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

* 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).

https://doi.org/10.1016/j.infbeh.2017.09.008
Received 12 January 2017; Received in revised form 14 September 2017; Accepted 16 September 2017
Available online 10 October 2017

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 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, 1932; Hock et al., 1978; Mandler & Johnson, 1976), and thus, it is likely that the top-down control of visual attention increases with age. So far, only little is known about the semantic scene processing development (Bornstein, Arterberry, & Mash, 2010; Bornstein, Mash, & Arterberry, 2011a; Bornstein, Mash, & Arterberry, 2011b; Duh & Wang, 2014; Hock et al., 1978). To our knowledge, only one study investigated the interaction of perceptual and semantic factors during natural scene viewing in young children (Duh & Wang, 2014). In this study, 15-month-olds were habituated with visual scenes by repeating the images until their attention decreased (i.e. they stopped looking at the screen). After the habituation phase, their looking times to the screen were measured in response to an object change. Children detected a perceptual (e.g., salient object) but not a semantic (e.g., object that disrupted the scene context) replacement when the scenes were presented shortly for 500 ms. However, presenting the images for 3000 ms—thereby allowing access to the scene meaning—they looked longer at the screen following a change that disrupted the meaning of the scene (e.g., replacing a beach umbrella by a table) compared to a perceptually salient change that preserved the gist (e.g., replacing a beach umbrella by a colourful beach umbrella). These results suggest that already 15-month-old infants take into account both low-level image features and semantic properties when processing visual scenes. Low-level features were processed faster than semantic properties, **which were processed** only when children were given more time to extract the gist of the scene (Duh & Wang, 2014). However, further empirical evidence using free exploration tasks is needed to better understand interactions between perceptual features and semantic properties in eye movement guidance in young children.

Recent evidence furthermore indicates that young children use implicit naming already during early language acquisition (Mani and Plunkett, 2010, 2011). The authors presented children at 18 and 24 months of age images (e.g., cup) followed by images of a phonologically similar target (e.g., cat) and an unrelated distractor (e.g., shoe). Looking times to the named target were affected by the prime image even it has not been named before. These findings indicate that children implicitly activated the label of the prime image. The spontaneous naming of objects has been also shown in 2-year-old children when playing freely with a set of novel objects (Samuelson & Smith, 2005). In this study, the tendency of children to name novel objects increased steadily as a function of their productive vocabulary size. Since object naming has shown to influence attention allocation during natural scene viewing in adults (Clarke et al., 2013; Coco et al., 2014) and young children are prone to name (Mani and Plunkett, 2010, 2011; Samuelson & Smith, 2005) it is plausible that children also use internal language for guiding their visual attention during scene viewing.

The main aim of the present study was to examine the influence and interaction of semantic consistency and saliency on attention allocation in 24-month-old children during a free exploration of visual scenes. We also investigated how other perceptual features such as centre distance, the clutter of the scene and object size influence gaze allocation in toddlers. Additionally, we asked whether implicit naming and language skills affected visual attention allocation during scene exploration in toddlers. To this end, we analysed the effect of linguistic properties of object labels and expressive vocabulary size on visual attention allocation. A group of 24-month-old children and a group of adults inspected typical indoor scenes, such as kitchens and bedrooms, while their eye movements were tracked. Half of the scenes displayed a target object that was inconsistent with the scene context (e.g., soap on a breakfast table) while the remaining half contained a target object that was consistent with the scene context. (e.g., soap in a bathroom). We measured both extrafoveal (latency of the first saccade to the object) and foveal (dwell time, first-pass gaze duration) parameters to examine possible influences of perceptual features, linguistic properties and semantic inconsistencies on eye movement guidance before or after the object was fixated. According to earlier findings on the development of visual attention (Helo et al., 2014, 2016) and semantic scene processing (Duh & Wang, 2014) in early infancy and childhood, we expected to find influences of scene context on eye guidance already in 24-month-olds, which should be expressed by amplified attention allocation to semantic inconsistencies. However, since semantic processing is still in development during early childhood, and in line with earlier findings on the influence of perceptual features on scene viewing in younger than in older children (Açik et al., 2010; Helo et al., 2014), we expected a pronounced influence of bottom-up factors on gaze allocation in toddlers when compared with adults. Likewise, based on previous evidence showing that at 2 years of age children are prone to name (Mani and Plunkett, 2010, 2011; Samuelson & Smith, 2005), we expected that linguistic properties of object labels had a stronger effect on toddlers than adults. To investigate the contribution of linguistic skills to the allocation of attention, we compared the gaze distribution in children with different expressive vocabulary skills. We predicted a direct influence of vocabulary skills on the allocation of visual attention. In particular, we expected that children with higher vocabulary skills were more likely to implicitly name objects during visual exploration affecting the distribution of their gaze within the scene.

2. Materials and Methods

2.1. Subjects

A total of 84 subjects participated in the experiment including 30 neurologically healthy adults (18 females, mean age 28 years, range 20-34 years) and 52 children. The data of 10 children were rejected due to “fussiness” during the experiment, calibration problems, or an insufficient number of trials. Therefore, data from 42 toddlers (21 girls, mean age 24 months, range 23-25 months) were analyzed in this study. Children were recruited from a database of parents who agreed to volunteer in child development studies and came from diverse socioeconomic backgrounds in the Parisian region. All children were born full-term and presented a typical development. Half of the children viewed image set 1 and half of them set 2. Ages and number of boys and girls were counterbalanced

across both sets (9 girls set 1, 12 girls set 2; mean age 24 months, range 23-25 in both sets). All adult participants had normal or corrected to normal vision and no history of psychiatric or neurological diseases. Participants (and their parents) were informed of the purpose of the study before signing the consent. The study was conducted in conformity with the declaration of Helsinki and approved by the Ethics Committee of the University of Paris Descartes.

2.2. Apparatus

Eye movements were sampled monocularly at 500 Hz using the desktop version of the EyeLink 1000 eye tracker system (SR Research, Ontario, Canada). In order to operate the system in the remote mode, a small target sticker was placed on the participants' forehead. The sticker allowed tracking of head position even when the pupil image was lost (i.e., during blinks or sudden movements). Fixations and saccades were defined using the saccade detection algorithm supplied by SR Research: Saccades were identified by deflections in eye position in excess of 0.1°, with a minimum velocity of $30°s^{-1}$ and a minimum acceleration of $8000°s^{-2}$, maintained for at least 4 ms. The first fixation in each trial was defined as the first fixation that began after the onset of the image. Pictures were displayed using a GeForce 7300 GT card and a CRT display (Sony GDM F520) at 1024 × 728 pixels at a refresh rate of 100 Hz viewed from a distance of 60 cm.

2.3. Stimuli and Design

Thirty-six color photographs of eighteen different indoor scenes served as stimuli. The images were photographed using a Nikon D5100 camera and had a resolution of 1024 × 768 pixels and 24-bit color depth. The images represented typical examples of Parisian indoor sceneries of four different home interior scenes (kitchens, bathrooms, bedrooms or living rooms), taken from five different homes. Each scene was shown either as semantically consistent or inconsistent. Semantically inconsistent scenes included a target object that did not conform to the scene context (e.g., soap on a breakfast table).

Semantically consistent scenes included a control object that was consistent with the scene context (e.g., a piece of bread on a breakfast table). Each object was shown twice, both in a consistent and in an inconsistent scene context (Fig. 1). In a particular scene, the objects in the consistent and inconsistent scenes were of similar size. Objects' locations were proportionately distributed across the four quadrants of a scene. An area of interest (AOI) was defined for each inconsistent object (Inconsistent AOI) and its control object (Consistent AOI). The mean area of AOIs was 64 cm^2. The AOI sizes for consistent and inconsistent objects in a given scene were identical (see, Fig. 1) but varied across the scenes within a range of 28–128 cm^2. Each participant viewed a set of 18 scenes. Each set contained a scene either in its consistent or inconsistent version to avoid repetition of a particular scene. To control that saliency levels between scene types (consistent and inconsistent AOIs) and sets were not different, objects within the scene were ranked using the MATLAB Saliency Toolbox (Walther & Koch, 2006). Based on low-level image features such as intensity, color and orientation, this toolbox creates a saliency map that allows estimating the saliency level of each region in an image. Repeated measures analyses of variance showed that the saliency of consistent AOIs (Set 1: $M = 2.7$, Set 2, $M = 4.2$) was not different from those of inconsistent AOIs (inconsistent condition, Set1: $M = 3.7$, Set 2: $M = 4.0$), $F < 1$. Likewise, no difference between sets $F(1,16) = 1.08$, $p = .31$ or interaction $F(1,16) = 1.26$, $p = .28$ between scene type and set were found.

