017/S0021963097001753

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology, 40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Geist, E. A., & Gibson, M. (2000). The Effect of Network and Public Television Programs on Four and Five Year Olds Ability to Attend to Educational Tasks. *Journal of Instructional Psychology, 27*(4), 250.

Gernsbacher, M A (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *Psychology of learning and motivation, 27*, 217–263. https://doi.org/10.1016/S0079-7421(08)60125-5

Hale, S., Bronik, M. D., & Fry, A. F. (1997). Verbal and spatial working memory in school-age children: developmental differences in susceptibility to interference. *Developmental Psychology, 33*(2), 364–371. https://doi.org/10.1037/0012-1649.33.2.364

Henderson, J M, Chanceaux, M, & Smith, T J (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of vision, 9*(1), 32. https://doi.org/10.1167/9.1.32

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243–271. https://doi.org/10.1146/annurev.psych.50.1.243

Hendry, A., Jones, E. J. H., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review, 42*, 1–33. https://doi.org/10.1016/j.dr.2016.06.005

Huston-Stein, A., Fox, S., Greer, D., Watkins, B., A., & Whitaker, J. (1981). The effects of TV action and violence on children's social behavior. *The Journal of Genetic Psychology, 2*(138), 183–191.

Ildirar, Kirbas, Sermin, & Smith, Tim J. (2018). How infants perceive animated films. In M. Uhrig (Ed.), *Emotion in Animated Films. Routledge Advances in Film Studies*. Abingdon, UK: Routledge. ISBN 9781138303287.

Itti, L (2005). Models of bottom-up attention and saliency. *Neurobiology of attention* (pp. 576–582). Academic Press.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203. https://doi.org/10.1038/35058500

Kirkorian, H. L., Wartella, E. A., & Anderson, D. R. (2008). Media and Young Children ' s Learning. *The Furture of Children, 18*(1), 39–61.

Kopatich, R. D., Feller, D. P., Kurby, C. A., & Magliano, J. P. (2019). The role of character goals and changes in body position in the processing of events in visual narratives. *Cognitive Research: Principles and Implications, 4*(1), 1–15. https://doi.org/10.1186/s41235-019-0176-1

Kostyrka-Allchorne, K., Cooper, N. R., Gossmann, A. M., Barber, K. J., & Simpson, A. (2017). Differential effects of film on preschool children's behaviour dependent on editing pace. *Acta Paediatrica, International Journal of Paediatrics, 106*(5), 831–836. https://doi.org/10.1111/apa.13770

Kostyrka-allchorne, K., Cooper, N. R., Kennett, S., Simpson, A., Cooper, N. R., & Kennett, S. (2019). The Short-Term Effect of Video Editing Pace on Children ' s Inhibition and N2 and P3 ERP Components during Visual Go/No-Go Task The Short-Term Effect of Video Editing Pace on Children ' s. *Developmental Neuropsychology, 44*(4), 385–396. https://doi.org/10.1080/87565641.2019.1630628

Kostyrka-Allchorne, K., Cooper, N. R., & Simpson, A. (2017). The relationship between television exposure and children's cognition and behaviour: A systematic review. *Developmental Review, 44*, 19–58. https://doi.org/10.1016/j.dr.2016.12.002

Kostyrka-Allchorne, K., Cooper, N. R., & Simpson, A. (2019). Disentangling the effects of video pace and story realism on children's attention and response inhibition (Retrieved from) *Cognitive Development, 49*(November 2018), 94–104. https://doi.org/10.1016/j.cogdev.2018.12.003.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory and Cognition, 40*(5), 812–826. https://doi.org/10.3758/s13421-011-0179-8

Landhuis, C. E., Poulton, R., Welch, D., & Hancox, R. J. (2007). Does childhood television viewing lead to attention problems in adolescence? Results from a prospective longitudinal study. *Pediatrics, 120*(3), 532–537. https://doi.org/10.1542/peds.2007-0978

Lang, A (2000). The limited capacity model of mediated message processing. *Journal of communication, 50*(1), 46–70. https://doi.org/10.1111/j.1460-2466.2000.tb02833.x. https://doi-org.ezproxy.lib.bbk.ac.uk/10.1111/j.1460-2466.2000.tb02833.x

