



**HAL**  
open science

# Scene exploration during development : influence of perceptual features and semantic context on visual attention

Andrea Helo

► **To cite this version:**

Andrea Helo. Scene exploration during development : influence of perceptual features and semantic context on visual attention. Psychology. Université Sorbonne Paris Cité, 2016. English. NNT : 2016USPCB205 . tel-02016309

**HAL Id: tel-02016309**

**<https://theses.hal.science/tel-02016309>**

Submitted on 12 Feb 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***Exploration de scènes visuelles au cours du développement: influence des caractéristiques perceptuelles et des contextes sémantiques sur l'attention visuelle***

***Scene exploration during development: influence of perceptual features and semantic context on visual attention***

Thèse de Doctorat, Neurosciences Cognitives

Andrea Helo Directrice de Thèse : Pia Rämä

Co-directeur : Sebastian Pannasch

Laboratoire Psychologie de la Perception

CNRS-UMR 8242



Date de soutenance Jeudi 10 Novembre, 2016

Membres de Jury :

Olivier Le Meur	Rapporteur
Olivier Pascalis	Rapporteur
Karine Doré-Mazars	Examineur
Melissa Le-Ho Võ	Examineur
Sebastian Pannasch	Co-directeur de thèse
Pia Rämä	Directrice de thèse

## Résumé

Bien que le développement des mécanismes sous-tendant les fonctions oculomotrices ait été largement exploré dans des tâches visuelles dites *stimuli-driven*, le développement des comportements oculomoteurs lors de l'exploration libre de scènes visuelles reste peu connu. L'objectif de cette thèse est d'étudier le développement des mécanismes sous-tendant l'exploration de scènes visuelles, naturelles et statiques au moyen des techniques d'*eye-tracking* et de potentiels évoqués (PEs). Les deux objectifs principaux étaient d'explorer le développement des stratégies d'attention visuelle du début à la fin de l'enfance ainsi que d'étudier l'interaction entre les contrôles *bottom up* et *top down* durant l'exploration de scènes visuelles chez les jeunes enfants.

Pour les études I et II, des analyses en *eye-tracking* ont été réalisées afin d'étudier le développement des fonctions oculomotrices (durée de fixation et amplitude des saccades) et des stratégies d'attention visuelle (ambiante *versus* focale) lors de l'exploration d'une scène visuelle de 3 mois à 10 ans. Les résultats ont montré que les durées de fixation diminuent de l'âge de 12 mois jusqu'à l'âge de 6 à 8 ans, tandis que l'amplitude des saccades atteignent des valeurs adultes vers l'âge de 4 à 6 ans. Des études précédentes ont montré qu'au cours de l'exploration d'une scène visuelle, les durées de fixation ont tendance à augmenter tandis que l'amplitude des saccades diminue. Ce décours temporel dans les durées de fixations et l'amplitude des saccades a été associé aux stratégies ambiante et focale durant l'exploration d'une scène visuelle. Un décours temporel similaire a d'ailleurs été observé pour les durées de fixation à l'âge de 12 mois mais pas chez les enfants plus jeunes, suggérant que les modes ambiant et focal émergent aux alentours de cet âge. A l'âge de 2 ans, le décours temporel associé aux modes ambiant et focal a été trouvé dans les deux paramètres: fixation et saccades. Cependant l'exploration visuelle était dominée par le mode focal jusqu'à l'âge de 6 ans. De plus, la saillance visuelle guidait davantage les comportements oculomoteurs des enfants âgés de moins de 6 ans, par rapport aux enfants plus âgés et aux adultes.

Dans l'étude III, les mouvements oculaires ont été enregistrés afin d'étudier l'influence du contrôle *top-down* (cohérence sémantique et linguistique) et *bottom-up* (saillance perceptive) sur l'orientation de l'attention visuelle durant l'exploration d'une scène visuelle à 24 mois et chez les adultes. Les résultats ont montré que les deux groupes regardent davantage les objets incohérents avec la scène et les objets de saillance élevée comparés aux objets cohérents et de saillance faible. Toutefois, seuls les enfants étaient attirés plus rapidement par les objets très saillants. Alors que les objets incohérents et très saillants attirés de façon égale le regard des enfants avec un niveau de vocabulaire en production moyen à fort et moyen à faible, les objets cohérents attirés significativement plus les enfants avec des compétences lexicales élevées.

Dans l'étude IV, la méthode des PEs a été utilisée pour comparer l'effet du contexte visuel sur le traitement des mots chez les enfants âgés de 24 mois à qui des paires mot/scène cohérentes ou incohérentes ont été présentées. De plus, l'influence des compétences langagières sur le traitement des mots en interaction avec une scène visuelle a été analysée. Les résultats ont révélé que l'amplitude de la composante N400 était plus prononcée pour les paires mot-scène cohérentes que pour celles incohérentes. Les enfants avec un niveau de vocabulaire en production moyen à faible présentaient une N400 d'apparence tardive vers les régions frontales droites, tandis que chez les enfants ayant un niveau lexical plus élevé, une N400 d'apparence plus précoce a été observée vers les régions frontales gauches. Ces résultats indiquent que les enfants âgés de 2 ans intègrent les informations tirées du contexte visuel au traitement des mots subséquent mais que les compétences langagières influencent la latence et la distribution topographique de la N400 pendant une tâche d'amorçage contextuel.

## *Abstract*

There is extensive evidence on the developmental mechanisms underlying oculomotor functions in stimulus-driven tasks. However, less is known about developing eye-movement behaviour during free exploration of scenes. The present dissertation investigates developmental mechanisms underlying scene exploration using eye-tracking and event-related potential (ERP) techniques. The two main aims were to investigate the development of visual attention strategies from early infancy to late childhood and to examine the interaction between bottom-up and top-down control during scene viewing in young children.

Studies I and II employed eye-tracking analyses to investigate development of oculomotor functions (fixation durations and saccade amplitudes) and visual attention strategies (ambient *versus* focal) during scene viewing from 3 months to 10 years of age. Results showed that during scene exploration, the fixation durations decreased from the age of 12 months up to 6 to 8 years, while saccade amplitudes reached adult values at 4 to 6 years of age. Previous studies have shown that fixation durations tend to increase, while saccade amplitudes decrease over time. This time course of fixation durations and saccade amplitudes has been associated to ambient and focal strategies during scene exploration. We found a similar time course for fixation durations at 12 months of age but not in younger infants, suggesting that ambient and focal modes emerge by this age. The time course associated to ambient and focal mode was found both the fixation and the saccade parameters from 2 years of age. However, scene exploration was dominated by the focal mode up to 6 years of age. Likewise, visual saliency guided eye movement behaviour more in children younger than 6 years compared with older children and adults.

In Study III eye movements were recorded to examine the influence of top-down (semantic consistency and linguistic) and bottom-up (perceptual saliency) guidance on visual attention during scene viewing in 24-month-old children and adults. Results showed that both 2-year-old children and adults looked longer to semantically inconsistent and high-salient objects than semantically consistent and low-salient objects. However, only children were attracted faster to high-salient objects. Even though semantically inconsistent and high-salient objects attracted equally the gaze of normal-to-high and normal-to-low producers, the consistent objects attracted significantly more the attention of toddlers with higher than lower vocabulary skills.

In Study IV, ERPs were used to compare the effect of visual scene-context on word processing in 24-month-olds when exposed to consistent and inconsistent scene-word pairs. In addition, the influence of language skills in scene-word interactions was analysed. Results revealed that amplitudes of the N400 component were more pronounced for inconsistent than for consistent scene-word pairs. Normal-to-low producers exhibited a later N400 effect over the right frontal recording sites whereas in the group of normal-to-high producers, the N400 effect was observed earlier over the left frontal sites. These findings indicate that 2-year-olds integrate contextual scene knowledge to subsequent word processing but language skills affect the latency and distribution of the N400 during contextual priming.

## *Acknowledgements*

Firstly, I would like to express my sincere gratitude to my supervisors Pia Rämä and Sebastian Pannasch for their continuous support during my PhD work. I would particularly like to thank Pia for her patience, motivation, and knowledge as well as for her guidance and constant encouragement. My sincere thanks also goes to Sebastian for his insightful comments and suggestions, helping me improve my research from various perspectives.

Besides my supervisors, I would like to thank my committee members, Olivier Le Meur, Olivier Pascalis, Karine Doré-Mazars and Melissa Le-Ho Võ for accepting to be part of my Jury. Thank you for your brilliant comments and suggestions.

I would also like to thank the members of the Speech team for providing me an opportunity to join a group of excellent researchers and great people and for giving me access to the laboratory and research facilities. Without their precious support it would not have been possible to conduct this research. I would also like to thank the vision team, especially Patrick Cavanagh and Thérèse Collins for allowing me to use the eye-tracking system that made possible this work. I am also thankful for Najla and Louah with whom I collaborated and who helped me with patience in the recruitment and testing process as well as with my French language impairments. I would like to thank David Meary and the Laboratoire de Psychologie et NeuroCognition of Grenoble for providing the infant data of Study1.

A huge thank to the girls of the 608 office for creating a warm and nice atmosphere and for their sense of humour that made days enjoyable, for the stimulating discussions, for every moment of "relaxation" despite the workload and for all the support in difficult times, particularly to Mélanie and Elena for their help during the writing period.

A special thanks to the *Departamento de Fonoaudiología de la Universidad de Chile* for all their support during these years, in particular to Zulema de Barbieri, Ximena Hormazábal and Carmen Julia Coloma who were always comprehensive and encouraging. I would also like to thank Pedro Maldonado who was my tutor during my master studies, whose teachings I have never forgotten.

This thesis was generously funded by the "*CONICYT Chile*". I am grateful for their financial support.

Last but not least, I am thankful to my family, especially my mother and brothers. I would also like to thank all my friends who supported me in writing and striving towards my goal. I am more than thankful for my little boy Vicente who offered me part of his own "mom-time" specially during the writing process and who also gave me a thousand of smiles that helped me to continue with joy. Fernando, thank you for being by my side during this process, for all your patience and support. Thank you for your always bright ideas and all your help through hard times.

## *TABLE OF CONTENTS*

<b>General Introduction .....</b>	<b>6</b>
<b>Theoretical Background .....</b>	<b>10</b>
Chapter I: Natural scene processing .....	11
Chapter II: Scene context effect in object processing: Semantic scene-object inconsistency .....	14
Chapter III: Eye movement behaviour during natural scene viewing .....	17
Chapter IV: Development of scene exploration from infancy to childhood.....	26
The aims of the thesis .....	34
<b>Experimental Work .....</b>	<b>36</b>
Study I: In press in Visual Neuroscience.....	37
Study II: Published in Vision Research .....	56
Study III: In preparation .....	86
Study IV: In preparation .....	115
<b>General Discussion .....</b>	<b>137</b>
Chapter I: Development of eye movement patterns during scene exploration from infancy to childhood .....	138
Chapter II: Interactions between cognitive and perceptual influences on gaze allocation during scene viewing .....	143
Chapter III: Effect of linguistic skills on scene processing and eye movement guidance in young children .....	146
Chapter IV: Conclusions and future perspectives.....	148
<b>References .....</b>	<b>150</b>

## *General Introduction*

Our visual world offers us an environment rich with semantic information and spatial configurations conforming typical sceneries. Frequently, the concepts of natural or real-world scenes are used for referring visual stimuli that represent this environment. As we gain experience in perceiving visual scenes, scene knowledge is built and stored in our memory. During natural scene viewing, observers activate the stored scene knowledge that allows for a fast identification of the scene, even in less than 100 ms (Henderson & Hollingworth, 1999; Oliva, 2005; Potter, 1975, 1976). Since the visual information of high quality is acquired from a small portion of the visual field at the center of the gaze only, to explore our visual environment in detail we need to move our eyes to relevant locations within the scene (e.g., Henderson, 2003; Henderson and Ferreira, 2004; Mills, 2001). Consequently, visual attention and gaze allocation are closely linked, resulting in attention orientation to fixated locations (Henderson, 2007; Tatler & Vincent, 2008). For example, when we identify a scene (e.g., kitchen) we generate expectations about the configuration of this particular scene and of the objects that would be presented in the scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport, 2007; Davenport & Potter, 2004; Friedman, 1979; Heise & Ansorge, 2014; Palmer, 1975). Thus, when a scene contains an object that is inconsistent with its context (e.g., a fire hydrant in the kitchen) our attention is directed to this object influencing the distribution of the gaze within the scene.