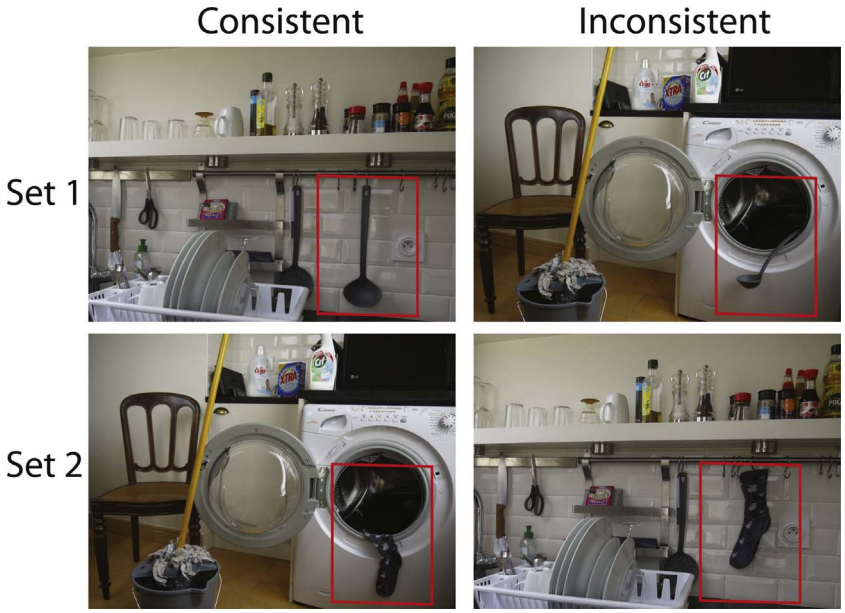

Fig. 1. Examples of consistent and inconsistent scenes in set 1 and set 2. Red rectangles (not shown during the experiment) illustrate the areas of interests (AOIs).

2.4. Predictors

2.4.1. Semantic consistency

The semantic consistency was tested in a pilot study with 16 adult participants who inspected all 36 images (i.e. both sets) but did not participate in the main experiment. Therefore, each participant had to rate the consistency of each scene using the scale from 1 to 9 (1 = highly consistent and 9 = highly inconsistent). The ratings differed significantly between consistent and inconsistent scenes, $F(1,16) = 352.07$, $p < .001$, suggesting that there was no ambiguity between scene types (set 1: M = 1.1 for consistent images and M = 5.96 for inconsistent images; set 2: M = 1.2 for consistent images and M = 5.95 for inconsistent images). Ratings did not differ between the images sets, $F < 1$.

2.4.2. Visual features

To test the effect of saliency, the saliency rank (Walther & Koch, 2006) was used to estimate the saliency level of each AOI within a scene. Each scene was ranked from 1 to 8 depending on its saliency in the scene (1 being the most salient). The effect of object size was also analysed, the area of the consistent and inconsistent objects was estimated by measuring their number of pixels. Visual attention is biased towards the centre of an image (Tatler, 2007), and thus, the centre distance that is the distance of the centroid of an object to the centre of the image was also measured. Finally, feature congestion scores were calculated for each scene as a measure of the clutter of the scene. A feature congestion map of visual clutter was computed for each scene using the algorithms described by Rosenholtz and co-workers (Rosenholtz, Li, & Nakano, 2007) and the MATLAB code provided at http://dspace.mit.edu/handle/1721.1/37593. Feature congestion scores for the clutter of a scene were defined as the mean over this feature map's values for the whole scene.

2.4.3. Linguistic properties

In order to determine whether linguistic features have an influence on eye movement behaviour, lexical frequencies and word lengths of the object labels were estimated. The lexical frequency of each label was obtained using the ChildFreq tool that extracts word frequencies from the Childes database (Bååth, 2010). Word frequency was reported by using the number of word tokens per million words. As a measure of word length, we used the number of syllables.

2.5. Procedure

All participants completed the experiment in a sound attenuated, dimly lit room. Children participants were seated on the laps of their parents. Each participant inspected a set of 18 scenes. Half of the participants inspected the images of set 1, and the other half the images of set 2. In order to keep adult participants motivated during the task, they were informed that a memory test would be administered at the end of the session.

Experimental sessions began with a 5-point calibration and validation of the eye-tracking system. Individual trials started with a drift check accompanied by the presentation of a full-screen scene image. Each image was shown for 7 seconds. The experiment consisted of three blocks with six trials in each, presented in a randomized order. Between the blocks, children were allowed to have a small break or to watch a short animation film of 2 minutes length. The total duration of the experiment was approximately 15 minutes. At the end of the experiment, adult participants were presented with images of eight scenes that they saw during the experiment but the target objects were hidden with a grey rectangle. They were asked to recall which object was shown previously in that location. The main purpose of the task was to ensure adult subjects' motivation during the experiment.

The French translation and adaptation of the MacArthur Communicative Development Inventory for Words (CDI) was used to measure comprehensive and productive vocabulary sizes of children (Fenson et al., 1993). The parents were asked to complete the CDI within two weeks following the experiment. Parents for 34 participants (out of 42) completed CDIs for their children. In addition, parents were asked to indicate, by filling an object-knowledge questionnaire, whether their child understood and/or produced the labels of target objects used in the experiment. All parents returned the completed second questionnaire.

2.6. Data Analysis

The eye-tracking data was used to generate the following eye movement parameters for our analyses. As a first step, we analysed extrafoveal and foveal measures to examine how semantic consistency, perceptual attributes, and linguistic features interact to modulate attention allocation in adults and toddlers. The latency of the first saccade served as extrafoveal measures by indicating to what extent eye movements—before the first fixation on the target—are modulated by these parameters. Foveal measures reveal the degree of attention allocated to the target. In particular, dwell time (i.e., looking time in the AOI divided by the total looking time of the trial) and first-pass gaze duration (i.e., looking time in the AOI immediately after the first fixation in this region) were used as a foveal measure to study the interaction of these parameters on visual attention allocation.

In order to determine whether language production correlated with eye measures previously associated with cognitive load in the toddler group the correlation of expressive vocabulary size with different parameters was analysed using Spearman correlation. These parameters were individual fixation duration, associated with visual processing effort, initial fixation durations associated to scene context retrieval, and total trial dwell time (a total looking time to the screen) associated to task engagement.

The statistical analyses were performed with the statistical framework of a linear mixed modeling (LME) as implemented by the R package lme4 (Baayen, Davidson, & Bates, 2008). Following Barr et al. (2013), we pursued maximal models for each of our dependent variables (Barr, Levy, Scheepers, & Tily, 2013) and we used the R package lmerTest (Kuznetsova et al., 2016) to estimate the

degrees of freedom and p-values for each model parameter. We analysed both extrafoveal (latency of the first saccade to AOIs) and foveal parameters (dwell time and first-pass gaze duration) and each model included age group, semantic consistency, saliency, centre distance and trial as predictors of eye movement behaviour. We also included other perceptual (object size, the clutter of the scene) and linguistic (word length and word frequency) features as predictors to assess their contribution to gaze distribution on toddlers and adults. To test our main hypothesis we included the interaction of age group with saliency and semantic consistency. We also added the interaction of age group, semantic consistency, and centre distance because previous studies have shown that central bias is a strong predictor of fixation landing in a scene (Nuthmann & Einhäuser, 2015; Tatler, 2007). Predictors were centred to have a mean value of 0 and scaled to a standard deviation of 1. Our random effects structure included intercepts for subjects, scenes, and objects, as well as slopes for semantic consistency and saliency for each of the random factors. Model assumptions of normally distributed residuals were met by log-transformation of the dependent variables and subsequent visual inspection of diagnostic plots of residuals for influential outliers, normality, and heteroscedasticity. Significant fixed effects are reported in the running text, full model results can be found in the appendix.

To study the effect of expressive vocabulary skills on the allocation of attention during scene processing when semantic consistency is manipulated, we ran the language model for the toddler group alone on the foveal parameters, including the same variables as the main model but replacing age group with expressive vocabulary size. This model additionally contained an interaction of expressive vocabulary size with linguistic features of the object label, to study its effect on eye movement behaviour. Because of expressive vocabulary size has a strong relation to by-subject variation, by-subject slopes were removed from the random effects structure in this model, leaving the subject, object and scene intercepts, and the by-object and by-scene slopes.

3. Results

Only the trials where the participants inspected the screen at least during 60% of total viewing time and participants fixated the target objects at least once were included in the analysis. In addition, trials in which toddler participants did not know the label of the target or control object were excluded from the analyses. In the adult group, 96% of the trials (522 out of 540 trials) fulfill the inclusion criteria and were analysed. A total of 18 trials were excluded because the participants did not reach the AOI (M = 0.6, SD = 0.81). In the toddler group, 60% of the trials (454 out of 756 trials) reached the inclusion criteria and were included in the analyses. A total of 112 trials (M = 2.66, SD = 2.30) were excluded because the total trial looking time was less than 60%, and 125 additional trials (M = 2.98, SD = 1.58) were excluded because the participants did not reach the AOI. In addition, 65 trials (mean = 1.55, SD = 1.61) were excluded because the participants, according to the parental reports, did not know the label of the object in the AOI. The number of fixations before reaching the AOI was six both in the adult (SD = 2.06) and the toddler (SD = 2.93) group. Fixation durations were significantly longer in toddlers (M = 338 ms, SD = 54.11) than in adults (M = 239 ms, SD = 30.71), $t(70) = 8.83$ $p < .001$. In the toddler group, there were no correlations between expressive vocabulary size and eye movement measures (fixation duration, $r_s = .072$, initial fixation duration, $r_s = .075$, and total trial time dwell time, $r_s = -.00$, all ps > .05), indicating that eye movement measures that have been earlier associated with cognitive load in visual processing were not affected by vocabulary skills. Adult participants performed a memory test after the experiment: the results showed that they were equally likely to remember consistent (M = 0.82, SD = 0.19) and inconsistent (M = 0.77, SD = 0.22) objects.

3.1. Extrafoveal processing

To study the influence of semantic consistency, perceptual attributes and linguistic features on early eye movements (prior to the first fixation to the AOIs) we ran the main model with latency of the first saccade to the AOI as a dependent variable.