Lang, A., Bolls, P., Potter, R. F., & Kawahara, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. *Journal of Broadcasting and Electronic Media, 43*(4), 451–475. https://doi.org/10.1080/08838159909364504

Levine, D., Hirsh-Pasek, K., Pace, A., & Golinkoff, R. M. (2017). A goal bias in action: The boundaries adults perceive in events align with sites of actor intent. *Journal of Experimental Psychology: Learning Memory and Cognition, 43*(6), 916–927. https://doi.org/10.1037/xlm0000364

Levine, L.E., & Waite, B.M. (2000). Television Viewing and Attentional Abilities in Fourth and Fifth Grade Children. 21(6), 667–679.

Lillard, A.S., Li, H., & Boguszewski, K. (2015a). Television and Children's Executive Function. 48, 219–248.

Lillard, A. S., Drell, M. B., Richey, E. M., Boguszewski, K., & Smith, E. D. (2015). Further examination of the immediate impact of television on children's executive function. *Developmental Psychology, 51*(6), 792–805. https://doi.org/10.1037/a0039097

Lillard, A. S., & Peterson, J. (2011). The immediate impact of different types of television on young children's executive function. *Pediatrics, 128*(4), 644–649. https://doi.org/10.1542/peds.2010-1919

Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2020). The Scene Perception & Event Comprehension Theory (SPECT) Applied to Visual Narratives. *Topics in Cognitive Science, 12*(1), 311–351. https://doi.org/10.1111/tops.12455

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing Space and Time in Film Understanding. *Applied Cognitive Psychology, 15*(5), 533–545. https://doi.org/10.1002/acp.724

Magliano, J. P., & Zacks, J. M. (2011). The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science, 35*(8), 1489–1517. https://doi.org/10.1111/j.1551-6709.2011.01202.x

Magliano, J P, Zwaan, R A, & Graesser, A C (1999). The role of situational continuity in narrative understanding. *The construction of mental representations during reading* (pp. 219–245). Psychology Press.

McCollum, J. F., Jr., & Bryant, J. (2003). Pacing in Children's Television Programming. *Mass Communication and Society, 6*(2), 115–136. https://doi.org/10.1207/S15327825MCS0602_1

Miller, C.J., Marks, D.J., Miller, S.R., Berwid, O.G., Kera, E.C., Santra, A., & Halperin, J.M. (2007). *Brief Report: Television Viewing and Risk for Attention Problems in Preschool Children*. 32(4), 448–452.

Mital, P K, Smith, T J, Hill, R L, & Henderson, J M (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive computation, 3*(1), 5–24.

Miyake, Akira, Friedman, Naomi P, Emerson, Michael J, Witzki, Alexaner H, & Howerter, Amy (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe'' Tasks: A Latent Variable Analysis. *Cognitive Psychology, 41*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Obel, C., Henriksen, T. B., Dalsgaard, S., Linnet, K. M., Skajaa, E., Thomsen, P. H., & Olsen, J. (2004). Does children's watching of television cause attention problems? Retesting the hypothesis in a Danish cohort. author reply 1373-4 *Pediatrics, 114*(5), 1372–1373. https://doi.org/10.1542/peds.2004-0954.

Ohman, A (1979). The orienting response, attention, and learning: An information-processing perspective. In H D Kimmel, E H van Olst, & J F Orlbeke (Eds.), *The orienting reflex in humans*. Hillsdale, NJ: Hillsdale.

Özmert, E., Toyran, M., & Yurdakök, K. (2002). Behavioral correlates of television viewing in primary school children evaluated by the child behavior checklist. *Archives of Pediatrics and Adolescent Medicine, 156*(9), 910–914. https://doi.org/10.1001/archpedi.156.9.910

Pannasch, S., Helmert, J. R., Roth, K., Herbold, A.-K., & Walter, H. (2008). Visual Fixation Durations and Saccade Amplitudes: Shifting Relationship in a Variety of Conditions. *Journal of Eye Movement Research, 2*(2), 1–19. https://doi.org/10.16910/jemr.2.2.4

Perez, J., & Feigenson, L. (2020). *Violations of expectation trigger infants to search for explanations*. 1–33. https://doi.org/10.31234/osf.io/eahjd.