It is well known that visual attention is influenced by the interaction between two attentional mechanisms, namely, bottom-up and top-down (Itti & Koch, 2000, 2001; Parkhurst, Law, & Niebur, 2002; Torralba, Oliva, Castelhana, & Henderson, 2006). In order to guide the eye movement behavior during scene exploration, both strategies coordinate: In the first stage of viewing exploration, the visual attention is automatically grasped by the salient information of the scene whereas in the latter endogenous control is activated (Castelhana, Mack, & Henderson, 2009; Mills, Hollingworth, & Dodd, 2011; Tatler & Vincent, 2008). Several studies have also shown 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). Several cognitive processing such as semantic knowledge, working memory and behavioural task demands, influence gaze allocation during scene exploration (Castelhana, Mack, & Henderson, 2009; Fischer et al., 2013; Mills, Hollingworth, & Dodd, 2011; Tatler & Vincent, 2008). Lately, language processing has also found to influence attention allocation during natural scene viewing (reviewed in Ferreira & Tanenhaus, 2007; Henderson & Ferreira, 2004).



Not only bottom-up and top-down strategies affect eye movement behavior during scene viewing but also there are systematic tendencies that are common across all types of scenes and behavior manipulations (Tatler, 2007, 2009, Tatler & Vincent, 2008, 2009). For instance, the eyes move more frequently in horizontal and vertical directions than in oblique directions (e.g. Bair & O’Keefe, 1998; Lappe, Pekel & Hoffmann, 1998; Lee, Badler & Badler, 2002; Moeller, Kayser, Knecht & König, 2004) and observers have a tendency to fixate central regions irrespective of the distribution of image features or tasks demands (Tatler, 2007). Recently, specific eye movement patterns in the time course of scene viewing have been associated with ambient (“where”) and the focal (“what”) modes of visual processing (Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005). The ambient mode dominates the early scene inspection and is expressed by short fixations and large saccade amplitudes. It has been associated with the orientation of the observers in the scene and the localization of objects. On the contrary, the focal mode dominates the later stages of scene exploration, engages increased attention and is reflected by longer fixations embedded in short saccades. It has been associated with the identification of object details.

To date, little is known about the developmental course of eye movement behavior and attentional strategies during natural scenes exploration. In particular, a systematic characterization of the eye movement behavior over different stages of development is lacking. Only a few studies have used scene viewing to examine the interaction of bottom-up and top-down strategies in attention guidance in young populations (Açik, Sarwary, Schultze- Kraft, Onat, & König, 2010; Duh & Wang, 2014). While language and vision are known to interact and influence each other (Ferreira & Tanenhaus, 2007; Henderson & Ferreira, 2004), and vocabulary skills improve extensively over the second year of life (reviewed in Ganger & Brent, 2004; Nazzi & Bertoncini, 2003), the influence of language development on visual processing, particularly on visual attention guidance, remains unclear.

The objective of the current thesis was, first, to investigate eye-movement patterns during scene perception in typically developing children in large range of age, covering the developmental period from 3 months to 10 years (Studies I and II). Second, to determine the developmental course of bottom-up saliency influence on eye guidance during scene viewing (Study I). Finally, to investigate the interaction between both the bottom-up and the top-down strategies during scene viewing in young children (Study III). More specifically, the following questions were addressed: How do semantic

consistency and saliency interact in the guidance of visual attention and how do linguistic skills contribute to semantic scene processing during scene exploration and during a scene-word priming task in 24-month-olds (Studies III and IV).

In the following section, principal theories and empirical findings related to the main aims of this thesis are summarized. Chapter I provides a general overview about natural scenes and mechanisms underlying scene processing. In the Chapters II and III, I review previous evidence in adults addressing the specific topics of this thesis: (1) scene context effect on semantic processing of natural scenes and (2) eye movement behavior during natural scene viewing. Chapter IV provides evidence about developmental findings of eye movement behavior during scene viewing and semantic scene processing.

## ***Theoretical Background***

## ***Chapter I: Natural scene processing***

A natural scene can be understood as a “*semantically coherent (and often nameable) human-scaled view of real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner. Background elements are taken to be larger-scale, immovable surfaces and structures, such as ground, walls, floors, and mountains, whereas objects are taken to be smaller-scale, discrete entities that move (e.g., animals), or can be moved (e.g., furniture) within the scene. The concept of real-world or natural scenes indicates that they are typically found during the natural course of everyday activity* (Henderson & Hollingworth, 1999a, p. 244).

Natural scenes have particular properties that differ from other visual stimuli: for instance, they fulfil the whole visual field and follow specific semantic and spatial structures that are constrained by real world configurations. Five types of relations that are present in natural scenes have been described: support (most objects rest on surfaces), interposition (objects interrupt their background), probability (the likelihood that certain objects will be present in a scene), position (the typical positions of objects within scenes) and familiar size (the adequate size relation among objects) (Biederman, 1981; Biederman, Mezzanotte, & Rabinowitz, 1982). Through lifetime experience, observers built mental representations of their visual world (Barlett, 1932; Friedman, 1979; Hock, Romanski, Galie, & Williams, 1978). These mental representations, known as *scene knowledge*, include information about expected objects in a specific scene type (e.g., sofa in a living room), spatial regularities (e.g., chairs do not float in space), and generic world knowledge (e.g., trees do not float in the air) (Henderson, 2003; Henderson & Ferreira, 2004).

When viewing a visual scene, observers use their scene knowledge to extract the identity—the so-called scene gist—of a scene (e.g., a kitchen) (Oliva, 2005; Potter, 1975, 1976). It has been proposed that besides the basic-level category or identity of the scene, the scene gist includes information about semantic features (objects that belong to the scene) and some aspects of the spatial configuration (e.g., arrangement of objects in the space) of the scene (Greene & Oliva, 2009; Henderson & Ferreira, 2004; Oliva, 2005; Schyns & Oliva, 1994). It has been proposed that scene gist is retrieved by contrasting visual input to stored semantic knowledge (Oliva, 2005; Wu, Wick, & Pomplun, 2014). This information allows the observer to understand the visual context and guide eye movement behaviour during the early stages of scene viewing (Castelhano & Henderson, 2007; Oliva, 2005; Tatler, 2009; Torralba, Oliva, Castelhano, & Henderson, 2006; Wu et al.,

2014). Behavioural studies have shown that observers need less than 100 ms to recognize the identity of a scene (Potter, 1975, 1976) as well as its spatial layout (Schyns and Oliva, 1994). Both perceptual and conceptual information retrieved from a short preview of a scene can modulate further eye movement behaviour and scene understanding. Therefore, it has been proposed that the gist of the scene can be explained at both perceptual and conceptual levels (Oliva, 2005). The perceptual level refers to the structural representation of a scene built during perception and includes low level of visual information such as global features (e.g., colour, contours) and intermediate information (e.g., shapes, texture regions) (Greene & Oliva, 2009; Oliva, 2005; Schyns & Oliva, 1994). The conceptual level refers to the semantic information inferred from scene based on high-level information (e.g., activation of semantic knowledge) during viewing time or shortly after the scene has disappeared from view (Oliva, 2005). Some studies have suggested that scene gist can be retrieved based on low-level image features (e.g., spatial layout, texture, volume) or low-level image frequencies (e.g., after a low pass filter) without object recognition (Schyns and Oliva, 1994; Oliva and Torralba, 2001; Torralba et al., 2006) while others have shown that scene gist is processed more accurately when representative objects in the scene are recognized (Davenport, 2007; Davenport & Potter, 2004; Friedman, 1979; Hollingworth & Henderson, 1999).

The scene gist contains the first semantic information about the scene meaning. However, in order to achieve a complete understanding of a scene more detailed information is needed. In this regard, it has been proposed that visual attention during scene exploration is guided by the semantic information contained in the scene. This semantic information, which allows observers to understand their visual environment, is contained in different semantic relations presented in the scene. These relations have been summarized by Wu and co-workers (2014): scene- object relations, spatial associations between objects, and semantic similarity between objects. Regarding object-scene relations, it has been proposed that semantic coherence guides attention to relevant objects for scene understanding or to inconsistent objects that violate expectations generated by the *gist* (Biederman, 1981; Biederman et al., 1982; Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978). Semantic information is also contained in the co-occurrence of objects (i.e. the probability of two or more objects to appear in the same scene type) and their local spatial layout in the scene (spatially close objects) (Bar, 2004; Mack and Eckstein, 2011). Object co-occurrence provides information about the probability of a spatially close object appearing in the scene. For instance, if a scene contains a table, it is

expected to find a chair next to it (Belke, Humphreys, Watson, Meyer, & Telling, 2008; Hwang, Wang, & Pomplun, 2011; Moores, Laiti, & Chelazzi, 2003). This effect, labelled as semantic guidance, has been found in search tasks even when two semantically related objects were shown on an inconsistent background (Castelhano & Heaven, 2011). Besides the object-object relation that includes spatial proximity, the conceptual semantic similarity among all the objects in the scene also influences attention and eye movements in real-world scenes (Hwang et al., 2011).

## ***Chapter II: Scene context effect in object processing: Semantic scene-object inconsistency***

### ***2.1. Semantic scene context and object processing: Behavioural evidence***

Several lines of evidence indicate that scene context influences the processing of objects contained in a scene (Bar, 2004; Biederman et al., 1982; Davenport, 2007; Davenport & Potter, 2004; Heise & Ansorge, 2014). Semantic object–scene consistency has been found to facilitate the detection and discrimination of target objects (Hock et al., 1974, Biederman et al., 1982, Oliva & Torralba 2007). For example, when scenes are presented briefly, objects appearing in consistent contexts are discriminated more accurately and faster than objects appearing in inconsistent background (Biederman et al., 1982; Davenport & Potter, 2004; Davenport, 2007; Heise & Ansorge, 2014). A short preview of a scene also increases search and memorization efficiency of objects that are consistent with the scene context (Hillstrom, Scholey, Liversedge, & Benson, 2012; Josephs, Draschkow, Wolfe, & Võ, 2016; Palmer, 1975; Võ & Henderson, 2010). Likewise, after a brief preview of a scene, targets objects are discriminated more accurately and detected faster when appeared in a natural scene than in a jumbled version of the same scene (Biederman, 1972; Biederman, Glass, & Stacy, 1973; Biederman, Rabinowitz, J, Glass, A, & Stacy, 1974). Moreover, increasing the time dedicated to integrating gist and object identity (i.e. increasing the time after the target object label that follow the scene preview and the start of the search task) improved search efficiency, even under conditions where scene information was degraded (Võ & Henderson, 2010). Based on this evidence, it has been proposed that semantic consistency effect reflect a high requirement of attentional resources either for the identification of the object in the scene or to solve the conflict given by the gist violation (Loftus & Mackworth, 1978; Davenport, 2007). It has been also proposed that object recognition might facilitate the recognition of scene background (Davenport, 2007; Davenport & Potter, 2004) indicating a bi-directional influence.