3.1.1. Main model for first saccade latency

The model for first saccade to the AOI revealed a main effect of centre distance, $b = 0.34$, $t(14) = 3.79$, $p = .002$, with shorter latencies to the AOIs closer to the centre and a main effect of age group, $b = 7.16$, $t(76) = -7.77$, $p < .001$, with shorter latencies for adults. The model also revealed an interaction of age group and centre distance, $b = 0.14$, $t(815) = 2.78$, $p = .006$, based on a stronger effect of centre distance in toddlers than in adults. In addition, there was an effect of trial, $b = 0.06$, $t(898) = -2.48$, $p = .013$, the latencies to AOIs became shorter with increasing trial number. There was also a marginal interaction of semantic consistency and saliency rank, $b = 0.17$, $t(10) = 2.24$, $p = .05$, based on shorter latencies to the AOI for higher saliency.

To further explore the interaction between age group and centre distance, we proceeded to split the model by age group, building two identical models for adults and children.

3.1.2. Adult model for latency of first saccade

The adult model for the latency of the first saccade to the AOI revealed a main effect of centre distance, $b = 0.48$, $t(26) = 4.62$, $p < .0001$, with shorter latencies to the AOIs closer to the centre and trial, $b = -0.08$, $t(467) = -2.59$, $p = .009$, with shorter latencies as the trial number increased. There was also an interaction of consistency and centre distance, $b = 1.88$, $t(467) = -2.59$, $p = .009$. A simple slope comparison moving centre distance by 2 SD in both directions showed a main effect of semantic consistency, $b = 0.45$, $t(19) = 2.73$, $p = .013$, only in central objects. Shorter latencies to inconsistent than to consistent objects were obtained for more central objects (Fig. 2). In addition, the simple slope showed a main effect of saliency, $b = -0.35$, $t(16) = -2.98$, $p = .009$, only in central objects, with shorter latencies to more salient objects. Trends suggested main effects of feature congestion, $b = 0.22$, $t(15) = 1.95$, $p = .069$, with shorter latencies for objects in less congested scenes, as well as for label-word frequency ($b = 0.09$, $t(19) = -1.89$, $p = 0.074$); with shorter latencies for more frequent words.

3.1.3. Children model for latency of the first saccade

The children model for the first saccade to the AOI only showed an effect of centre distance, $b = 0.22$, $t(17) = 2.26$, $p = .037$, with shorter latencies to the central than peripheral (Fig. 2).

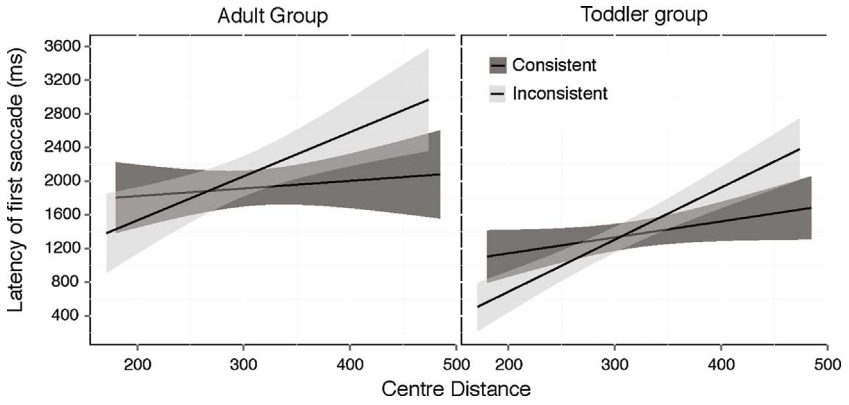

Fig. 2. The effect of centre distance (in pixels) on first saccade latencies for consistent and inconsistent AOIs in adults (left) and toddlers (right). Dispersion depicts 95% of confidence interval.

3.2. Foveal processing

The influence of semantic consistency, perceptual features and linguistic properties on eye movements after the first fixation was examined by running models for dwell time and first-pass gaze duration.

3.2.1. Main model for dwell time

The main model for dwell time revealed a main effect of semantic consistency, $b = 0.35$, $t(63) = 5.84$, $p < .001$, with longer dwell times on inconsistent objects. There was a main effect of age group, $b = -0.15$, $t(65) = -2.68$, $p = .009$, with longer dwell times for toddlers than adults. Furthermore, trial had a main effect, $b = 0.05$, $t(131) = 4.66$, $p < .001$, expressed by longer dwell times as trial number increased. Regarding visual and linguistic features, the model showed a main effect of centre distance, $b = -0.19$, $t(15) = -2.85$, $p = .011$, with longer dwell times for more central objects. There was a main effect of feature congestion, $b = -0.13$, $t(8) = -2.32$, $p = .047$, with longer dwell times for objects in less congested scenes. There was a main effect of word frequency, $b = 0.16$, $t(130) = 4.63$, $p < .001$, based on longer dwell times for objects with more frequent labels. The interaction effect of age group and semantic consistency, $b = 0.39$, $t(61) = 3.68$, $p < .001$, was based on a stronger effect of semantic consistency in adults than toddlers (Fig. 3). Furthermore, trends were obtained for word length, $b = 0.06$, $t(61) = 1.84$, $p = .071$, with longer dwell times for longer object labels; the interaction between semantic consistency and saliency rank, $b = -0.15$, $t(14) = -1.93$, $p = .075$, with a stronger effect of consistency for salient objects; and for the interaction between group and centre distance, $b = 0.08$, $t(853) = 1.89$, $p = .059$, with a stronger effect of centre distance in toddlers (Fig. 4).

To further explore the interaction of age group with semantic consistency we proceeded to split the model by age group, building two identical models for adults and children.

3.2.2. Adult model for dwell time

The adult model for dwell time showed a main effect of semantic consistency, $b = 0.55$, $t(14) = 6.75$, $p < .001$, with longer dwell times for inconsistent objects (Fig. 3) and a main effect of trial, $b = 0.05$, $t(449) = 2.12$, $p = .034$, with longer dwell times as trial increased. There was a trend for word-label frequency, $b = 0.08$, $t(10) = 1.92$, $p = .087$, with longer dwell times to objects with more frequent labels (Fig. 5).

3.2.3. Children model for dwell time

The children model for dwell time showed main effects of centre distance, $b = -0.20$, $t(12) = -3.01$, $p = .011$, with longer dwell times to more central objects; feature congestion, $b = -0.18$, $t(5) = -2.98$, $p < .029$, with longer dwell times to objects in less congested scenes (Fig. 4); and word frequency, $b = 0.24$, $t(12) = 4.54$, $p < .001$, with longer dwell times to objects with more frequent labels (Fig. 5). A trend suggested a main effect of saliency rank, $b = -0.15$, $t(11) = -2.05$, $p = .064$, with longer dwell times to more salient objects and word length, $b = 0.14$, $t(14) = 1.89$, $p = .08$, with longer dwell times to objects with longer labels.

When language was included to the model, dwell time revealed a main effect of expressive vocabulary size, $b = 0.42$, $t(299) = 2.49$, $p = .013$, with longer dwell times for high producers than low producers (Fig. 6). For visual and linguistic features this model showed similar results as the children model: the effects of centre distance, $b = -0.18$, $t(54) = -2.09$, $p = .041$; feature congestion, $b = -0.17$, $t(25) = -3.03$, $p = .006$, and word frequency, $b = 0.34$, $t(55) = 5.02$, $p < .001$. Additionally, this model revealed a main effect of saliency rank, $b = -0.19$, $t(24) = -2.09$, $p = .047$, with longer looking times with increasing saliency. Importantly, the language model revealed an interaction of language production and semantic consistency, $b = -0.19$, $t(307) = -2.03$,. $p = .039$, with a decreasing effect

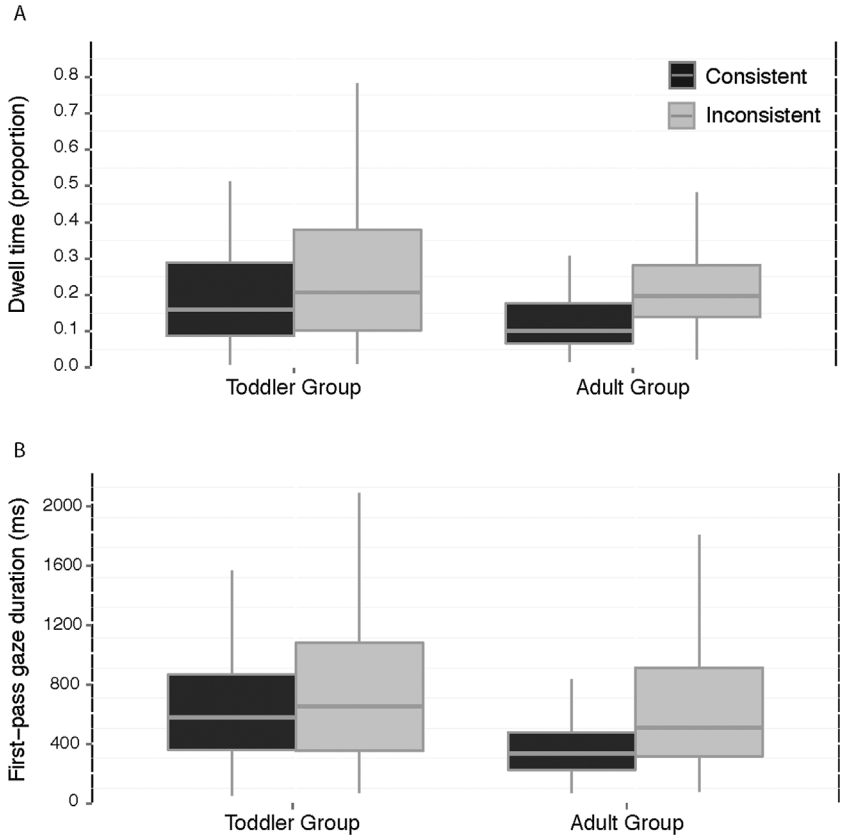

Fig. 3. Dwell times (A) and first-pass gaze durations (B) for consistent and inconsistent scenes in toddlers and adults. Error bars depict interquartile intervals.