Pempek, T A, Kirkorian, H L, Richards, J E, Anderson, D R, Lund, A F, & Stevens, M (2010). Video comprehensibility and attention in very young children. *Developmental psychology, 46*(5), 1283–1293. https://doi.org/10.1037/a0020614

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 608–620. https://doi.org/10.1002/wcs.133

Ribner, A. D., Barr, R. F., & Nichols, D. L. (2021). Background media use is negatively related to language and literacy skills: indirect effects of self-regulation. *Pediatric Research, 89*(6), 1523–1529. https://doi.org/10.1038/s41390-020-1004-5

Rothbart, M., & Posner, M. (2001). Mechanism and variation in the development of attentional networks In Nelson C., Luciana M. (Eds.), Handbook of developmental cognitive neuroscience (pp. 353–363).

Rideout, Victoria, & Common Sense Media. (2017). *The Common Sense census: Media use by kids age zero to eight*. San Francisco: Common Sense Media.

Ruff, H. A., Capozzoli, M., & Weissberg, R. (1998). Age, individuality, and context as factors in sustained visual attention during the preschool years. *Developmental Psychology, 34*(3), 454–464. https://doi.org/10.1037/0012-1649.34.3.454

Ruff, H. A., & Rothbart, M. K. (2001). *Attention in early development: Themes and variations*. Oxford University Press,.

Salt, B (1992). *Film style and technology: History and analysis*. Starword.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition, 129*, 241–255.

Setliff, A. E., & Courage, M. L. (2011). Background television and infants' allocation of their attention during toy play. *Infancy, 16*(6), 611–639. https://doi.org/10.1111/j.1532-7078.2011.00070.x

Sisson, S. B., Broyles, S. T., Baker, B. L., & Katzmarzyk, P. T. (2010). Screen time, physical activity, and overweight in U.S. Youth: National survey of children's health 2003. *Journal of Adolescent Health, 47*(3), 309–311. https://doi.org/10.1016/j.jadohealth.2010.02.016

Smith, T J (2012). The attentional theory of cinematic continuity. *Projections, 6*(1), 1–27. https://doi.org/10.3167/PROJ.2012.060102

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

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes, 16–16 *Journal of Vision, 13* (8). https://doi.org/10.1167/13.8.16.

Stahl, A. E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition, 163*, 1–14. https://doi.org/10.1016/j.cognition.2017.02.008

Stevens, T., & Mulsow, M. (2006). There is no meaningful relationship between television exposure and symptoms of attention-deficit/hyperactivity disorder. *Pediatrics, 117*(3), 665–672. https://doi.org/10.1542/peds.2005-0863

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138*(2), 236–257. https://doi.org/10.1037/a0015631

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

Watt, J H, & Welch, A J (1983). Effects of static and dynamic complexity on children's attention and recall of televised instruction. *Children's understanding of television: Research on attention and comprehension* (pp. 69–102). New York: Academic Press Inc.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in Reading and Film Comprehension. *Journal of Experimental Psychology: General* (Vol. 138,, 307–327. https://doi.org/10.1037/a0015305

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the alzheimer type. *Psychology and Aging, 21*(3), 466–482. https://doi.org/10.1037/0882-7974.21.3.466

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research, 1076*(1), 150–162. https://doi.org/10.1016/j.brainres.2005.12.122

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of Situation Model Construction in Narrative Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 386–397. https://doi.org/10.1037/0278-7393.21.2.386

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language and memory. *Psychological Bulletin, 123*(2), 162–185 (Retrieved from) 〈https://search-proquest-com.proxy.library.carleton.ca/docview/614398538/fulltextPDF/F8C6586362BD435FPQ/1?accountid=9894〉.

https://apps.who.int/iris/bitstream/handle/10665/325147/WHO-NMH-PND-2019.4-eng.pdf. 2019. (Accessed 01 March 2021).

https://www.statista.com/statistics/397833/hours-of-media-consumption-by-children-by-media-uk/. 2020. (Accessed 01 March 2021).