### ***2.2. Semantic scene context and object processing: Event-related potential evidence***

The event-related potential (ERP) is a technique widely applied to investigate neural and cognitive mechanisms underlying cognitive processing in different modalities (review in, Kutas & Federmeier, 2011) both in adult and developing populations. Lately, this technique has also been used to investigate object facilitation in scene context (Ganis & Kutas,

2003; Mudrik, Lamy, & Deouell, 2010; Võ & Wolfe, 2013). The ERPs are electrophysiological measures resulting of averaging the electrical brain activity recorded at the scalp using an electroencephalogram (EEG), and time-locked to an event of interest (reviewed in, Kutas & Federmeier, 2011). This averaged measure reflects a temporally accurate time series of changes in the brain activity in response to sensory or cognitive events. Different ERP components have been reported to be associated with different cognitive process. These components are usually defined by the polarity of their waveform and the time of their occurrence after the stimulus presentation. For instance, the P100 is a positive response elicited around 100 ms after stimulus onset while the N400 is a negative response around 400 ms associated with semantic integration both in visual and auditory modalities (review in, Kutas & Federmeier, 2011). In the following chapter, previous research related to the occurrence and significance of the N400 component will be described more in details.

### ***2.2.1. N400 component***

The N400 component is a negative response associated with semantic processing often appearing around 200 to 600 ms after a stimulus onset (Kutas & Federmeier, 2011). Kutas and Hillyard (1980) provided the first evidence of a negative ERP component around 400 ms in response to semantically inconsistent sentence endings (e.g., take coffee with cream and *dog*) compared with consistent sentences (e.g., take coffee with cream and *sugar*). After this first description, the N400 response to semantic processing has been found in a large amount of studies. The N400 component is typically defined by calculating the point-by-point difference between two conditions (e.g., subtraction of the ERP response of a consistent from the response to an inconsistent condition). The amplitude and distribution of the N400 response is modulated by semantic inconsistencies, semantic complexity and semantic errors (reviewed in Kutas & Federmeier 2000; Kutas & Van Petten 1994). The N400 component is generalized across input modalities and it can be elicited by a wide range of stimulus types: words, sentences, pictures (e.g. objects, faces, natural scenes), environmental sounds, and even odors (reviewed in Kutas & Federmeier, 2011). However, the distribution and latency of this component can vary between modalities. For instance, the N400 effects for linguistic stimuli are often evoked over the centro-parietal recording sites, with a slight lateralization to the right hemisphere while pictures elicit a more frontally distribution (Ganis et al.1996), suggesting that the N400 component is modality sensitive but not modality specific.



### ***2.2.2. N400 and semantic inconsistencies during scene perception***

As mentioned above, the N400 response to scene-object inconsistencies have been reported in some studies in adults. In two of these studies, participants saw a short preview of visual scenes (e.g., soccer players in a field) following by a presentation of either semantically consistent (soccer ball) or inconsistent (toilet paper) object at a predetermined location within the scene while the brain response was measured (Ganis & Kutas, 2003; Võ & Wolfe, 2013). Both of these studies found the N400 effect over the central recording sites. In another study (Mudrik, Lamy, & Deoell, 2010), participants were presented with visual displays of consistent and inconsistent scene-object without a preview of scenes (e.g., a man playing a violin vs. a man “playing” a broomstick). In this study, the N400 effect was found over the frontal and central recording sites. This finding provides additional evidence that scene context influences object processing very quickly even when inconsistent objects were displayed simultaneously with the scene (Mudrik, Lamy, & Deoell, 2010). In comparison with other studies using the visual modality where the distribution of the N400 has been more pronounced over the frontal areas (for review, Kutas & Federmeier, 2011), the results of studies using visual scenes showed that the N400 effect was more centrally than frontal distributed. These late findings show that ERP measure is a useful technique to study semantic processing during scene viewing providing a new tool to understand the underlying mechanisms associated to this processing.

### ***Chapter III: Eye movement behaviour during natural scene viewing***

Given the optical and anatomical structure of the eye and retina, high-resolution visual information is acquired from a small portion of the visual field corresponding to about 5 visual degrees surrounding the centre of the gaze. The quality of this visual representation declines from the centre of the gaze to the periphery. In addition, the visual system amplifies the representation of central visual information in the visual cortex and more cortical resources participate in the processing of fixated regions (Yu, Chaplin, & Rosa, 2015). Therefore, even the gist of a scene can be extracted within a single fixation, eye-movements are needed to extract specific information such as local visual details, or object identities (Henderson & Hollingworth, 1999a; Hollingworth, Williams, & Henderson, 2001; Wu et al., 2014). During visual exploration, the gaze is re-oriented within a scene approximately three times per second via fast saccadic eye movements up to 900 deg/s (Carpenter, 1988). Visual information is extracted during fixations, periods when the gaze is relatively static (Matin, 1974; Volkman, 1986). Correspondingly, several studies revealed that identifying an object requires a direct or close by fixation (Henderson & Hollingworth, 1999b; Hollingworth, Schrock, & Henderson, 2001; Hollingworth, Williams, et al., 2001).

#### ***3.1. Visual attention guidance during scene viewing***

Usually, our attention is directed to a location that is currently fixated but also to the locations that will be fixated next (Henderson, 2007; Tatler & Vincent, 2008). Thus, online decision of eye movements and gaze locations (i.e. gaze control) are critical for supporting ongoing cognitive and behavioural activity (Henderson, 2007; Henderson & Ferreira, 2004; Tatler & Vincent, 2008). Moreover, fixation durations reflect online visual processing during scene viewing (Henderson & Pierce, 2008). It is a widely held view that gaze distribution under natural viewing conditions is influenced by interactions between bottom-up, stimulus-driven (e.g., saliency) (Henderson, 2003; Itti & Koch, 2000, 2001; Parkhurst, Law, & Niebur, 2002; Tatler, Hayhoe, Land, & Ballard, 2011) and top-down, cognitive controlled (e.g., semantic schema knowledge, working memory and behavioural task demands) (Henderson, 2003; Henderson & Hollingworth, 1999a; Spotorno, Tatler, & Faure, 2013; Tatler, 2009; Torralba et al., 2006) factors. However, the mechanisms underlying online eye-movement control during scene viewing are not yet fully understood and the weighting of perceptual and cognitive factors in gaze allocation is still under debate.

### *3.1.1. Bottom-up guidance in scene perception*

There is extensive evidence demonstrating that low-level properties (e.g., colour, contrast, edges) capture and guide visual attention during scene exploration (Henderson, 2003; Itti & Koch, 2000, 2001; Le Meur, Le Callet, & Barba, 2007; Le Meur & Liu, 2015; Parkhurst et al., 2002; Tatler et al., 2011). For example, it has been shown that regions with greater edge density and local contrast attract fixations more than at other locations (Kayser, Nielsen, & Logothetis, 2006; Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Reinagel & Zador, 1999). Particularly, saliency (i.e. the difference between low properties of a visual stimulus compared to the near visual input) is a determining factor in gaze allocation (Itti & Koch, 2000, 2001; Koch & Ullman, 1985; Treue, 2003; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006).

The most influential saliency model to examine fixation distribution during scene exploration derives from Koch & Ullman (1985) proposal (Itti & Koch, 2000). This model generates saliency maps based on the properties of primary visual cortex using different visual dimensions of images such as colour, intensity, contrast, and edge orientation over multiple spatial scales (Itti & Koch, 2000; Koch & Ullman, 1985; Le Meur, Le Callet, Barba, & Thoreau, 2006; Torralba et al., 2006). One map is generated for each dimension and then combined to create a single saliency map including several salient points in the image. These salient points are used as predictors of the spatial distribution of fixations (Itti & Koch, 2000, 2001). In fact, experimental evidence using natural scenes has shown that salient regions are more fixated than control locations or locations expected by chance (Foulsham & Underwood, 2008; Parkhurst et al., 2002).

Previous studies have shown that the influence of bottom-up saliency on fixation distribution is more significant during the early than late stages of viewing time (Mannan, Ruddock, & Wooding, 1995; Parkhurst et al., 2002). Moreover, it has been suggested that initial fixation placements might be controlled only by low-level features (Mannan et al., 1995). In this study, viewers examined different type of filtered (unfiltered, high-pass and low-pass filtered) grey-scale images of natural scenes while their eye movements were tracked. Fixation positions were found to be similar on the unfiltered and low-pass filtered scenes during the early stage of viewing (first 1.5 s) even when viewers were unable to extract semantic information of filtered scenes, suggesting that perceptual information was enough to guide early eye movements.

### *3.1.2. Top-down guidance in scene viewing*

Buswell (1935) and Yarbus (1967) provided the first evidence of top-down control during scene perception. Buswell (1935) showed that when examining paintings, viewers' fixations tended to land on informative areas of the scenes (e.g., people) rather than on background regions. Later, Yarbus (1967) observed qualitative differences in fixation distribution depending on both visual information and the viewing task. For example, when participants had to estimate the ages of family members in a painting, they tended to fixate on faces but when they had to estimate the material circumstances of the family, they tended to land on objects. Since these early findings, several studies have shown that the gaze is directed to regions that are relevant either for scene understanding or for achieving the task goals (Castelhano, Mack, & Henderson, 2009; De Graef, Christiaens, & D'Ydewalle, 1990; Einhäuser, Rutishauser, & Koch, 2008; Henderson et al., 1999; Loftus & Mackworth, 1978; M. Mills, Hollingworth, & Dodd, 2011; Vö & Henderson, 2009, 2011). For example, relevant visual information such as human figures, faces or animals has a stronger effect on gaze allocation than neutral images (Buswell, 1935; Crouzet, Kirchner, & Thorpe, 2010; Kirchner & Thorpe, 2006) and gaze distribution has found to be different for search and memory tasks (Castelhano et al., 2009; Einhäuser et al., 2008; M. Mills et al., 2011). Likewise, during complex and well-learned tasks such as making a tea (Land, Mennie, & Rusted, 1999) or a sandwich (Land & Hayhoe, 2001), participants tend to fixate task-relevant objects and areas related to future actions associated with the task. Based on these findings, it has been proposed that online decisions about gaze distribution depend on the interaction of scene and task related knowledge (Henderson & Ferreira, 2004). Scene knowledge would facilitate gaze allocation to informative regions of the scenes while task-related knowledge would allow the observers to adjust their eye-movements to the requirements of the task (reviewed in, Henderson & Ferreira, 2004). Lately, a third relevant component, language, has been associated to visual attention guidance during scene exploration (Ferreira & Tanenhaus, 2007; Henderson & Ferreira, 2004; Meyer, Belke, Telling, & Humphreys, 2007; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Previous studies have shown that gaze allocation and perceptual processing are influenced by linguistic processing (Meteyard, Bahrami, & Vigliocco, 2007; Meyer et al., 2007; Spivey et al., 2002; Tanenhaus et al., 1995).

### *3.1.2.1. Semantic guidance of visual attention*

As discussed above, based on top-down mechanisms, visual attention is directed to informative regions of a scene (De Graef et al., 1990; Henderson, 2007; Henderson et al., 1999; Loftus & Mackworth, 1978; Vö & Henderson, 2009, 2011). Informativeness has been defined as a relative contribution of a particular region to the total information provided by the scene (Antes, 1974), and it has been estimated using two different approaches: (1) the agreement between participants about the contribution of a particular region of the scene to the global scene understanding (Antes, 1974) and (2) the degree to which a given object is predictable within the scene, with unpredictable objects taken for more informative (De Graef et al., 1990; Loftus & Mackworth, 1978). Studies manipulating these two parameters have shown that informative regions attract more fixations compared with the rest of the scene (De Graef et al., 1990; Henderson, 2007; Henderson et al., 1999; Loftus & Mackworth, 1978; Vö & Henderson, 2009, 2011).