of semantic consistency by increasing vocabulary size. A simple slope comparison moving vocabulary size by 2 SD in both directions showed that there was a main effect of semantic consistency for toddlers with low expressive vocabulary size, $b = 0.55$, $t(246) = 2.04$, $p = .043$, but not for toddlers with higher vocabulary size. Expressive vocabulary size did not affect proportion of looking time to inconsistent objects but higher producers looked more to the consistent objects than lower producers. Additionally, there was a significant

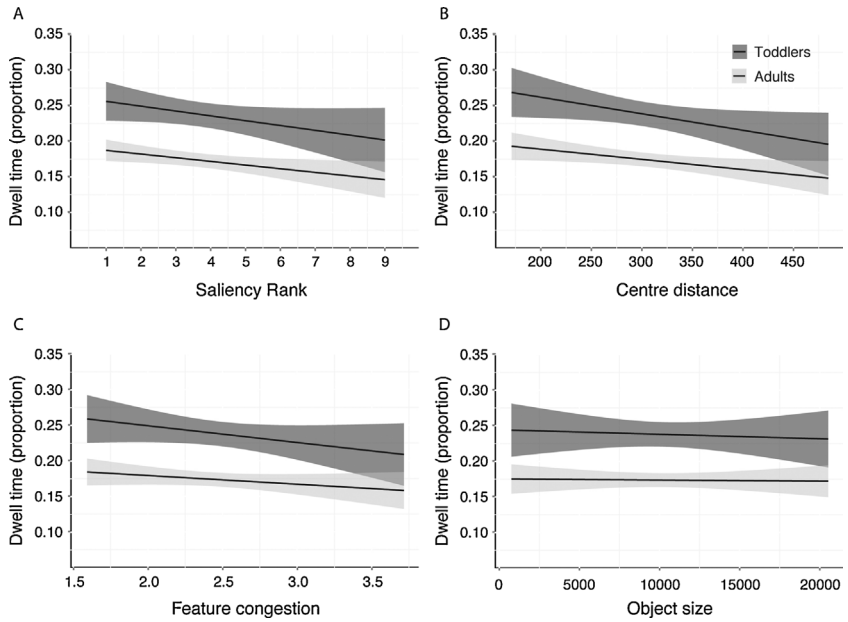

Fig. 4. The effects of (A) saliency rank, (B) centre distance (in pixels), (C) feature congestion score, and (D) object size (in pixels) on dwell times at the AOIs in toddlers and adults. Dispersion depicts 95% of confidence interval.

interaction of language production and word frequency, $b = -0.08$, $t(325) = -2.00$, $p = .046$. Even if all toddlers looked longer to objects with more frequent labels, the simple slope comparison revealed that this effect was stronger for low producers, $b = 0.48$, $t(221) = 3.93$, $p < .001$, than for high producers, $b = 0.19$, $t(60) = 2.97$, $p = .004$. All toddlers looked longer to frequent objects but high producers looked longer to less frequent objects compared to low producers. Finally, a trend suggested an interaction between semantic consistency and saliency rank, $b = -0.36$, $t(24) = -1.75$, $p = .093$, with a stronger effect of consistency for salient objects.

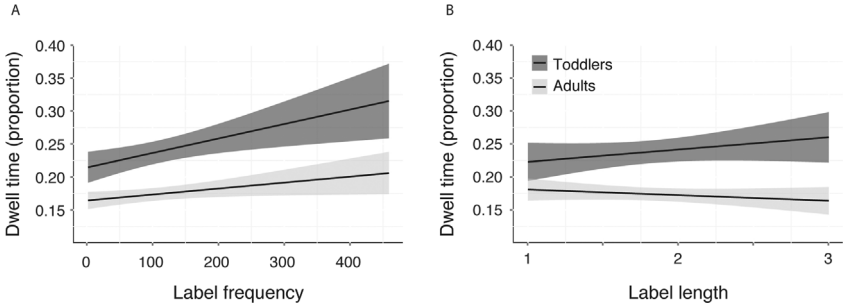

Fig. 5. The effects of label frequency (A) and label length (B) on dwell times in toddlers and adults. Dispersion depicts 95% of confidence interval.

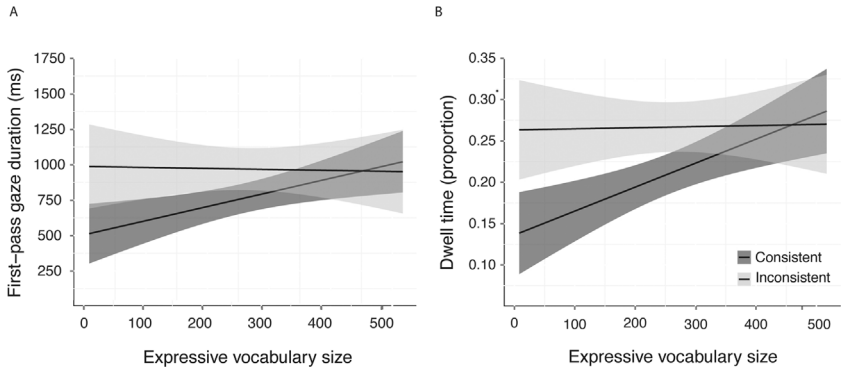

Fig. 6. The effect of expressive vocabulary size on first-pass gaze durations (A) and dwell times (B) for consistent and inconsistent AOIs in toddlers. Dispersion depicts 95% of confidence interval.

3.2.4. Main model for first-pass gaze duration

The main model for first-pass gaze duration revealed a main effect of semantic consistency, $b = -0.29$, $t(12) = 4.66$, $p < .001$, with longer first visits to consistent than inconsistent objects; a main effect of age group, $b = -0.31$, $t(68) = -4.79$, $p < .001$, with longer first visits for toddlers than adults. There was also an interaction of age group with saliency rank, $b = -0.09$, $t(178) = -2.16$, $p = .028$, and centre distance, $b = 0.15$, $t(789) = 3.58$, $p < .001$, with a stronger effect of saliency rank and centre distance on toddlers than adults. The main model for first-pass gaze duration also showed significant effects for linguistic features. There was a main effect of word frequency, $b = 0.07$, $t(57) = 2.51$, $p = .015$, and word length, $b = 0.09$, $t(33) = -2.49$, $p = .018$, with longer first visits to objects with more frequent and longer labels. The model also revealed an interaction of group and semantic consistency, $b = 0.24$, $t(104) = 2.74$, $p = .007$, with a stronger effect of consistency for adults than toddlers; and an interaction between semantic consistency and saliency rank, $b = -0.16$, $t(14) = -2.37$, $p = .035$, showing that with increasing saliency inconsistent objects were looked longer than consistent objects.

To further explore the interaction of age group with saliency rank and centre distance we proceeded to split the model by age group, building two identical models for adults and children.

3.2.5. Adult model for first-pass gaze duration

The adult model showed a main effect of semantic consistency, $b = 0.40$, $t(8) = 5.13$, $p < .001$, with longer first visit to inconsistent objects (Fig. 3), centre distance, $b = 0.20$, $t(13) = 2.88$, $p < .013$, with longer first visits to more central objects. There was also an interaction of semantic consistency and saliency rank, $b = -0.21$, $t(11) = -2.43$, $p < .033$. A simple slope comparison moving saliency rank by 2 SD in both directions showed that there was a main effect of semantic consistency only in highly salient objects, $b = 0.80$, $t(14) = 4.03$, $p = .001$. The simple slope comparison additionally showed an effect of centre distance in low-salient objects only, $b = 0.39$, $t(33) = 3.84$, $p < .001$. In addition, a trend suggested an effect of feature congestion, $b = 0.13$, $t(16) = 1.78$, $p = .09$, with longer first looks to objects in less congested scenes.

3.2.6. Children model for first-pass gaze duration

The children model revealed a main effect of visual and linguistic features. There was a main effect of centre distance, $b = -0.12$, $t(18) = -2.35$, $p = .011$, with longer first visits to more central objects; feature congestion, $b = -0.09$, $t(28) = -2.24$, $p = .026$, with

longer first visits to objects in more congested scenes; and word frequency, $b = 0.18$, $t(98) = 4.69$, $p < .001$, with longer first visits to objects with more frequent labels. Additionally, a trend suggested an effect of word-label length, $b = 0.09$, $t(23) = 1.87$, $p = .075$, with longer first looks to longer words; and an interaction between consistency and saliency rank, $b = -0.27$, $t(17) = -1.47$, $p = .098$.

