Infant Behavior and Development 49 (2017) 248–266

Contents lists available at ScienceDirect

Infant Behavior and Development

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

# Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers

Andrea Helo[a,b,*], Sandrien van Ommen[a], Sebastian Pannasch[c], Lucile Danteny-Dordoigne[a], Pia Rämä[a,d]

[a] *Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France*
[b] *Departamento de Fonoaudiología, Universidad de Chile, Santiago, Chile*
[c] *Department of Psychology, Engineering Psychology and Applied Cognitive Research, Technische Universitaet Dresden, Germany*
[d] *CNRS (UMR 8242), Paris, France*

ARTICLE INFO

*Keywords:*
Scene viewing
semantic knowledge
vocabulary skills
saliency
eye movement development

ABSTRACT

Conceptual representations of everyday scenes are built in interaction with visual environment and these representations guide our visual attention. Perceptual features and object-scene semantic consistency have been found to attract our attention during scene exploration. The present study examined how visual attention in 24-month-old toddlers is attracted by semantic violations and how perceptual features (i. e. saliency, centre distance, clutter and object size) and linguistic properties (i. e. object label frequency and label length) affect gaze distribution. We compared eye movements of 24-month-old toddlers and adults while exploring everyday scenes which either contained an inconsistent (e.g., soap on a breakfast table) or consistent (e.g., soap in a bathroom) object. Perceptual features such as saliency, centre distance and clutter of the scene affected looking times in the toddler group during the whole viewing time whereas looking times in adults were affected only by centre distance during the early viewing time. Adults looked longer to inconsistent than consistent objects either if the objects had a high or a low saliency. In contrast, toddlers presented semantic consistency effect only when objects were highly salient. Additionally, toddlers with lower vocabulary skills looked longer to inconsistent objects while toddlers with higher vocabulary skills look equally long to both consistent and inconsistent objects. Our results indicate that 24-month-old children use scene context to guide visual attention when exploring the visual environment. However, perceptual features have a stronger influence in eye movement guidance in toddlers than in adults. Our results also indicate that language skills influence cognitive but not perceptual guidance of eye movements during scene perception in toddlers.

## 1. Introduction

Our everyday visual environment is predictable even if particular aspects often vary across situations. For instance, objects are most likely to appear in certain contexts, conforming visual scenes (e.g., kitchen). However, the position of objects can vary depending on the layout of a particular scene (e.g., two different kitchens). The concepts of natural or real-world scenes are often used to refer to representations of the real visual world that are constrained by its semantic and spatial configurations (Henderson & Ferreira, 2004; Henderson & Hollingworth, 1999). Evidence from studies using natural scenes indicates that with visual experience viewers store information about different scene types in the long-term memory and establish a "scene knowledge" (Barlett, 1932; Hock, Romanski, Galie, & Williams, 1978; Mandler & Johnson, 1976; Potter, 1975). Scene knowledge allows quick extraction of the

* Corresponding author at: Laboratoire Psychologie de la Perception (CNRS UMR 8242), Université Paris Descartes, 45, rue des Saints-Pères, 75006 Paris, France.
*E-mail address:* ahelo@med.uchile.cl (A. Helo).

http://dx.doi.org/10.1016/j.infbeh.2017.09.008
Received 12 January 2017; Received in revised form 14 September 2017; Accepted 16 September 2017

global meaning of a scene, i.e. the so-called gist. After the gist extraction, viewers generate expectations about possible objects and their locations within a scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Hock et al., 1978; Mandler & Johnson, 1976; Oliva, 2005; Potter, 1976). These expectations guide further visual exploration.

Visual attention allocation—as reflected by fixation locations (Yarbus, 1967)—is affected by low-level properties of the images during scene exploration (Henderson, 2003; Itti & Koch, 2000;Le Meur, Le Callet, & Barba, 2007; Parkhurst, Law, & Niebur, 2002). In particular, saliency has been proven to be a determining factor in gaze allocation (Koch, 2000, 2001; Koch, 2000, 2001;Koch & Ullman, 1985; Treue, 2003; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006). Studies using saliency as a predictor of gaze distribution within a scene have shown that salient regions are fixated more than control locations or locations expected by chance (Foulsham & Underwood, 2008;Itti & Koch, 2001; Parkhurst et al., 2002). Recently, also other image features have shown to influence eye movement behavior during scene exploration. The global clutter of an image has shown to influence first saccade latencies and fixation durations during a searching task (Henderson & Smith, 2009). Further, high edge density and clutter of scenes patches attract the gaze during a memory task (Nuthmann & Einhäuser, 2015). Additionally, object size (Clarke, Coco, & Keller, 2013) and central location (Nuthmann & Einhäuser, 2015; Tatler, 2007) have shown to influence gaze allocation. Adults tend to make more fixations at the center than the periphery of a scene. Likewise, larger objects attract the fixation more than smaller objects.