Loftus and Mackworth (1978) provided the first evidence that semantic inconsistency attracted the gaze of observers (semantic consistency effect). In this study, participants were presented with line drawings of natural scenes containing semantically consistent (e.g., a tractor in a farmyard) or inconsistent (e.g., an octopus in a farmyard) objects in the context of a memory task. Results showed that inconsistent objects were fixated earlier and for a longer time than consistent objects. The authors proposed that these results might be explained by memorization strategies used for the participants to distinguish informative and no informative regions in the scene. After this first report many studies have shown that more fixations are directed to inconsistent than consistent objects (e.g., Henderson, 2007; Tatler & Vincent, 2008; Vo, 2009, De Graef 1900; Henderson 1999; Friedman 1979). Furthermore, many studies have shown the first pass gaze duration (the sum of all fixations from first entry to exit in a region) is also longer for inconsistent than consistent objects (Loftus & Mackworth and Friedman, 1970, De Graef, 1990, Henderson, 1999, Vo, 2009).

While it is widely accepted that semantic inconsistencies attract visual attention during scene viewing, it is still unclear whether they influence the early (extrafoveal) eye movement control, that is, prior to foveal processing. Some studies argue that semantic inconsistencies are detected rapidly, i.e. within the first 200 ms, influencing early eye movements (Becker, Pashler, & Lubin, 2007; Loftus & Mackworth, 1978; Underwood, Humphreys, & Cross, 2007; Geoffrey Underwood & Foulsham, 2006; Geoffrey Underwood, Templeman, Lamming, & Foulsham, 2008). Alternatively, it has been reported that a foveal process is

needed in order to detect the semantic inconsistencies (e.g., De Graef, Christiaens, & d'Ydewalle, 1990; Gareze & Findlay, 2007; Henderson, Weeks, & Hollingworth, 1999; Rayner, Castelhana, & Yang, 2009). Studies using line drawings in different searching (De Graef et al., 1990) and memory (Henderson & Hollingworth, 1999b) tasks have failed to find an extrafoveal semantic consistency effect. Furthermore, Võ and Henderson (2009) presented participants with real-world scenes where low-level features were controlled. Once more, no early effect of scene inconsistencies either during scene memorization or visual search tasks was found. Based on these findings, it has been suggested that extrafoveal effect found in previous studies might related to low-level features of stimuli rather than semantic inconsistency detection (Henderson & Hollingworth, 1999b; Võ & Henderson, 2009).

### *3.1.2.2. Linguistic guidance of visual attention*

In everyday life, linguistic inputs usually occur within a visual context, and information from both modalities is integrated very quickly (Eberhard, Spivey-Kowlton, Sedivy, & Tanenhaus, 1995; Ferreira & Tanenhaus, 2007; Henderson & Ferreira, 2004; Spivey, Tyler, Eberhard, & Tanenhaus, 2001; Tanenhaus et al., 1995; Tanenhaus & Trueswell, 2006). For instance, language guides visual attention and visual perceptual processing. One of the first finding showing language effect on gaze allocation was provided by Cooper (1974). In his study, participants heard spoken stories while looking at objects illustrated in a grid while their eye movements were tracked. The results showed that participants directed their gaze toward objects named in the stories and towards semantically related pictures. Also, in tasks where participants followed spoken instructions to manipulate real objects (Tanenhaus & Trueswell, 2006) or objects presented on a screen (Allopenna, Magnuson, & Tanenhaus, 1998) their eye movements were closely time-locked directed to the referred objects. Recently, language-driven eye-movements have been shown in the context of natural scene exploration (Andersson, Ferreira, & Henderson, 2011). In this study, the complexity of the linguistic stimuli (high and low-speed of a spoken sentence) and the visual scene (high and low density of object) was manipulated. Results showed that linguistic complexity had an effect in the probability of fixation of a target. The probability of fixating an object increased when it was mentioned in a sentence, even when this linguistic stimulus and the scene were complex. However, objects mentioned in less complex low-speed condition were more likely to be fixated and were fixated earlier than those mentioned in the high-speed condition. It has been also shown that linguistic inputs can affect perceptual visual processing (Meteyard et al., 2007; Spivey et al., 2001). For example, in a study where participants heard verbs describing upward and downward motion while performed a motion-detection task, incongruent auditory-visual pairs interfered their perceptual sensitivity (Meteyard et al., 2007). Altogether, these finding indicate that language processing directly influence gaze allocation and visual processing

Language processing is also activated automatically during visual exploration even when no linguistic information is provided (Matlock & Richardson, 2004; Richardson & Matlock, 2007, Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus et al., 1995, Görge et al., 2013; Meyer et al., 2007; Chabal & Marian, 2015) and implicit naming can influence gaze allocation (Matlock & Richardson, 2004; Richardson & Matlock, 2007, Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus et al., 1995, Görge et al., 2013;

Meyer et al., 2007; Chabal & Marian, 2015). Additionally, when naming tasks are performed in the context of natural scene viewing, these two processes influence each other. For instance, the probability of an object to be fixated is affected by linguistic parameters such as semantic proximity (i.e. similarity between words based on the co-occurrence on similar context) and word frequency (Clarke, Coco, & Keller, 2013; Coco, Malcolm, & Keller, 2014). Interestingly, in a study where the exploration of consistent and inconsistent scenes was combined with a simultaneous naming task, consistent objects were more fixated than inconsistent objects in contrast to the classical findings found in purely visual tasks (Coco, Malcolm, & Keller, 2014). This finding suggests that naming processing can modify attention allocation during semantic processing of scenes.

### ***3.1.3. Top-down and bottom up interaction during scene viewing***

While it has been commonly agreed that both low- and high-level interactions play an important role in gaze guidance during natural scene viewing, there is no agreement about the weighting of these components in the fixation distribution (Tatler et al., 2011; Theeuwes, 2010). Two different theoretical positions have been established: one line of evidence claims a stronger effect of saliency guidance on gaze allocation (Itti & Koch, 2001; Zelinsky, 2008) while the other argues that there is a stronger effect of semantic-guidance (Hwang et al., 2011; Torralba et al., 2006).

Empirical evidence suggests that perceptual and cognitive interactions during gaze guidance depend on the time course of viewing time, visual features of the stimuli, and characteristic of the task (Einhäuser, Rutishauser, & Koch, 2008; Hayhoe & Ballard, 2005; Parkhurst, Law, & Niebur, 2002). Saliency guidance dominates during the early stages of scene inspection while later in time top-down control becomes more significant (Castelhano et al., 2009; Mannan et al., 1995; Parkhurst et al., 2002; Tatler & Vincent, 2008). In addition, studies using saliency maps have shown that the semantically informative stimuli decrease the influence of perceptual guidance of gaze (Henderson et al., 2007; Nyström & Holmqvist, 2008; Parkhurst et al., 2002). Parkhurst and colleagues (2002) showed that even if saliency map predictions correlated with real fixations, these correlations decreased when meaningful stimuli were presented. Likewise, Nyström and Holmqvist's (2008) reported a saliency effect for images with a relatively neutral semantic content (e.g., nature) whereas in images including semantically relevant visual information (e.g., faces or man-made elements) these regions were prioritized irrespectively of the saliency. With regard to task demands



(Henderson et al., 2007; Underwood & Foulsham, 2006; Underwood et al., 2006), it has been shown that salience models predict better gaze allocation in free viewing (Parkhurst et al., 2002) or memorization tasks (Underwood & Foulsham, 2006) than in tasks where visual attention is cued to a specific target such as searching tasks (Einhäuser et al., 2008; Underwood & Foulsham, 2006; Underwood et al., 2006). Altogether, these findings indicate that the strength of bottom-up saliency effect on gaze control is modulated by cognitive control (Parkhurst et al., 2002; Treue, 2003).

### ***3.2. Scanning strategies during scene viewing: Ambient and focal modes***

Recent evidence has shown systematic tendencies that are common across all type of scenes and behaviour manipulations (Tatler, 2007, 2009, Tatler & Vincent, 2008, 2009). For instance, saccade amplitudes usually show a positively skewed, long-tailed distribution when complex scenes are viewed (Bahill, Adler, & Stark, 1975; Tatler, Baddeley, & Vincent, 2006; Tatler & Vincent, 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005). Saccade directions are more frequently executed in horizontal and vertical directions than in oblique directions (Le Meur & Liu, 2015; Lee, Badler, & Badler, 2002; Moeller, Kayser, Knecht, & König, 2004) and observers present a central fixation bias irrespective of the distribution of image features or tasks demands (Tatler, 2007). Moreover, it has been shown that a model based on systematic motor biases (blind to current visual information) predicted fixation locations better than chance (Tatler & Vincent, 2009).

Antes (1974) showed a particular time course of eye movement patterns during scene exploration. Fixation durations increased while saccadic amplitudes became shorter over first several seconds of exploring a scenes. Later, the same time course of fixation durations and saccades amplitude has been shown using different type of stimuli and tasks (Follet, Le Meur, & Baccino, 2011; Pannasch, Helmert, Roth, & Walter, 2008; Scinto & Pillalamarri, 1986; Unema et al., 2005). Additionally, a direct relationship between successive fixations and saccades has been described (Velichkovsky, 2002; Unema 2005, Tatler 2008). Several studies have shown that short fixation durations are often followed by long saccade amplitudes while long fixations are followed by short saccades. For instance, a strong no-linear relationship between fixation durations and the following saccade amplitudes was shown by Unema and co-workers (2005). In their study fixations shorter than 180 ms were followed by saccades longer than 5 visual degrees. A similar pattern was shown by Tatler (2008): fixations between 80 and 180 ms were followed by large amplitude

saccades whereas longer fixations were followed by smaller amplitude saccades. Based on these findings, it has been proposed that successive fixations and saccades are not independent events and that both parameters are controlled by common mechanisms (Tatler & Vincent, 2008; Unema et al., 2005).

These systematic eye movement patterns in scene viewing have been associated to two different scanning strategies during scene viewing (Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005; Velichkovsky, Rothert, Dornhoefer, & Joos, 2002). Different authors have called these strategies differently: ambient-focal (Trevarthen, 1968), noticing-examining (Weiskrantz, 1972), spatial-figural (Breitmeyer & Ganz, 1976), and ambient-foveal (Stone, Dreher, & Leventhal, 1979). In this dissertation, ambient-focal nomenclature is used. It has been proposed that during initial scanning of a scene, in the first 2 seconds of viewing time, eye-movements are mainly guided by an explorative behavior, characterized by the dominance of short fixation durations (< 180 ms) and large saccade amplitudes (> 5 visual degrees) (Pannasch et al., 2008; Tatler & Vincent, 2008; Unema et al., 2005; Velichkovsky et al., 2005; Velichkovsky, Rothert, Kopf, Dornhöfer, & Joos, 2002). These eye movement patterns would support the orientation of observers in the visual environment allowing for a fast scanning of large areas (long saccades) of the visual field but in a superficial manner (short fixation durations) (Norman, 2002; Pannasch & Velichkovsky, 2009; Velichkovsky et al., 2005). After this orienting phase, a change in the scanning strategy occurs associated to the exploration of relevant details within the scene. Recognition and identification of details would require longer fixation durations and shorter saccade amplitudes (several fixations within a same region) (Norman, 2002; Pannasch, Helmert, Roth, et al., 2008; Velichkovsky et al., 2005; Velichkovsky, Rothert, Kopf, Dornhöfer, & Joos, 2002). Recent empirical evidence has confirmed this assumption, For instance, in a study using a driving task, ambient fixations were related to detection of hazardous events whereas focal fixation were associated to identification of the these events (Velichkovsky et al., 2002). In another study, observers showed a better performance in recognizing cut-outs as part of a previously explored scene when these patch had been fixated by focal fixations than ambient fixations (Velichkovsky et al., 2005). Likewise, focal fixations have shown stronger distractor effects (i.e. prolongation of visual fixation after a sudden change in the stimuli by the appearance of a visual distractor) than ambient fixations - (Pannasch & Velichkovsky, 2009).