The language model for first-pass gaze duration revealed a main effect of expressive vocabulary size, $b = 0.44$, $t(304) = 2.69$, $p = .007$, with longer first visits for high producers. Like the main children model, the toddler language model revealed an effect of word frequency, $b = 0.29$, $t(134) = 5.11$, $p < .001$, with longer first visits to objects with higher-frequency labels, and a trend suggesting an effect of label length, $b = 0.13$, $t(35) = 1.75$, $p = .089$. Interestingly, when language skills were included in the analysis of first-pass gaze duration the main effect of visual features showed by the children model disappeared except for a trend of feature congestion, $b = -0.87$, $t(224) = -1.75$, $p = .082$. In contrast, the language model revealed an interaction of semantic consistency and saliency rank, $b = -0.39$, $t(26) = -2.17$, $p = .039$. A simple slope comparison moving saliency rank by 2 SD in both directions showed that there was a main effect of semantic consistency, $b = 0.87$, $t(23) = 2.32$, $p = .029$, only for high-salient objects. In addition, expressive vocabulary size interacted with word frequency, $b = -0.09$, $t(321) = -2.39$, $p = .017$. A simple slope comparison moving expressive vocabulary size by 2 SD in both directions showed that the effect of word frequency was higher for low producers, $b = 0.45$, $t(299) = 4.08$, $p < .001$, than high producers, $b = 0.14$, $t(129) = 2.43$, $p = .016$. As in dwell time, results revealed that both toddlers with higher and lower vocabulary size looked longer to more frequent objects but higher producers looked longer to less frequent objects compared to lower producers.

4. Discussion

In the present study, we investigated how semantic consistency and perceptual features influence attention allocation in 24-month-old children and adults during a free exploration of visual scenes. We furthermore examined the effect of vocabulary skills and linguistic properties of the object labels on the guidance of visual attention. Toddlers and adults fixated objects closer to the image centre faster than peripherally located objects, but the central position of objects facilitated semantic inconsistency detection only in adults. Once an object was fixated, gaze allocation was much stronger influenced by perceptual parameters in toddlers than in adults. Adults revealed a strong semantic consistency effect while the consistency effect in toddlers depended on perceptual features and linguistic skills. Toddlers looked longer at inconsistent than consistent objects only if those were highly salient. In addition, while inconsistent objects attracted the gaze of toddlers regardless of their vocabulary skills, gaze allocation to consistent objects increased with better vocabulary skills.

The probability of reaching the AOIs containing either consistent or inconsistent objects was lower in toddlers than adults. A less explorative behaviour during scene viewing in children than adults has been reported previously (Açik et al., 2010). Likewise, a dominance of focal upon ambient attentional mode during scene exploration in young children compared to adults has been shown recently (Helo et al., 2014). Ambient and focal attentional modes are attentional strategies associated with specific eye movement patterns to explore the visual environment. The ambient mode is expressed by short fixations and large saccade amplitudes and it serves to the orientation in the visual environment. The focal mode is indicated by longer fixations embedded in short saccades and is associated with the identification of object details. Thus, ambient mode it is associated with a more explorative behavior than the focal mode.

Additionally, the latencies of the first saccades to the AOIs were longer in toddlers while both age groups exhibited an equal amount of fixations before reaching them. Therefore, the difference in latency between groups is due to longer fixation durations in toddlers. Previous studies have shown longer fixation durations in children compared to adults during scene perception (Helo et al., 2014, 2016). Longer fixations have been associated with higher cognitive effort during visual processing (Colombo, Mitchell, Coldren, & Freeseman, 1991; Wass & Smith, 2014) and with longer time intervals for the programming of the following saccade (Fukushima, Hatta, & Fukushima, 2000; Gredebäck, Örnkloo, & von Hofsten, 2006; Irving, Steinbach, Lillakas, Babu, & Hutchings, 2006 ;Klein & Foerster, 2001; Luna et al., 2008; Matsuzawa & Shimojo, 1997; Munoz, Broughton, Goldring, & Armstrong, 1998). Altogether, our findings suggest that longer latencies and lower rates to reach an AOI in toddlers refer to lower speed of visual processing and the maturation of visual exploration strategies rather than delayed skills in the processing of semantic scene context.

The main effect of consistency was found for extrafoveal measures neither in toddlers nor in adults. In contrast, we found an effect of centre distance in both age groups (see, Fig. 2). In adults, there was also an interaction of centre distance and semantic inconsistency detection: inconsistent objects were reached faster than consistent objects but only when they were located close to the screen centre, not when they were shown in the periphery. It is worth to note that although the first fixation occurred at the centre of the image it took about 500 ms to reach central inconsistent objects; this duration corresponds to approximately two fixations after the image onset. Further, the latency of the first saccade increased with increasing distance from the centre. These results suggest that, even for central AOIs, the shorter latencies to inconsistent objects might be due to the guidance of proximate fixations to the target rather than an extrafoveal processing during the first fixation. Earlier, it has been suggested that semantic object–scene inconsistencies processing occurs only within a relatively limited area near to the fovea, during viewing of complex scenes (De Graef et al., 1990; Gareze & Findlay, 2007; Võ & Henderson, 2009,2011) and that the previous findings of early effects of semantic inconsistencies might be due to the effect of perceptual features such as object conspicuity, centre distance and image density (Gareze & Findlay, 2007; Henderson et al., 1999; Võ & Henderson, 2009). Our results support this claim in adults but also suggest that proximate fixations are not enough to guide the gaze to inconsistent objects in toddlers since there was no interaction obtained in this group. Previous studies have shown that the size of peripheral visual field increases with age during infancy and early childhood (e.g., Dobson, Brown, Harvey, & Narter, 1998;Sireteanu, Fronius, & Constantinescu, 1994; for review see Maurer, & Lewis, 1991). Thus, it is possible that the lack of the central effect on extrafoveal parameter is due to a reduced peripheral visual field in young children compared to adults.

Perceptual features had a stronger influence on gaze allocation in toddlers than in adults (see, Fig. 4). When the dwell time was analysed, toddlers presented longer looking times to objects that were more salient and closer to the centre of the image than to less salient

and peripheral objects. Toddlers also looked more at objects that are embedded in less cluttered scenes. The effects of centre distance and clutter of the scene were already present during the first visit to the AOIs in toddlers. In contrast, perceptual features had no effect on gaze allocation in adults for dwell time and only the centre distance had an effect during the first visit to the AOIs in the adult group. These results support previous developmental findings showing a stronger influence of saliency on gaze guidance in children than in adults (Açik et al., 2010; Helo et al., 2014; Kooiker, Van Der, & Pel, 2016). They also suggest that centre distance and feature congestion have a strong influence on eye movement guidance in toddlers. In addition, our findings in adult participants corroborate earlier findings showing that low-level features have a stronger effect on gaze guidance during the early stages of scene exploration whereas in the later stages top-down control dominates visual attention guidance (Castelhano et al., 2009; Mannan et al., 1995; Parkhurst et al., 2002; Tatler & Vincent, 2008). Furthermore, our findings suggest that the interaction between bottom-up and top-down factors in the guidance of visual attention during viewing time differs in toddlers and adults. In toddlers, perceptual features have a dominating role during the whole viewing time confirming a stronger influence of bottom-up factors on visual attention guidance.

Consistency effects on foveal parameters were found only in the adult group (see, Fig. 3). However, both age groups showed an interaction of semantic consistency and saliency rank for first-pass gaze duration. All participants looked longer to the inconsistent than the consistent objects with a high saliency rank but the consistency effect diminished with decreasing saliency. Together with the lack of consistency effect in toddlers, this finding suggests that even though saliency enhanced the consistency effect in both age groups, toddlers detected the inconsistencies only when the objects were salient. It is possible that at the age of two years the visual system is not yet capable of processing visual saliency and semantic object-scene relations in parallel, and processing of semantic inconsistencies is slower than that of perceptual features. Thus, in young children the processing of semantic object-scene relations might be dependent on visual saliency, that is, visual attention might be attracted by saliency first, facilitating semantic inconsistency detection. In fact, it has been previously shown that 15 month-old children process faster perceptual than semantic properties of natural scenes (Duh and Wang, 2014).

In addition to object saliency, vocabulary skills interacted with semantic consistency for dwell time in the toddlers. Toddlers with high vocabulary skills exhibited longer looking times to consistent objects than toddlers with lower vocabulary skills resulting in a reduced consistency effect (see, Fig. 6). However, visual features attracted the gaze of toddlers independently of their linguistic skills. This finding indicates that language skills influence more cognitive than perceptual guidance during scene perception. Vocabulary skills did not correlate with initial fixation durations or total trial dwell times, suggesting that differences between high and low producers were not explained by different cognitive efforts in understanding the scene context or by their engagement to the task. In adults, the semantic consistency effect has been shown using different type of stimuli (e.g., line drawings, 3D pictures, real-world scenes) and different visual tasks (e.g., object or letter searching and memory tasks) (Henderson et al., 1999; Hwang, Wang, & Pomplun, 2011; Underwood & Foulsham, 2006; Underwood et al., 2007; Võ & Henderson, 2009,2011) indicating the robustness of the effect. However, when participants had to name objects in a scene, their gaze allocation to consistent objects increased (Coco et al., 2014), suggesting that a simultaneous linguistic process covered the semantic inconsistency guidance. At the end of the second year of life there is an extensive improvement in vocabulary skills − the so-called vocabulary spurt (Ganger & Brent, 2004; Nazzi & Bertoncini, 2003), 2003). This process is not homogenous, and as a result, the productive vocabulary of a 2-year-old child can vary from few words to hundreds of words. During this developmental stage, toddlers might be attracted to objects and its labels guiding their visual exploration. In effect, young children are shown to activate the labels of objects silently during exploration of visual displays (Mani and Plunkett, 2010, 2011) and silent naming has shown to be stronger in 24-month-olds than in adults (Khan, 2013). Additionally, young children are shown to be prone to spontaneous naming during the manipulation of novel objects and the amount of naming has been associated with their productive vocabulary skills (Samuelson & Smith, 2005). Consequently, it is possible that toddlers with higher vocabulary skills were silently naming the objects more often than those with lower vocabulary skills, thus directing their attention to objects in general.