Regarding cognitive mechanisms, semantic knowledge of the scene and behavioural task demands have shown to influence gaze allocation during scene exploration (Castelhano, Mack, & Henderson, 2009; Fischer, Graupner, Velichkovsky, & Pannasch, 2013; Mills, Hollingworth, & Dodd, 2011; Tatler & Vincent, 2008). Several lines of evidence indicate that scene-object consistency influence visual attention during scene exploration (Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Võ & Henderson, 2009). For instance, objects semantically inconsistent with the scene context—and thereby violating the expectations of the viewer—attract the gaze of observers (semantic consistency effect) increasing number of fixation landings (Henderson et al., 1999; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Võ & Henderson, 2009). It has been proposed that semantic consistency effects reflect a high requirement of attentional resources either for the identification of the object in the scene or for solving the conflict given by the semantic violation (Davenport, 2007; Loftus & Mackworth, 1978). Currently, there is no consensus on whether semantic inconsistencies guide the eye movements before or only after the object is fixated. While some studies revealed that semantic inconsistencies are detected within the first 200 milliseconds, i.e. influencing eye movements before the inconsistent object has been fixated (Becker, Pashler, & Lubin, 2007; Loftus & Mackworth, 1978; Underwood, Humphreys, & Cross, 2007; Underwood, Templeman, Lamming & Foulsham, 2008), others have shown that the inconsistent objects need to be fixated in order to be detected (De Graef et al., 1990;Gareze & Findlay, 2007; Henderson et al., 1999; Võ & Henderson, 2009; Võ & Henderson, 2011). In particular, studies where participants were presented with complex real-world scenes (Henderson et al., 1999) and with real-world scenes where low-level features were controlled (Võ & Henderson, 2009) failed to find an extrafoveal semantic consistency effect either during scene memorization or visual search. Based on these results, it has been suggested that the findings of extrafoveal effect of semantic inconsistency might be related to visual features of stimuli such as image density as well as object conspicuity and eccentricity rather than semantic inconsistency detection (Henderson & Hollingworth, 1999; Henderson et al., 1999; Võ & Henderson, 2009)

There is also extensive evidence demonstrating that visual attention allocation is influenced by the interaction between the bottom-up (i.e. perceptual features) and top-down (i.e. cognitive control) mechanisms (Koch, 2000, 2001; Koch, 2000, 2001; Parkhurst et al., 2002; Torralba, Oliva, Castelhano, & Henderson, 2006). Previous studies have shown that the influence of saliency on fixation distribution is more significant during the early than late stages of viewing time (Mannan, Ruddock, & Wooding, 1995; Parkhurst et al., 2002). Based on these studies, it has been proposed that early in viewing exploration, visual attention is mainly guided by salient areas within a scene whereas during the later stages top-down control dominates visual attention guidance (Castelhano et al., 2009; Fischer et al., 2013; Mills et al., 2011; Tatler & Vincent, 2008). In addition, studies using saliency maps have shown that the semantically informative stimuli decrease the influence of saliency on gaze allocation (Castelhano & Henderson, 2007;Nyström & Holmqvist, 2008; Parkhurst et al., 2002). Based on these findings, it has been also proposed that top-down control modulates the strength of bottom-up saliency contribution to attention guidance (Einhäuser, Rutishauser, & Koch, 2008; Parkhurst et al., 2002; Theeuwes, 2010;Treue, 2003).

Recently, it has been shown that language processing can also guide visual attention during natural scene exploration (Andersson, Ferreira, & Henderson, 2011; Clarke et al., 2013; Coco, Malcolm, & Keller, 2014). In previous work, the complexity of linguistic stimuli, differentiated by the speed of spoken sentences (high vs. low) was manipulated (Andersson et al., 2011). An object was more likely fixated when it was mentioned in a sentence, but also the linguistic complexity influenced the fixation probability. Objects that were mentioned in low complexity condition were fixated more likely and earlier compared to those in the high complexity condition, suggesting that linguistic processing influences the gaze distribution within a scene. Furthermore, it has been found that during scene exploration, naming and gaze allocation influenced each other when participants had to name seen objects (Clarke et al., 2013; Coco et al., 2014). Particularly, fixation landings and perceptual properties of the objects enhanced their probability of being named. At the same time, the fixation distribution was affected by linguistic properties of objects labels such as semantic proximity (i.e. a similarity between words based on their co-occurrence in a similar context) or word frequency (Clarke et al., 2013; Coco et al., 2014).

The development of eye movement control (e.g. Açik, Sarwary, Schultze-Kraft, Onat, & König, 2010; Helo, Pannasch, Sirri, & Rämä, 2014;Helo, Rämä, Pannasch, & Meary, 2016; Karatekin, 2007; Luna, Velanova, & Geier, 2008) and of cognitive resources (Gathercole, 1999;Gathercole, Pickering, Ambridge, & Wearing, 2004; Hitch, Halliday, Schaafstal, & Schraagen, 1988;Klenberg, Korkman, & Lahti-Nuuttila, 2001; Pearson & Lane, 1991; Pickering, 2001; Sanders, Stevens, Coch, & Neville, 2006) during infancy and childhood makes it plausible to assume that the interaction between perceptual (e.g., saliency) and cognitive (e.g., semantic) gaze guidance is different in young children and adults. For example, it has been shown that saliency guides eye movements to a larger extent in younger children compared to older children and adults (Açik et al., 2010; Helo et al., 2014). More specifically, fixations of children from 2 to 6-years of age were shown to be more attracted to salient areas of images than of eight to ten-year-olds and adults (Helo et al., 2014). These findings indicate that perceptual gaze guidance is more pronounced in young children. Nevertheless, it is also known that semantic scene knowledge is built through visual experience (Barlett,