## ***Chapter IV: Development of scene exploration from infancy to childhood***

### ***4.1. Development of oculomotor functions***

During scene viewing gaze has to be distributed over different regions of the visual environment. Consequently, maturation of the motor aspect (e.g. saccade latency, fixation steadiness) and cognitive control (e.g. voluntary inhibition of reflexive saccades, fixation durations) of oculomotor functions are needed to allow an efficient exploration of scenes. In the following sections the developmental course of the principal oculomotor parameters will be reviewed.

#### ***4.1.1. Oculomotor functions during infancy***

It is known that many oculomotor functions or skills needed for visual scene exploration are mature already by the end of the first year of life (Karatekin, 2007; Luna, Velanova, & Geier, 2008). Saccade amplitudes are associated with the distribution of gaze in the visual environment, and thus, a proficient scene exploration requires of certain maturity of this parameter. Saccades are hypometric during infancy, that is, shorter than the optimal location for centering the stimuli into the foveal region and the number of saccades that are needed to fixate a peripheral target decreases during the first year of life (Aslin & Salapatek, 1975; Luna et al., 2008; Regal, Ashmead, & Salapatek, 1983; Roucoux, Culee, & Roucoux, 1983). Similarly, saccade latencies and inhibition of return to a previous target, an ability that has been related to exploration efficiency, have been found to develop from birth until 8 months of age (Gredebäck, Örnkloo, & von Hofsten, 2006). Likewise, eye-movements during smooth pursuit - i.e. the ability to closely match pursuit eye movements with a moving target - can be elicited already in newborns, but they improve rapidly during the first seven months of age (Hainline, Turkel, Abramov, Lemerise, & Harris, 1984; Regal et al., 1983; Rosander & Von Hofsten, 2002; Roucoux et al., 1983; Rüttsche, Baumann, Jiang, & Mojon, 2006; Shea & Aslin, 1990; Von Hofsten & Rosander, 1997). Furthermore, the coordination between head and eye movements, which is needed to shift a gaze from a fixated target to another during visual exploration, emerges at two months of age, but improves during the first year of life. During this period infants switch gradually from a pattern of head-moves-

first followed by eye-movements to an adult-like pattern were eyes-move-first followed by head motion (Bloch & Carchon, 1992; Daniel & Lee, 1990; Nakagawa & Sukigara, 2013; Roucoux et al., 1983).

The ability to maintain a fixation by inhibiting reflexive saccades reflects cognitive control of eye movements and it is needed to maintain the gaze during detailed exploration of scenes. This capability is present already during the first months of life but it improves during the first year. Four-month-old infants are able to inhibit orienting to a peripheral target for short periods of time but their performance is highly dependent on the visual input (Johnson, 1994, 1995). However, by the age of six months, infants can inhibit reflexive saccade more consistently and for longer periods of time (Gilmore & Johnson, 1995). Also, the facility to disengage from a stimulus and to attend to another, which is necessary to distribute the gaze over a scene, increases during the first year of life (Hood & Atkinson, 1993; Matsuzawa & Shimojo, 1997). Likewise, fixation durations during free exploration, associated with arousal and visual processing speed in infants, decrease with increased age during infancy (Bronson, 1991; Colombo, Mitchell, Coldren, & Freeseaman, 1991; Wass & Smith, 2014).

Concomitantly, visual attention is notably developing during the first year of life (for review, Colombo, 2001; Johnson, 2002). Colombo (2001) proposed that three important stages of attention develop during the first year of life: (1) *alertness*, (2) attention to *object features and spatial orienting*, and, (3) *endogenous attention*. The first stage, *alertness*, is present at birth and develops very fast during the first two months of age. During the first month, the phases of alertness are short and usually caused by exogenous stimulation (Colombo & Horowitz, 1987; reviewed in Colombo, 2001) while by the second or third month, alert states become consolidated and infants are able to attain, and maintain more extended periods of alertness. The second state, attention to *object features and spatial orienting*, develops during the second month of life up to six months of age. During this period there is an important development of objects features and spatial orienting attention. The third level of attention, *endogenous attention*, develops during the latter parts of the first year, infant learns to inhibit attentional distractors, alternate between two stimuli, and cognitive requirement such as maintain vigilance (Colombo, 2001).

#### ***4.1.2. Oculomotor functions during childhood***

Although oculomotor functions present an accelerated development during the first year of life, most of these functions continue to develop during childhood with a relatively slow and gradual progression up to the adolescence. In this regard, the development of saccade parameters, such as latency and accuracy, has been shown to mature during childhood and most of these parameters seem to reach adult-like performance near to 10 years of age or even later (reviewed in Luna et al., 2008). For instance, saccade amplitudes are shorter and less accurate in children compared with adults and they tend to stabilize around the age of 10 years (Fioravanti, Inchingolo, Pensiero, & Spanio, 1995; Irving, Tajik-Parvinchi, Lillakas, González, & Steinbach, 2009). Likewise, different studies have shown that cognitive control of saccade execution in pro- and anti-saccades tasks reaches adult-like performance about 10 to 12 years of age (Fukushima, Hatta, & Fukushima, 2000; Irving et al., 2009; Klein & Foerster, 2001) while saccade latencies are decreasing until the age of 14 to 15 years (Fukushima et al., 2000; Irving, Steinbach, Lillakas, Babu, & Hutchings, 2006; Klein & Foerster, 2001; Luna & Seeney, 2004; Munoz, Broughton, Goldring, & Armstrong, 1998). Pursuit accuracy also matures up to adolescence (Karatekin, 2007; Luna et al., 2008). Concerning fixation development, the time of fixation maintenance has been shown to increase and the number of obtrusive saccades to decrease from 4 to 15 years old (Aring, Grönlund, Hellström, & Ygge, 2007; Ygge, Aring, Han, Bolzani, & Hellström, 2005). Likewise, children at the age of 8 years are less proficient than 10 to 12-year-olds in maintaining a fixation on a central target when exposed to a peripheral distractor. However, these age differences disappeared when the maintenance of fixation was required to play a game, suggesting that increasing motivation to achieve the task goal is affecting eye movement control (Paus, 1989).

#### ***4.2. Exploration of natural scenes in infants and children***

Most of the developmental studies concerning eye movement behaviour have used artificial stimuli (e.g., lines or dots) and stimulus-driven tasks (e.g., smooth pursuit or saccade to a peripheral target) while only a few studies have used more naturalistic stimuli or free exploration tasks (Açık, Sarwary, Schultze-Kraft, Onat, & König,

2010; Bronson, 1990, 1991, 1994; Egami et al., 2009; Vurpillot, 1968; Wass & Smith, 2014).

Earlier studies presenting infants with geometric shapes and natural scenes indicate that shorter fixations are associated with better cognitive control and exploration proficiency (Bronson, 1990, 1991, 1994; Wass & Smith, 2014). Infants at 1 to 2 months of age present longer fixations and tend to group them in a small area of the stimuli while by 3 to 4 months of age; fixations are shorter and more scattered (Bronson, 1990, 1991, 1994). Additionally, at the age of 3 months infants with shorter fixations presented a more spread scan pattern (Bronson, 1991). Moreover, previous studies have shown that infants with shorter looking times during visual exploration process visual stimuli more in an adult-like manner, that is, from global-to-local sequence, while infants with longer look durations, used more local visual information to process visual stimuli (Colombo, Freeseaman, Coldren, & Frick, 1995; Colombo, Frick, Ryther, & Gifford, 1996). These earlier findings using single objects or shapes suggest that shorter fixations are related to a more efficient exploration of images in infants. One of the first studies that addressed directly eye movement behaviour during free exploration of natural scenes in infants was conducted by Was and Smith (2014). In this study, 11 month-old infants freely explored naturalistic scenes using static and dynamic settings while their eye movements were recorded. The results showed individual consistency on mean fixation duration measured in the first session and 15 days later. Results also showed that within-participant variation in fixation durations was strongly related to cognitive control (estimated by the proportion of correct anticipatory looks in a visual task) and arousal, (indexed by measuring tonic pupil size). These findings indicate that fixation duration is a reliable measure of cognitive processing in infants and that gaze control is associated to attentional processing in infants as shown previously in adults.

To date, only a few developmental studies have investigated semantic scene-object or background-object inconsistencies in infants and in young children (Bornstein, Mash, & Arterberry, 2011a, 2011b; Duh & Wang, 2014; Richmond & Nelson, 2009). Bornstein et al. (2011a, 2011b) investigated object-context relations in four-month-old infants. The findings of these studies showed that infants presented different eye movement patterns depending on the object-context relation. More fixations were

directed to the objects shown in a natural context than in a white background. Also, objects were more fixated than their background in pictures containing consistent object-context relations whereas the background was more fixated in inconsistent object-context pictures. These results suggest that infants take in account the context to which an object is embedded in (Bornstein et al., 2011a, 2011b). In another study, 9-month-olds were tested in a face-scene paradigm in which they were presented with faces superimposed on three different arbitrary scenes. In an encoding phase, they explored three different face-scenes pairings successively. Then, children were presented with all the faces together with one of the scenes seen previously. The infants were able to discriminate between familiar and novel pairings and looked preferentially at the correct face-scene pairing, suggesting that they encoded the faces in conjunction with the other elements in the scene (Richmond & Nelson, 2009). The first study addressing semantic inconsistency detection in young children was conducted by Duh & Wang (2014). In this study, 15-month-old infants were habituated to visual scenes by repeating the images until their attention decreased (i.e. they stopped looking). After the habituation phase their looking times to the screen were measured in response to an object change. Results show that children detected salient object replacement when the scenes were presented shortly (during 500 ms). However, when the images were presented longer (during 3000 ms), which was assumed to allow an access to the scene meaning, they looked longer at the screen when a change that disrupted the scene meaning occurred (e.g., a beach umbrella replaced by a table) than when a perceptually salient change that preserved the gist (e.g., a beach umbrella replaced by a colourful beach umbrella) was presented. These results suggest that by the age of 15-months, children have acquired at least a preliminary knowledge about scenes to detect semantic inconsistencies. In addition, the results indicate that low-level features attract the attention very fast but when more time is given, semantic processing overrides perceptual effects (Duh & Wang, 2014).

Studies of eye movement patterns in older children using line drawings and natural scenes have shown that fixation distribution is less scattered in children younger than 6 year-old than in older children (Vurpillot, 1968 Egami et al., 2009). In a comparative visual task of line drawings of houses, 3 to 6 year-old children scanned fewer regions of the pictures than 6 to 9 years olds. In another study, based on visual

exploration of line drawings of faces, younger children (4-6 year-olds) presented a fewer number of total gaze points and reduced distribution of the fixations within the image than older children (10-14 year-olds) achieving adult-like performance at 16 years of age (Egami et al., 2009). To our knowledge, Ackik et al. (2010) were the first to investigate eye movement behaviour in children with a typical development during free exploration of natural scenes. In their study, children from 7 to 9 year-old performed less fixations and shorter saccades compared with adults. Likewise, children showed a trend for being less explorative (i.e. presented less scattered distribution of fixations within the scene) and showed stronger bottom-up influences on fixation selection than adults.

#### ***4.3. Language development influence in visual guidance in young children***

Lately, studies using cross-modal paradigms (looking-while-listening paradigms) have shown that language affects looking behaviour in infants and in young children. Looking times analysis during priming tasks has shown a link between spoken word recognition or semantic relations and looking time to visual objects (Arias-Trejo & Plunkett, 2009, 2013; Styles & Plunkett, 2009). For instance, looking times to a named target (e.g., horse) increase when the target is preceded by a semantically related (e.g., rabbit) but not by an unrelated (e.g., door) prime word (Arias-Trejo & Plunkett, 2009, 2013; Styles & Plunkett, 2009). There is also recent evidence indicating that young children use implicit naming when exploring visual images (Khan, 2013; Mani & Plunkett, 2010, 2011). A study in 18-month-olds showed that when children were presented with unlabelled images (e.g., cup) following by a pair of images containing a phonologically similar target (e.g., cat) and unrelated distractor (e.g. shoe), children looked longer to the named target even when the prime image was not named (Mani & Plunkett, 2010). Later, the same paradigm was used in 24-month-old children: the results showed that similarly to adults, the unnamed prime object interfered with the looking time to the related target (Mani & Plunkett, 2011).