However, our further analysis did not fully support our naming hypothesis. Earlier empirical evidence in adults has shown that linguistic properties such as frequency of object labels influence the gaze allocation during an explicit naming task of objects presented in a natural scene (Clarke et al., 2013; Coco et al., 2014). Therefore, if high producers were more likely to name than low producers, we would expect that linguistic properties of object labels would have a stronger influence on visual attention allocation in toddlers with higher vocabulary skills. However, when linguistic properties of object labels were analysed we observed that all toddlers, high and low producers, looked longer to objects with more frequent than less frequent labels. We also found that all toddlers tended to look longer to objects with longer label lengths during their first visit to the AOI. These findings suggest that toddlers, regardless of their vocabulary size, were silently naming objects during scene exploration. An alternative explanation is that toddlers with advanced vocabulary skills were more attracted to object shapes in their visual environment but this “object bias” was not directly linked to naming. Attention to object shape and the rate of noun acquisition have shown to be dependent on each other: as children's vocabularies grow, their attention to object shape increases (Gershkoff-Stowe & Smith, 2004; Pereira, Smith, & Yu, 2014; Smith, 2003). Also, objects that appear most frequently in 8- to 10-month-old infants’ egocentric views (in everyday visual environments) have shown to correspond with those objects labels that are learned first suggesting that visual environment may guide early acquisition of words (Clerkin, Hart, Rehg, Yu, & Smith, 2017).

Even naming and “object bias” are possible explanations for differences in gaze distribution in high and low producers, we are not able to fully clarify why high producers looked longer to consistent objects based on our current findings. On one hand, we cannot completely rule out the possibility of stronger implicit naming in high producers since the objects shown in our experiment were all very familiar objects with relatively short labels. If the object labels were more variable in terms of frequency and lengths, differences in linguistic guidance between high and low producers might have been stronger. On the other hand, we cannot confirm the presence of the object bias in high producers since gaze allocation was compared only between the consistent and inconsistent AOIs, and looking times to other objects within the scenes were not analysed. Nevertheless, our results indicate that by the age of two years, toddlers are able to use conceptual schemes of their everyday environment to guide further exploration of complex visual scenes. The

results also indicate that expressive vocabulary size affects gaze allocation in 2-year-old children demonstrating an interaction between visual attention guidance and linguistic skills during scene exploration, even in the absence of a linguistic task. A further study using more complex objects labels and controlling the number of objects in each scene is needed to further study the cognitive mechanisms underlying this phenomenon.

Acknowledgements

We express our gratitude to the infants and parents who participated to the study, we thank for their kindness and cooperation. A.H. was supported by a doctoral fellowship from CONICYT, Chile.

Appendix A. Parameter tables

Tables A1 and A2

Table A1
Extrafoveal models: all parameters. Grey marks significant coefficients.

	Extrafoveal		
	First Saccade Start		
	Overall	Adult	Infant
N (Subjects, Scenes, Objects)	968 (72,18,18)	520 (30,18,18)	448 (42,18,18)
Intercept	b=7.164, t(16.70)=36.645, p=0.000***	b=7.169, t(24.30)=31.527, p=0.000***	b=7.335, t(10.20)=28.581, p=0.000***
Age Group	b=-0.443, t(75.60)=-7.771, p=0.000***		
Consistency	b=0.173, t(9.60)=2.236, p=0.050.	b=0.082, t(15.00)=0.926, p=0.369	b=0.184, t(35.10)=1.685, p=0.101
Saliency Rank	b=-0.048, t(3.40)=-0.842, p=0.455	b=0.029, t(10.30)=0.486, p=0.637	b=-0.032, t(7.70)=-0.356, p=0.731
Centre Distance	b=0.336, t(14.40)=3.799, p=0.002**	b=0.481, t(26.40)=4.624, p=0.000***	b=0.221, t(16.80)=2.265, p=0.037*
Feature Congestion	b=0.142, t(9.20)=1.607, p=0.142	b=0.219, t(15.00)=1.951, p=0.070.	b=0.106, t(9.00)=1.208, p=0.258
AOI size	b=-0.026, t(12.00)=-0.438,	b=-0.001, t(15.80)=-0.009,	b=-0.004, t(8.70)=-0.053,

(continued on next page)

Table A1 (*continued*)

	p=0.669		p=0.993		p=0.959	
Log Frequency	b=-0.047, p=0.310	t(12.40)=-1.060,	b=-0.097, p=0.074.	t(18.70)=-1.894,	b=-0.042, p=0.500	t(7.20)=-0.710,
Number of syllables	b=0.012, p=0.771	t(23.10)=0.294,	b=0.097, t(5.40)=1.748, p=0.136		b=-0.050, p=0.483	t(3.00)=-0.797,
Trial	b=-0.063, p=0.013*	t(898.60)=-2.478,	b=-0.084, p=0.010**	t(467.10)=-2.594,	b=-0.027, p=0.474	t(401.10)=-0.717,
Age Group * Consistency	b=-0.031, p=0.769	t(78.30)=-0.295,				
Age Group * Saliency Rank	b=-0.081, p=0.134	t(224.00)=-1.505,				
Consistency * Saliency Rank	b=0.149, t(6.10)=1.041, p=0.337		b=0.080, t(4.80)=0.557, p=0.603		b=0.199, t(7.30)=1.231, p=0.257	
Age Group * Centre Distance	b=0.143, p=0.006**	t(815.20)=2.783,				
Consistency * Centre Distance	b=0.097, p=0.189	t(27.30)=1.348,	b=0.188, p=0.048*	t(9.70)=2.267,	b=0.061, p=0.573	t(20.50)=0.572,
Age Group * Consistency * Saliency Rank	b=-0.130, p=0.227	t(139.70)=-1.214,				
Age Group * Consistency * Centre Distance	b=0.126, p=0.229	t(826.30)=1.203,				

Table A2
Foveal models: all parameters. Grey marks significant coefficients.

	Foveal							
	Proportion of looking time				First Fixation Duration			
	Overall	Adult	Infant	Language	Overall	Adult	Infant	Language
N (Subjects, Scenes, Objects)	976 (72,18,18)	522 (30,18,18)	454 (42,18,18)	354 (33,18,18)	976 (72,18,18)	522 (30,18,18)	454 (42,18,18)	354 (33,18,18)
Intercept	b=-2.551, t(57.20)=-15.562, p=.000***	b=-2.315, t(12.30) =-11.486, p=0.000***	b=-2.814, t(12.50)=-12.760, p=0.000***	b=-3.307, t(59.00)=-11.447, p=0.000***	b=5.916, t(50.00)=47.505, p=0.000***	b=6.065, t(19.80)=34.899, p=0.000***	b=5.585, t(85.60)=33.237, p=0.000***	b=5.109, t(130.10)=20.525, p=0.000***
Expressive vocabulary score				b=0.426, t(309.40)=2.492, p=0.013*				b=0.436, t(304.20)=2.696, p=0.007**
Age Group	b=-0.154, t(65.10)=-2,679, p=0.009**				b=-0.313, t(68.10)=-4.791, p=0.000***			
Consistency	b=0.352, t(62.90)=5.839, p=.000***	b=0.553, t(14.90) =6.745, p=0.000***	b=0.049, t(24.80)=0.440, p=0.664	b=0.213, t(54.60)=1.566, p=0.123	b=0.292, t(11.50)=4.659, p=0.001***	b=0.404, t(7.90)=5.125, p=0.001***	b=-0.021, t(21.60)=-0.206, p=0.838	b=0.115, t(57.80)=0.914, p=0.365
Saliency Rank	b=0.020, t(11.80)=0.495,	b=-0.035, t(18.40) =-0.639,	b=-0.149, t(11.40)=-2.053,	b=-0.190, t(25.30)=-2.107,	b=-0.025, t(23.70)=-0.707,	b=-0.076, t(7.90)=-1.465,	b=-0.094, t(13.80)=-1.741,	b=-0.051, t(29.20)=-0.645,

(*continued on next page*)