During the second year of life, there is an important increasing in vocabulary size (Reznick & Goldfield, 1992; Ganger & Brent, 2004; Mills, Plunkett, Prat, & Schafer, 2005). However, 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. By this



age children have also shown to be prone to spontaneous naming even when no linguistic task is presented (Khan, 2013; Samuelson & Smith, 2005). Furthermore, there is evidence that children at the age of 24-month-olds are more likely to silently name objects than adults (Khan, 2013). Moreover, spontaneous labelling has been found in 2 year-old children when playing freely with novel objects (Samuelson & Smith, 2005). This phenomenon has been further associated to productive vocabulary skills (Samuelson & Smith, 2005). There is no previous evidence of silent naming during visual exploration in young children or whether language skills affect scene processing. However, there is previous evidence that bilingualism has an impact in face processing in adults. In a face recognition task bilingual were more efficient than monolinguals in identify faces. The authors suggested that these differences might have relied in different distribution of the gaze between language groups (Kandel et al., 2016). Based on these previous findings, it is likely that by 2 years of age children spontaneously label objects that are present in their visual environment, which in turns would guide their visual attention to these objects. Further, it is possible that toddlers with higher vocabulary size are more likely to silent naming during scene exploration reflected in their gaze distribution.

#### ***4.4. Semantic inconsistency detection in young children: Event-related potential evidence in visual-linguistic cross-modal studies***

Semantic consistency detection has been also examined in young children by measuring the N400 at object level. In several ERPs studies children were presented with semantically consistent and inconsistent object-word pairs (e.g. Friedrich & Friederici, 2004, 2005a, 2005b, 2010; Torkildsen et al., 2006, 2008; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007). The results of these studies have shown a larger N400-like response for inconsistent than for consistent object labels from the first year of life. Nevertheless, the distribution and the lateralization of N400 component vary across developmental studies (e.g. Friedrich & Friederici, 2004, 2005a, 2005b; Torkildsen et al., 2006, 2007). In addition, vocabulary skills have been shown to contribute to the occurrence and the distribution of the N400 in different priming tasks (Friedrich & Friederici, 2004; Torkildsen et al., 2006, 2008). For instance, the N400 effect was found at 12 months of age in response to word-picture violation only in children with

particularly high level of vocabulary skills (e.g., Friedrich & Friederici, 2010). In 20-month-olds, the N400 effect was found for between category violations in children with both low and high vocabulary skills, while the effect for within category violation was found only in children with a high vocabulary level (Torkildsen et al., 2006). Altogether, these results indicate that by the age of two year-old children have sufficient language experience to build an interconnected lexical-semantic network and that productive vocabulary skills contribute to this process.

## *The aims of the thesis*

There is extensive evidence on the developmental mechanisms underlying oculomotor functions using stimulus-driven tasks. However, less is known about developing eye-movement behaviour during free exploration of scenes. The present dissertation investigates developmental mechanisms underlying scene exploration using eye-tracking and event-related potential (ERP) techniques. The two main aims were to investigate the development of visual attention strategies from early infancy to late childhood and to examine the interaction between bottom-up and top-down control during scene viewing in young children.

Studies I and II aimed at investigating development of oculomotor parameters and visual attention strategies during scene viewing from 3 months to 10 years of age. Participants were presented with visual scenes while their eye movements were recorded. Oculomotor parameters (fixation durations and saccade amplitudes) and scanning strategies (ambient and focal) were compared between age groups. Based on previous research on the development of attention and eye movement control in infants, it was expected that attentional scanning strategies are present already at early stages of development. However, as the distribution of attention and gaze allocation develops during childhood, it was expected that children might be engaged more to focal than to ambient strategies. It was also expected that visual saliency guidance decreases by age.

In Study III eye movements were recorded to examine the influence of top-down (semantic consistency and linguistic) and bottom-up (perceptual saliency) guidance on visual attention during scene viewing in 24-month-old children and adults. Children and adults were presented with real world scenes containing either semantically consistent or inconsistent objects in two saliencies (high and low) conditions. Based on previous evidence on scene context processing in young children, it was expected that young children are able to use scene context on eye movement guidance also during free viewing tasks. However, it was anticipated that the influence of saliency on visual attention guidance is higher in children than in adults. Moreover, it was hypothesized that gaze allocation of children with high and low vocabulary skills may differ since high producers are more likely to activate verbal processes (as naming) during scene exploration.

In Study IV, the ERPs were used to compare the effect of visual scene-context on word processing in 24-month-olds. Children were presented with consistent (e.g., kitchen scene and word 'knife') and inconsistent (e.g., kitchen scene and word 'bus') scene-words

pairs and the ERPs were recorded in response to target words. The aim was to investigate whether contextual information facilitates word processing as manifested by more enhanced N400 component in response to incongruent than to congruent target words. In addition, the influence of language skills in scene-word interactions was investigated. Based on previous literature on language-related ERPs, it was expected that semantic scene context facilitates word processing. Furthermore, it was hypothesized that the N400 effect in children with higher vocabulary skills is more pronounced and occurs earlier than in children with lower vocabulary skills.

## *Experimental Work*

*Study I: In press in Visual Neuroscience*

*Eye movement patterns and visual attention during scene viewing in 3- to 12- month-olds*

Andrea Helo<sup>1,2</sup>, Pia Rämä<sup>1,3</sup>, Sebastian Pannasch<sup>4</sup>, David Meary<sup>5</sup>,

<sup>1</sup>Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France

<sup>2</sup>Departamento de Fonoaudiología, Universidad de Chile, Santiago, Chile

<sup>3</sup>CNRS (UMR 8242), Paris, France

<sup>4</sup>Department of Psychology, Engineering Psychology and Applied Cognitive Research, Technische Universität Dresden, Germany

<sup>5</sup>Laboratoire de Psychologie et NeuroCognition, Université Grenoble Alpes, CNRS, Grenoble, France

Corresponding author: Andrea Helo

Address: Laboratoire Psychologie de la Perception (CNRS UMR 8158), Université Paris Descartes,

45, rue des Saints-Pères, 75006 Paris,

France. Fax: +33 (0)1 42 86 33 22

Email: [andrea.helo@parisdescartes.fr](mailto:andrea.helo@parisdescartes.fr)

## *Abstract*

Recently, two attentional modes have been associated with specific eye movement patterns during scene processing. Ambient mode, characterized by short fixations and long saccades during early scene inspection, is associated with localization of objects. Focal mode, characterized by longer fixations, is associated with more detailed object feature processing during later inspection phase. The aim of the present study was to investigate the development of these attentional modes. More specifically, we examined whether indications of ambient and focal attention modes are similar in infants and adults. Therefore, we measured eye-movements in 3- to 12-months-old infants while exploring visual scenes. Our results show that both adults and 12-month-olds had shorter fixation durations within the first 1.5 seconds of scene viewing compared with later time phases ( $> 2.5$  seconds); indicating that there was a transition from ambient to focal processing during image inspection. In younger infants, fixation durations between two viewing phases did not differ. Our results suggest that at the end of the first year of life infants have developed an adult-like scene viewing behavior. The evidence for the existence of distinct attentional processing mechanisms during early infancy furthermore underlines the importance of the concept of the two modes.

**Keywords:** Scene perception; Visual development; Ambient and focal processing; Eye-movements; Infants

## 1. Introduction

Highest visual acuity is achieved only within a small region in the foveal area. Thus, in order to perceive objects at different locations in visual scenes, eyes execute saccades to center the visual target into the foveal region. Two different attentional visual modes, ambient and focal have been proposed to explain scene exploration through viewing time (Trevarthen, 1968). It has been suggested that the ambient mode dominates the early part, first 2 seconds of scene viewing, and it is characterized by short fixations ( $< 180$  ms) followed by large saccade amplitudes ( $> 5^\circ$ ). This mode allows the localization of objects in the environment (e.g., Scinto & Pillalamarri, 1986). The focal mode appears later in time and is characterized by longer fixations ( $> 180$  ms) surrounded by saccades of reduced amplitude ( $< 5^\circ$ ). This mode has been associated to the process of objects in detail (Henderson, 2007; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Rothert, Dornhoefer, Joos, et al., 2002). Ambient and focal processing strategies have been demonstrated using different types of stimuli (e.g., high *versus* low density, emotional pictures, landscapes) during memorization and free viewing tasks (Follet, Le Meur, & Baccino, 2011; Pannasch, Helmert, Roth, & Walter, 2008; Pannasch, Schulz, & Velichkovsky, 2011; Unema et al., 2005). However, there is evidence that saccade lengths during initial viewing (Castelhano, Mack, & Henderson, 2009) and also their relationship to fixation durations (Castelhano et al., 2009; Mills, Hollingworth, & Dodd, 2011) are influenced by task instruction. It has been proposed that these two attentional mechanisms might be fundamental for our survival. This assumption is supported by the evidence that these attentional mechanisms have been also found in nonhuman primates and various other animal species (Ingle, 1967; Trevarthen, 1968). Recently, we also showed that ambient and focal modes process were present in children as early as 24-months of age (Helo, Pannasch, Sirri, & Rämä, 2014) providing further evidence for their importance as a fundamental mechanism in scene viewing.

Eye-movement control improves extensively already during the first year of life, even though the adult-like levels are reached later during childhood (reviewed in Luna, Velanova, & Geier, 2008). Visual acuity increases rapidly from birth to 7 months of age following by a slower improvement until preschool ages (reviewed in Chandna, 1991). The capability to fixate a target binocularly is present from birth (Slater & Findlay, 1975), but fixation durations decrease with increasing age up to late childhood (Bronson, 1991;



Colombo, Mitchell, Coldren, & Freesean, 1991; Helo et al., 2014; Wass & Smith, 2014). The number of saccades needed to fixate a peripheral target also decreases during the first year of life (Aslin & Salapatek, 1975; Regal, Ashmead, & Salapatek, 1983; Roucoux, Culee, & Roucoux, 1983; Salapatek, Aslin, Simonson, & Pulos, 1980). In addition, smooth pursuit improves rapidly in precision during the first three or four months of age (Phillips, Finocchio, Ong, & Fuchs, 1997; Rosander & Von Hofsten, 2002; Shea & Aslin, 1990; Von Hofsten & Rosander, 1997). The latencies of reflexive saccades also decrease and facility of disengagement from a stimulus increases during the first year of life (Hood, B & Atkinson, 1993; Matsuzawa & Shimojo, 1997). Likewise, the ability to inhibit a reflexive saccade is also present in early infancy but it improves from four to six months of age (Gilmore & Johnson, 1995; Johnson, Posner, & Rothbart, 1994; Johnson, 1995). Furthermore, the coordination between head and eye movements – which is needed to shift a gaze from a fixated target to another during visual exploration - emerges at two months and improves across the first year of life (Bloch & Carchon, 1992; Daniel & Lee, 1990; Nakagawa & Sukigara, 2013; Roucoux et al., 1983).

Despite of extensive amount of evidence on the development of oculomotor functions during the first year of life and early childhood, much less is known about developing eye-movement control during scene viewing. As indicated above, many ocular functions or skills are operative by the end of the first year of life allowing an infant to explore the visual environment proficiently. Moreover, visual attention is notably developing during the first year of life (for reviews, see Colombo, 2001; Johnson, 2002). It has been proposed that there are at least three important phases in the development of visual attention during the first year (for review, see Colombo, 2001). During the first period, from birth to two months of age, alert system develops and infants attain alertness periods more frequently and for longer durations. During the second period, from two to six months of age, visual acuity and eye-movement control improve significantly allowing children to shift attention from one particular location to another. In this period infants also start to attend more to object features. The third period, from 5 or 6 months and beyond, is characterized by significant changes in endogenous attention. During this period, the eye- movement control further increases, infants learn to inhibit attentional shifts, alternate between two stimuli, and also maintain vigilance (for review, see Colombo, 2001). Development of both eye-movement and attention control during the first year of life most probably allows an infant to explore visual scenes efficiently and attentively by the end of this period.

The first aim of the current study was to determine whether eye-movement patterns change from 3- to 12-months of age during a scene exploration task. The second aim was ascertain whether ambient and focal attentional modes are present during the first year of life and when they emerge. We tested four age groups of infants: 3-, 6-, 9-, and 12-month-olds, and compared the results with those of adult participants. Based on earlier findings on studies showing that fixation durations decrease and saccade lengths increase with age (Aslin & Salapatek, 1975; Bronson, 1991; Colombo et al., 1991; Helo et al., 2014; Regal et al., 1983; Roucoux et al., 1983; Wass & Smith, 2014) we expected the same results during free exploration of visual scenes. We also expected the two visual attentional modes (ambient and focal) to emerge by the end of the first year of life. We based our assumption on three developmental evidences: First, several important aspects of eye-movement control are well established by the first year. Second, spatial orienting, associated with ambient mode, and attention to object features needed to inspect objects in detail and related to focal mode are already developed (for reviews, see Colombo, 2001; Johnson, 2002). Third, endogenous attention, fundamental for the guidance of gaze allocation and effective exploration of the scene also appears during the latter half of the first year. Based on findings that the proficiency of motor-ocular functions continues improving after the first year of life (Luna et al., 2008) we predicted a less scattered scan pattern of scenes in infants. Therefore, even if both processing modes are present, ambient processing might be less prominent in infants compared with adults. Accordingly, we assumed stronger dominance of focal processing in infants since a predominance of this mode has been already found in 2-years-old children (Helo et al., 2014). To answer these questions, all the participants freely explored colorful visual scenes while their eye movements were recorded. We analysed eye movement patterns during the whole viewing time and fixation durations and saccade amplitudes during the early and the late phases of scene viewing. We also analyzed the proportion of ambient and focal fixations in all age groups to explore which fixation types are more present in each age group.

## ***2. Materials and Methods***

### ***2.1. Subjects***

A total of 214 subjects participated in the experiment including 44 adult student volunteers (23 females and 21 males, mean age 23 years, range 20-45 years) and 170 infants. All participants were tested in the *Laboratoire de Psychologie et NeuroCognition* of Grenoble. The infants were assigned to four groups according to their age: 3-month-olds (26 girls and 17 boys, mean age 117 days, range 86-127 days), 6-month-olds (23 girls and 21 boys, mean age 191 days, range 181-198 days), 9-month-olds (22 girls and 18 boys, mean age 284 days, range 273-314 days), and 12-month-olds (14 girls and 28 boys, mean age 375 days, range 364-402 days). In order to facilitate reading, we will henceforth use the following labels for the groups of different ages: 3m, 6m, 9m, 12m and adults.

All adult subjects had normal or corrected to normal vision with no hearing impairment. All children were born full-term and presented a typical development. The infants were recruited from a database of parents who volunteered to participate in child development studies, and came from diverse socio-economic backgrounds in the region of Grenoble. The study was conducted in conformity with the declaration of Helsinki and approved by the Ethics Committee of the University of Grenoble-Alpes. Formal written consent was obtained from all parents before testing and the parents filled a short questionnaire concerning their infant daily activity and skills. Overall effective testing time was around 6-7 minutes. The present study lasted around 1 minute.

### ***2.2. Apparatus and Stimuli***

Eye movements were sampled monocularly at 250 Hz using the EyeLink 1000 Remote eye trackers system (SR Research, Ontario, Canada) with a spatial resolution below  $0.01^\circ$  and a spatial accuracy of better than  $0.5^\circ$ . In order to operate the system in head-free mode, a small target sticker was placed on the participants' forehead. The sticker allowed tracking of head position and subsequent correction of gaze position. The Eye-tracker settings were the same in both the infant and the adult group. We used a 5-point calibration procedure (following by a 5-point validation to test calibration accuracy) with small animations ( $<2^\circ$  of visual angle) instead of the classical static EyeLink calibration point. Saccades and

fixations were identified using the saccade detection algorithm supplied by SR Research: Saccades were identified by deflections in eye position in excess of  $0.1^\circ$ , with a minimum velocity of  $30^\circ\text{s}^{-1}$  and a minimum acceleration of  $8000^\circ\text{s}^{-2}$ , maintained for at least 4ms. Six digitized scenes from the Kodak lossless true color image suite (images 3, 7, 11, 14, 22 and 23, <http://r0k.us/graphics/kodak/>). Images were resized to 1024 x 768 pixels to match display resolution. Stimuli were displayed on a CRT display (Iama Vision Master) with a resolution of 1024x728 pixels and a refresh rate of 100 Hz. Screen size was 40.5 x 30 cm. When viewed from 60 cm, the image subtended around 37 degree of visual angle horizontally and 28 degrees vertically.

### ***2.3. Procedure***

Adult participants as well as the parents were informed of the purpose of the study before signing the consent. All participants viewed the scenes from a distance of 60 cm. Infant participants seated on a small chair in front of the screen, in a sound attenuated dimly lit room, with their parents staying in close contact behind them. With the exception of the use of the small chair, adults and infants were tested in the same conditions. Participants of all groups were shown the images without any instructions while the eye movements were recorded. Each trial started with a fixation target located on the top center of the image. Once the participant fixated the target the experimenter launched the presentation of a full-screen scene image for 5 seconds. Scene images were shown in randomized order. The total duration of the experiment, including the preparation for recording and calibration, was less than two minutes per participant.

### ***2.4. Data Analysis***

The data analyses were carried out using SPSS 21.0 and MATLAB 8.4. The first fixation in each trial was defined as the first fixation that began after the onset of the image; as first saccade we considered the saccade following this fixation. All fixations outside the presentation screen, fixations shorter than 80 ms and fixations around eye blinks were discarded. According to earlier findings (e.g. Velichkovsky, Dornhoefer, Pannasch, & Unema, 2000), fixation durations were anticipated to show a right skewed distribution where the median represents a more reliable value than the mean. Therefore, the median

(fixation durations and saccade amplitudes) of each subject was calculated and used for the further analysis. To compare different age groups the mean of medians was used.

For the statistical testing the respective values were subjected to one-way analysis of variance (ANOVA), repeated measures ANOVA, and repeated measures multivariate analysis of variance (MANOVA) and pairwise t-test. Bonferroni correction was used in the post-hoc analyses.

### **3. Results**

#### **3.1. Eye-movement patterns during scene viewing**

In order to assess age-related characteristics of eye movement patterns, median values for fixation durations and saccade amplitudes, and mean values for first saccade latencies, total scan path lengths and total looking times were compared between age groups. A MANOVA revealed a significant multivariate main effect for age group on fixation durations and saccade amplitudes,  $F(2,8) = 21.36$ ,  $p < .001$ . No effect of gender or interactions between factors,  $F = 1.12$  were found. Follow-up ANOVAs confirmed significant differences between the age groups for fixation durations,  $F(4,213) = 46.78$ ,  $p < .001$  (means and standard deviations are presented in ms; 3m: 392 (68), 6m: 371 (88), 9m: 377 (56), 12m: 340 (49), adults: 237 (30)), as well as for saccade amplitudes,  $F(4,213) = 4.58$ ,  $p < .01$  (means and standard deviations are presented in deg; 3m: 5.8 (2.4), 6m: 6.1 (2.8), 9m: 6.4 (2.8), 12m: 4.6 (1.9), adults: 4.9 (1.3)). Post-hoc analyses revealed longer fixations in all infant groups compared to the adults, all  $p < .01$  (Fig. 1A) and fixations in the 3m-group were significantly longer than fixations in the 12m-group,  $p < .01$ . No significant differences were obtained among 3-, 6-, and 9-m groups. Saccade amplitudes were significantly shorter in the 12m-group compared with 6m- and 9m-groups, all  $ps < .05$  (Fig. 1B).

A one-way-ANOVA analysis for the first-saccade latency revealed a significant age group effect,  $F(4,213) = 31.33$ ,  $p < .001$ , while no main effect of gender was found,  $p > .05$ ,  $F < 1$ . Bonferroni post-hoc tests showed that the first-saccade latencies (ms) were longer for the two youngest groups (3m: 941 (246) and 6m: 974 (304)) compared with all the other age groups,  $ps < .05$ . Likewise, the two oldest groups (9m: 848 (161) and 12m: 791 (171)) presented longer latencies compared to adults: 528 (92),  $ps < .05$ . No differences between

3m and 6m ( $p > .05$ ) or between 9m and 12m ( $p > .05$ ) months were found.

Age group had also a significant effect on total scan path lengths,  $F(4,213) = 15.51$ , and total looking times,  $F(4,213) = 10.45$ ,  $ps < .05$ , while no main effects of gender were found,  $ps > .05$ ,  $Fs < 1$ . Bonferroni post-hoc analysis showed that the scan paths were shorter (deg) for all infant groups compared to adults, (3m: 45.6 (17.1), 6m: 48.7 (17.7), 9m: 51.1 (15.12), 12m: 48.5 (17.9), adults: 72.4 (22.5)),  $ps < .05$ . Also total looking times were shorter (ms) for all infant groups compared to adults (3m: 2975 (434), 6m: 2908 (548), 9m: 2636 (426), 12m: 2949 (419), adults: 3406 (257)),  $ps < .05$ . No significant differences between infant groups were found for either measure,  $F < 1$ .

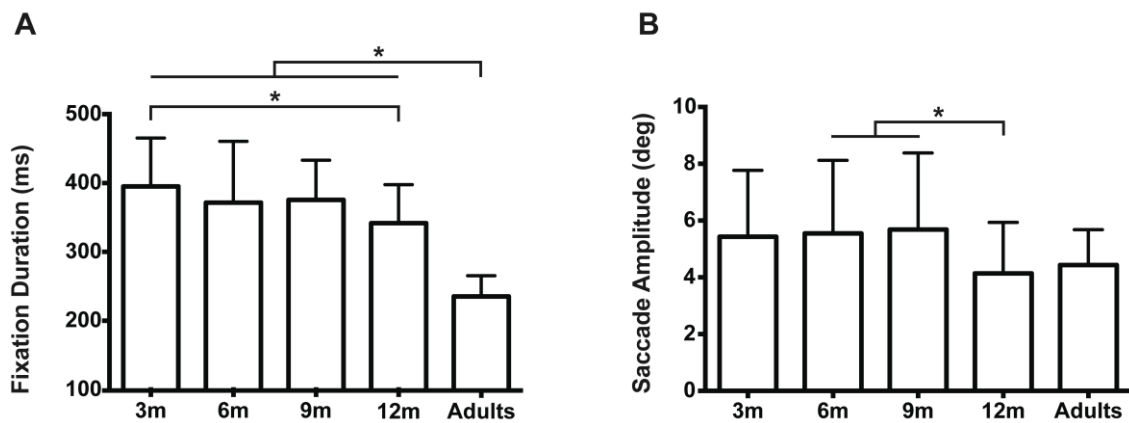


Figure 1. Mean fixation durations (A) and saccade amplitudes (B). All data in mean  $\pm$  SD.

\* indicates  $p < 0.05$ , MANOVA Test.