Table A2 (*continued*)

	p=0.630	p=0.531	p=0.064.	p=0.045*	p=0.486	p=0.182	p=0.104	p=0.524
Centre Distance	b=-0.194, t(15.30)=-2.850, p=0.012*	b=-0.135, t(17.30) =-1.534, p=0.143	b=-0.203, t(11.60)=-3.017, p=0.011*	b=-0.177, t(54.20)=-2.108, p=0.040*	b=0.020, t(17.90)=0.375, p=0.712	b=0.199, t(13.30)=2.878, p=0.013*	b=-0.121, t(179.10)=-2.350, p=0.020*	b=-0.091, t(213.60)=-1.223, p=0.223
Feature Congestion	b=-0.127, t(8.20)=-2.323, p=0.048*	b=-0.095, t(8.90) =-1.039, p=0.326	b=-0.177, t(5.20)=-2.979, p=0.029*	b=-0.172, t(25.30)=-3.022, p=0.006**	b=-0.037, t(14.60)=-0.617, p=0.546	b=0.133, t(15.80)=1.767, p=0.096.	b=-0.098, t(278.30)=-2.240, p=0.026*	b=-0.087, t(224.20)=-1.749, p=0.082.
AOI size	b=-0.039, t(127.80)=-1.003, p=0.318	b=-0.062, t(17.70) =-0.949, p=0.356	b=0.003, t(27.20)=0.043, p=0.966	b=0.015, t(54.00)=0.236, p=0.814	b=0.019, t(123.80)=0.482, p=0.631	b=-0.031, t(12.60)=-0.540, p=0.598	b=0.001, t(148.50)=0.020, p=0.984	b=-0.012, t(138.80)=-0.215, p=0.830
Log Frequency	b=0.162, t(130.70)=4.663, p=.000***	b=0.085, t(9.80) =1.902, p=0.087.	b=0.236, t(11.60)=4.535, p=0.001***	b=0.338, t(57.50)=5.010, p=0.000***	b=0.070, t(57.00)=2.514, p=0.015*	b=-0.007, t(16.30)=-0.180, p=0.859	b=0.186, t(98.40)=4.698, p=0.000***	b=0.295, t(133.80)=5.106, p=0.000***
Number of syllables	b=0.063, t(60.80)=1.840, p=0.071.	b=-0.024, t(12.90) =-0.566, p=0.581	b=0.139, t(14.10)=1.885, p=0.080.	b=0.139, t(27.90)=1.535, p=0.136	b=0.093, t(32.90)=2.491, p=0.018*	b=0.009, t(5.90)=0.196, p=0.851	b=0.098, t(23.00)=1.864, p=0.075.	b=0.126, t(35.10)=1.749, p=0.089.
Trial	b=0.048, t(879.60)=2.252, p=0.025*	b=0.047, t(449.90) =2.122, p=0.034*	b=0.040, t(403.60)=1.100, p=0.272	b=0.047, t(308.00)=1.105, p=0.270	b=0.028, t(877.00)=1.330, p=0.184	b=0.034, t(455.70)=1.363, p=0.173	b=0.005, t(400.40)=0.157, p=0.875	b=0.018, t(316.40)=0.448, p=0.655
Age Group * Consistency	b=0.397, t(61.10)=3.676, p=0.001***				b=0.247, t(103.50)=2.740, p=0.007**			

(*continued on next page*)

Table A2 (*continued*)

Expressive vocabulary score * Consistency				b=-0.191, t(307.10)=-2.032, p=0.043*				b=-0.133, t(306.80)=-1.542, p=0.124
Age Group * Saliency Rank	b=-0.032, t(169.40)=-0.694, p=0.489				b=-0.097, t(177.90)=-2.216, p=0.028*			
Expressive vocabulary score * Saliency Rank				b=-0.001, t(321.50)=-0.015, p=0.988				b=-0.031, t(319.60)=-0.626, p=0.532
Consistency * Saliency Rank	b=-0.148, t(13.90)=-1.928, p=0.075.	b=-0.130, t(5.40) =-1.333, p=0.236	b=-0.190, t(10.40)=-1.101, p=0.296	b=-0.361, t(24.10)=-1.746, p=0.094.	b=-0.159, t(13.60)=-2.337, p=0.035*	b=-0.208, t(10.90)=-2.426, p=0.034*	b=-0.273, t(17.10)=-1.747, p=0.099.	b=-0.396, t(26.80)=-2.170, p=0.039*
Age Group * Centre Distance	b=0.083, t(853.00)=1.886, p=0.060.				b=0.151, t(789.40)=3.575, p=0.000***			
Expressive vocabulary score * Centre Distance				b=-0.027, t(306.70)=-0.541, p=0.589				b=-0.025, t(307.10)=-0.541, p=0.589
Consistency * Centre Distance	b=0.030, t(53.90)=0.573, p=0.569	b=0.034, t(9.10) =0.548, p=0.597	b=-0.015, t(22.10)=-0.143, p=0.887	b=-0.015, t(96.50)=-0.097, p=0.923	b=-0.044, t(23.60)=-0.761, p=0.454	b=-0.015, t(7.90)=-0.197, p=0.849	b=0.006, t(87.60)=0.059, p=0.953	b=-0.028, t(145.00)=-0.201, p=0.841
Age Group * Consistency *	b=-0.049, t(313.90)=-0.548,				b=-0.153, t(71.80)=-1.613,			

(*continued on next page*)

Table A2 (*continued*)

Saliency Rank	p=0.584		p=0.111	
Expressive vocabulary score * Log Frequency		b=-0.080, t(321.90)=-2.005, p=0.046*		b=-0.089, t(321.20)=-2.393, p=0.017*
Expressive vocabulary score * Number of syllables		b=-0.014, t(323.60)=-0.288, p=0.773		b=-0.064, t(323.50)=-1.382, p=0.168
Expressive vocabulary score * Consistency * Saliency Rank		b=0.145, t(318.00)=1.384, p=0.167		b=0.083, t(315.30)=0.860, p=0.390
Age Group * Consistency * Centre Distance	b=0.063, t(850.80)=0.707, p=0.480		b=-0.010, t(856.00)=-0.117, p=0.907	
Expressive Vocabulary Score * Consistency * Centre Distance		b=0.119, t(304.60)=1.229, p=0.220		b=0.099, t(320.00)=1.083, p=0.280

References

Açik, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental changes in natural viewing behavior: Bottomup and top-down differences between children, young adults and older adults. *Frontiers in Psychology, 1*, 1–14. http://dx.doi.org/10.3389/fpsyg.2010.00207.

Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica, 137*(2), 208–216. http://dx.doi.org/10.1016/j.actpsy.2011.01.007.

Bååth, R. (2010). ChildFreq: An online tool to explore word frequencies in child language. *LUCS Minor, 16*, 1–6.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Barlett, F. C. (1932). *Remembering: A study in experimental and social psychology (Cambridge)*. UK: Cambridge.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001.

Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades: Journal of Experimental Psychology. *Human Perception and Performance, 33*(1), 20–30. http://dx.doi.org/10.1037/0096-1523.33.1.20.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177. http://dx.doi.org/10.1016/0010-0285(82)90007-X.

Bornstein, M. H., Arterberry, M. E., & Mash, C. (2010). Infant object categorization transcends diverse object-context relations. *Infant Behavior and Development, 33*(1), 7–15. http://dx.doi.org/10.1016/j.infbeh.2009.10.003.

Bornstein, M. H., Mash, C., & Arterberry, M. E. (2011a). Perception of object-contet relations: Eye-movement analyses in infants and adults. *Developmental Psychology, 47*(2), 364–375. http://dx.doi.org/10.1037/a0021059.Perception.

Bornstein, M. H., Mash, C., & Arterberry, M. E. (2011b). Young infants' eye movements over natural scenes and experimental scenes. *Infant Behavior and Development, 34*(1), 206–210. http://dx.doi.org/10.1016/j.infbeh.2010.12.010.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 753–763. http://dx.doi.org/10.1037/0096-1523.33.4.753.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3), 1–15. http://dx.doi.org/10.1167/9.3.6.Introduction.

Clarke, A. D. F., Coco, M. I., & Keller, F. (2013). The impact of attentional, linguistic, and visual features during object naming. *Frontiers in Psychology, 4*, 1–12. http://dx.doi.org/10.3389/fpsyg.2013.00927.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160055. http://dx.doi.org/10.1098/rstb.2016.0055.

Coco, M. I., Malcolm, G. L., & Keller, F. (2014). The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. *Quarterly Journal of Experimental Psychology, 67*(6), 1096–1120. http://dx.doi.org/10.1080/17470218.2013.844843.
Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual differences in infant visual attention: are short lookers faster processors or feature processors? *Child Development, 62*(6), 1247–1257. http://dx.doi.org/10.1111/j.1467-8624.1991.tb01603.x.
Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition, 35*(3), 393–401. http://dx.doi.org/10.3758/BF03193280.
De Graef, P., Christieaens, D., & D'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*, 317–329.
Dobson, V., Brown, A., Harvey, E., & Narter, D. (1998). Visual field extent in children 3. 5-30 months of age tested with a double-arc LED perimeter. *Vision Research, 38*(18), 2743–2760.
Duh, S., & Wang, S.-H. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology, 72*, 142–161. http://dx.doi.org/10.1016/j.cogpsych.2014.03.001.
Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2), 2.1–19. http://dx.doi.org/10.1167/8.2.2.
Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., & Reilly, J. S. (1993). *The MacArthur communicative development inventories: User's guide and technical manual. (P. H, Ed.)*. Baltimore: Brokes Publishing Co.
Fischer, T., Graupner, S.-T., Velichkovsky, B. M., & Pannasch, S. (2013). Attentional dynamics during free picture viewing: Evidence from oculomotor behavior and electrocortical activity. *Frontiers in Systems Neuroscience, 7*(17), 1–9. http://dx.doi.org/10.3389/fnsys.2013.00017.
Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(2), 1–17. http://dx.doi.org/10.1167/8.2.6.
Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements. I. Age-related changes in normal children. *Brain & Development, 22*(3), 173–180. http://dx.doi.org/10.1016/S0387-7604(00)00101-7.
Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology, 40*(4), 621–632. http://dx.doi.org/10.1037/0012-1649.40.4.621.
Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. Hill (Eds.). *Eye movements: A window on mind and brain* (pp. 617–637). Amsterdam: Elsevier.
Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410–419. http://dx.doi.org/10.1016/S1364-6613(99)01388-1.
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. http://dx.doi.org/10.1037/0012-1649.40.2.177.
Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development, 75*(4), 1098–1114. http://dx.doi.org/10.1111/j.1467-8624.2004.00728.x.
Gredebäck, G., Örnkloo, H., & von Hofsten, C. (2006). The development of reactive saccade latencies. *Experimental Brain Research, 173*(1), 159–164. http://dx.doi.org/10.1007/s00221-006-0376-z.
Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research, 103*, 83–91. http://dx.doi.org/10.1016/j.visres.2014.08.006.
Helo, A., Rämä, P., Pannasch, S., & Meary, D. (2016). Ambient and focal visual attention during scene viewing in 3- to 12-month-olds. *Visual Neuroscience* in press.
Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498–504. http://dx.doi.org/10.1016/j.tics.2003.09.006.
Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World, 2004*, 1–58. http://dx.doi.org/10.4324/9780203488430.
Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243–271. http://dx.doi.org/10.1146/annurev.psych.50.1.243.
Henderson, J. 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*(2009), 1–8. http://dx.doi.org/10.1167/9.1.32.Introduction.
Henderson, J. M., Weeks, P. A. J., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228. http://dx.doi.org/10.1037/0096-1523.25.1.210.
Hitch, G. J., Halliday, S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition, 16*(2), 120–132. http://dx.doi.org/10.3758/BF03213479.
Hock, H. S., Romanski, L., Galie, A., & Williams, C. S. (1978). Real-world schemata and scene recognition in adults and children. *Memory & Cognition, 6*(4), 423–431. http://dx.doi.org/10.3758/BF03197475.
Hwang, A. D., Wang, H. C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research, 51*(10), 1192–1205. http://dx.doi.org/10.1016/j.visres.2011.03.010.
Irving, E. L., Steinbach, M. J., Lillakas, L., Babu, R. J., & Hutchings, N. (2006). Horizontal saccade dynamics across the human life span. *Investigative Ophthalmology and Visual Science, 47*(6), 2478–2484. http://dx.doi.org/10.1167/iovs.05-1311.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489–1506. http://dx.doi.org/10.1016/S0042-6989(99)00163-7.
Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience, 2*(3), 194–203. http://dx.doi.org/10.1038/35058500.
Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review, 27*(3), 283–348. http://dx.doi.org/10.1016/j.dr.2007.06.006.
Khan, M. (2013). *Thinking in Words: Implicit Verbal Activation in Children and Adults*.
Klein, C., & Foerster, F. (2001). Development of prosaccade and antisaccade task performance in participants aged 6 to 26 years. *Psychophysiology, 38*(2), 179–189. http://dx.doi.org/10.1017/S0048577201981399.
Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential Development of Attention and Executive Functions in 3- to 12-Year-old Finnish Children. *Developmental Neuropsychology, 20*(1), 407–428. http://dx.doi.org/10.1207/S15326942DN2001.
Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiolgy, 4*, 219–227.
Kooiker, M. J. G., Van Der, Steen J., & Pel, J. J. M. (2016). Development of salience-driven and visually-guided eye movement responses. *Journal of Vision, 16*(5), 1–11. http://dx.doi.org/10.1167/16.5.18.
Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research, 47*(19), 2483–2498. http://dx.doi.org/10.1016/j.visres.2007.06.015.
Loftus, G. R., & Mackworth, N. H. (1978). LoftusMackworth1978.pdf. *Journal of Experimental Psychology. Human Perception and Performance, 4*(4), 565–572.
Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition, 68*(3), 293–308. http://dx.doi.org/10.1016/j.bandc.2008.08.019.
Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology. Human Learning and Memory, 2*(5), 529–540. http://dx.doi.org/10.1037//0278-7393.2.5.529.
Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: evidence for implicit naming in 18-month-olds. *Research Report, 21*(7), 908–913. http://dx.doi.org/10.1177/0956797610373371.
Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition, 121*(2), 196–206. http://dx.doi.org/10.1016/j.cognition.2011.06.013.
Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9*(3), 363–386. http://dx.doi.org/10.1163/156856895X00052.
Matsuzawa, M., & Shimojo, S. (1997). Infants' fast saccades in the gap paradigm and development of visual attention. *Infant Behavior and Development, 20*(4), 449–455. http://dx.doi.org/10.1016/S0163-6383(97)90035-7.
Maurer, D., & Lewis, T. L. (1991). The development of peripheral vision and its physiological underpinnings. In M. J. S. Weiss, & P. R. Zelazo (Eds.). *Newborn attention: Biological constraints and the influence of experience* (pp. 218–255). Westport, CT: Ablex Publishing.
Mills, M., Hollingworth, A., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision, 11*(8), 1–15. http://dx.doi.org/10.1167/11.8.17.Introduction.

Munoz, D. P., Broughton, J. R., Goldring, J. E., & Armstrong, I. T. (1998). Age-related performance of human subjects on saccadic eye movement tasks. *Experimental Brain Research, 121*(4), 391–400. http://dx.doi.org/10.1007/s002210050473.
Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science, 6*(2), 136–142. http://dx.doi.org/10.1111/1467-7687.00263.
Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences, 1339*(1), 82–96. http://dx.doi.org/10.1111/nyas.12705.
Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing – both initially and overall. *Journal of Eye-Movement Research, 2*(2), 2:1–2:11.
Oliva, A. (2005). Gist of the scene. *Neurobiology of Attention,* 251–256. http://dx.doi.org/10.1016/B978-012375731-9/50045-8.
Parkhurst, D., Law, K., & Niebur, E. (2002). Modelling the role of salience in the allocation of visual selective attention. *Vision Research, 42*(1), 107–123.
Pearson, D., & Lane, D. M. A. (1991). Auditory attention switching: a developmental study. *Journal of Experimental Child Psychology, 51*(2), 320–334. http://dx.doi.org/10.1016/0022-0965(91)90039-U.
Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review, 21*(1), 178–185. http://dx.doi.org/10.3758/s13423-013-0466-4.
Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory, 9*(4-6), 423–432. http://dx.doi.org/10.1080/09658210143000182.
Potter, M. C. (1975). Meaning in visual search. *Science, 187*(4180), 965–966.
Potter, M. C. (1976). Journal of Experimental Psychology: Human Learning and Memory. *Journal of Experimental Psychology. Human Learning and Memory, 2*(5), 509–522.
Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 1–22. http://dx.doi.org/10.1167/7.2.17.
Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: Spontaneous naming and attention to shape. *Developmental Science, 8*(2), 182–198. http://dx.doi.org/10.1111/j.1467-7687.2005.00405.x.
Sanders, L. D., Stevens, C., Coch, D., & Neville, H. J. (2006). Selective auditory attention in 3- to 5-year-old children: An event-related potential study. *Neuropsychologia, 44*(11), 2126–2138. http://dx.doi.org/10.1016/j.neuropsychologia.2005.10.007.
Sireteanu, R., Fronius, M., & Constantinescu, D. (1994). The development of visual acuity in the peripheral visual field of human infants: binocular and monocular measurements. *Vision Research, 34*(12), 1659–1671.
Smith, L. B. (2003). Learning to recognize objects. *Psychological Science, 14*(3), 244–250.
Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 4.1–17. http://dx.doi.org/10.1167/7.14.4.
Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2*(2), 1–18.
Theeuwes, J. (2010). Top-down and bottom-up control of visual selection: Reply to commentaries. *Acta Psychologica, 135*(2), 133–139. http://dx.doi.org/10.1016/j.actpsy.2010.07.006.
Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review, 113*(4), 766–786. http://dx.doi.org/10.1037/0033-295X.113.4.766.
Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology, 13*(4), 428–432. http://dx.doi.org/10.1016/S0959-4388(03)00105-3.
Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology (2006), 59*(11), 1931–1949. http://dx.doi.org/10.1080/17470210500416342.
Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology, 18*(3), 321–342. http://dx.doi.org/10.1080/09541440600604248.
Underwood, G., Humphreys, L., & Cross, E. (2007). Congruency, saliency and gist in the inspection of objects in natural scenes. *Eye Movements: A Window on Mind and Brain,* 564–579. http://dx.doi.org/10.1016/B978-008044980-7/50028-8.
Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17*(1), 159–170. http://dx.doi.org/10.1016/j.concog.2006.11.008.
Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9*(3), 1–15. http://dx.doi.org/10.1167/9.3.24.
Võ, M. L.-H., & Henderson, J. M. (2011). Object-scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm Attention. *Perception & Psychophysics, 73*(6), 1742–1753. http://dx.doi.org/10.3758/s13414-011-0150-6.
Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395–1407. http://dx.doi.org/10.1016/j.neunet.2006.10.001.
Wass, S. V., & Smith, T. J. (2014). Individual Differences in Infant Oculomotor Behavior During the Viewing of Complex Naturalistic Scenes. *Infancy, 19*(4), 352–384. http://dx.doi.org/10.1111/infa.12049.
Yarbus, A. L. (1967). Eye movements and vision. *Neuropsychologia, 6*(4), 389–390. http://dx.doi.org/10.1016/0028-3932(68)90012-2.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R Package Version, *3.0.0*, https://cran.r-project.org/package=lmerTest. Retrieved from https://cran.r-project.org/package=lmerTest(/C ((/C(.