### 3.2. Time course of fixation durations and saccade amplitudes

Possible influences of time course of visual exploration on the gaze behaviour were analysed by comparing two different viewing phases. Previous studies have shown differences in fixation durations and saccade amplitudes between the early viewing phase comprising the time interval from 0 to 2 seconds and the late viewing phase covering from 4 to 6 seconds (Pannasch et al., 2008). However, the total viewing time per image was shorter (5 seconds) than in previous studies (10 seconds) and thus, we chose slightly different time windows for our study (0-1.5s versus 2.5-4s)

For both viewing phases, we calculated the median for fixation durations and for

saccade amplitudes *per* subject; these individual medians were used for statistical testing. Since no effects or interactions of gender were found in the previous analyses, this variable was excluded from the following analyses. Fixation durations and saccade amplitudes were examined for differences along the time course by conducting a 5 (age group) x 2 (viewing phase) repeated measures ANOVA for each parameter.

The analysis for fixation durations during early and late phases (means and standard deviations are presented in ms; 3m: early 414 (91), late 395 (89), 6m: early 371 (94), late 380 (96), 9m: early 377 (63), late 392 (110), 12m: early 332 (70), late 364 (74), adults: early 321 (31), late 254 (43)) revealed main effects for age group,  $F(4,202) = 47.7$ ,  $p < .001$ , and for viewing phase  $F(1,202) = 4.33$ ,  $p < .05$ . Fixation durations increased from early to late phase in all, but 3m group, and thus, the interaction between age group and phase was only close to significance,  $F(4,202) = 2.07$ ,  $p = .086$ . When the two age groups at the middle (6- and 9-month-olds) were combined, the interaction between age group and phase was significant,  $F(3,203) = 2.76$ ,  $p < .05$ . Following pairwise *t*-tests revealed that fixation durations were significantly longer during the late than the early viewing phase in 12 month-olds,  $t(40) = -2.35$ ,  $d = 0.35$ , and in adults,  $t(43) = -6.56$ ,  $d = 1.06$ , both  $p < .05$ .

ANOVA for saccade amplitudes during early and late phases (means and standard deviations are presented in deg; 3m: early 5.9 (2.8), late 5.9 (2.9), 6m: early 6.5 (2.9), late 6.4 (3.3), 9m: early 6.7 (3.1), late 7.5 (3.1), 12m: early 4.3 (3.1), late 5.8 (3.6), adults: early (1.8), late 4.9 (1.4)) revealed main effects for age group,  $F(4,202) = 7.06$ , and for viewing phase  $F(1,202) = 4.79$ ,  $p < .05$ . Saccades amplitudes increased, or were similar between two phases in all age groups, and thus, no significant interaction was found,  $F = 1.53$ ,  $p > .05$ . Even when two age groups (6 and 9-months) were combined, the interaction between age group and phase did not reach significance  $F(3,203) = 1.61$ ,  $p < .05$ . However, a paired *t*-test showed that saccade amplitudes were significantly longer during the late than the early phase only in 12-month-olds  $t(40) = -2.45$ ,  $p < .05$ ,  $d = 0.46$ , but not in other age groups or in adults ( $p > .05$ ).

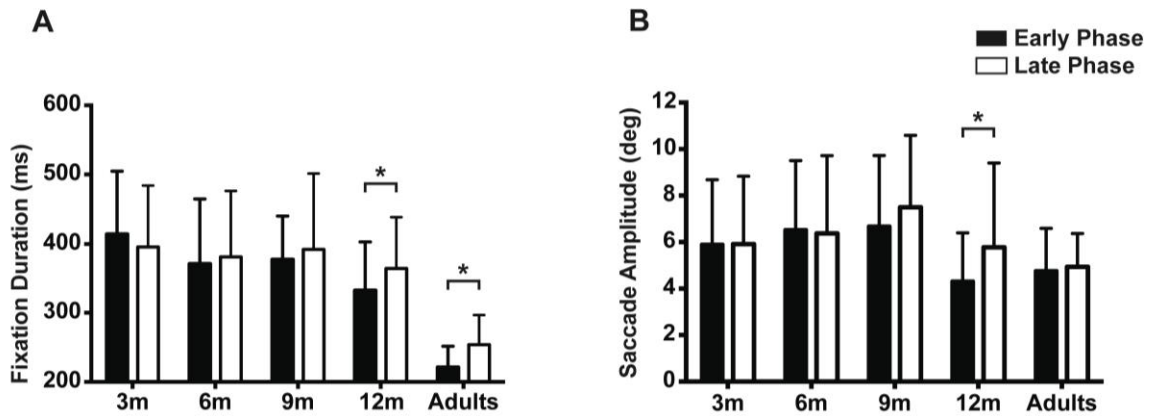


Figure 2. Mean fixation duration (A) and saccade amplitude (B) during early and late viewing phases. All data in mean  $\pm$  SD. \* indicates  $p < 0.05$ , Repeated measure ANOVA Test.

### 3.3. Proportion of “ambient” and “focal” fixations in each age group

In order to know whether the proportion of ambient and focal fixations was different between the age groups, we categorized the fixations based on fixation duration and *following* saccade amplitude. Short fixations ( $< 180$  ms) followed by long saccades ( $> 5$  deg) were considered as ambient fixations, whereas long fixations ( $> 180$  ms) followed by short saccades ( $< 5$  deg) were considered as focal fixations (Unema et al., 2005). To compare the proportion of ambient and focal fixation between groups, a 5 (age group)  $\times$  2 (fixation type) repeated measures ANOVA was conducted. Statistical testing revealed a significant main effect of fixation type,  $F(1,209) = 947.34$ ,  $p < .001$ , group effect,  $F(1,209) = 3.72$ . Furthermore, there was a significant interaction between group and fixation type,  $F(4,209) = 10.85$ ,  $p < .001$ .

All groups showed higher proportion of focal than ambient fixations. Post-hoc analyses revealed the lowest proportion of ambient fixations for all groups of infants compared with adults; all  $p$ -values  $< .05$ ). The proportions furthermore did not differ between infant groups.



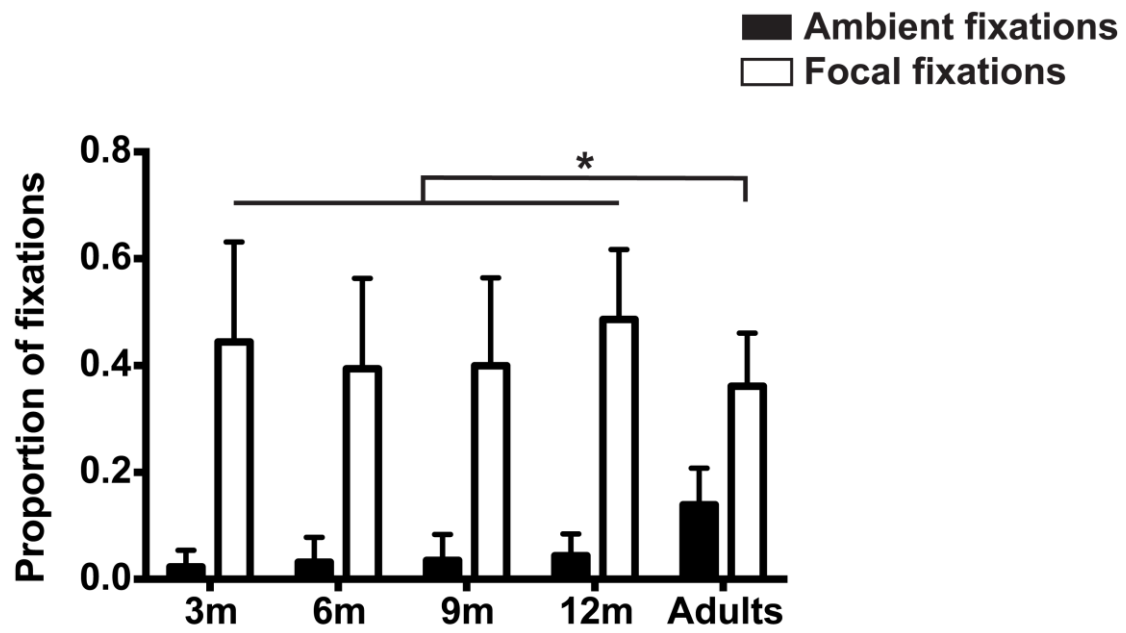


Figure 3. Proportion of ambient and focal fixations for the different age groups. All data in mean  $\pm$  SD. \* indicates  $p < 0.05$ , Repeated measure ANOVA Test.

#### *4. Discussion*

Our aim was to investigate whether eye-movement characteristics in infants during scene viewing reveal indications for processing according to the two modes approach. The results showed that fixation durations decreased with growing age while saccade amplitudes were similar in infants and adults. Fixation durations were longer in the 3- to 9- month-olds compared with 12-month-olds and adults. Previous evidence using standardized experimental assessments has shown that fixation durations decrease with age (Bronson, 1991; Colombo et al., 1991). We have previously shown that fixation durations decrease by age from 2- to 10-years of age (Helo, Pannasch, Sirri, & Rämä, 2014). Altogether our previous and the current results indicate that there is age-related decrease in fixation durations from 3-months to 6-years of age, at which age adult values are reached. Longer fixations have been associated with higher cognitive effort, difficulties in information extraction, and with lesser amount of visual exploration in infants (Bronson, 1990, 1991, 1994; Colombo, Freeseaman, Coldren, & Frick, 1995; Colombo et al., 1991; Wass & Smith, 2014). Accordingly, our results indicate that visual information processing becomes easier by the end of the first year of life.

First-saccade latencies were longer for 3-month- and 6-month-olds compared with all the other age groups, and also for 9-month- and 12-month-olds presented longer latencies compared to adults, indicating that even if the latencies decrease, they do not reach adult values during the first year of life. It has been suggested that the initial fixation activates scene category or schema (Potter, 1976; Schyns & Oliva, 1994) and provides information for planning the subsequent eye movements (Castelhano & Henderson, 2007, 2008). Our results suggest that the youngest infants might devote a higher cognitive effort to activate scene category or planning the following saccade than the older infants and adults. In contrast, saccade lengths (as measured during the whole viewing time) in infants were not different from adults but the two youngest group presented longer saccades lengths compared with 12-month-olds. This finding differs from previous results, which showed that infants typically perform hypometric saccades, at least in tasks requiring shifting of gaze toward peripheral targets (Aslin & Salapatek, 1975; Regal et al., 1983; Roucoux et al., 1983).

The second aim of our study was to determine whether and when infants exhibit ambient and focal attentional modes during exploration of scenes. We compared fixation

durations and saccade amplitudes at the early and late phases of viewing. We also compared the proportion of ambient and focal fixations between age groups. The oldest infant group (12-month-olds) and the adult group had shorter fixations during the early than the late phase of scene viewing whereas fixation durations of the three youngest age groups (3- to 9-month-olds) were equally long during both phases of viewing time. These findings suggest that the transition between ambient and focal modes during viewing time appears at the end of the first year of life. Short fixation durations during the early phase have been previously associated to ambient mode during which a viewer orientates to object locations within the scene while longer fixation durations during the later viewing phase have been associated with focal processing mode, allowing processing of objects in detail (Henderson, 2007; Unema et al., 2005; Velichkovsky, Rothert, Dornhoefer, & Joos, 2002). However, contrary to previous findings, even fixation durations increased with viewing time, both infants and adults exhibited either similar saccade lengths in two viewing phases or increased lengths from the early to late viewing phase. This result is not in accordance with previous reports supporting the two visual attentional modes hypothesis that show longer saccade lengths at the beginning of viewing compared with shorter lengths at the late viewing time (Henderson, 2007; Unema et al., 2005; Velichkovsky, Rothert, Dornhoefer, & Joos, 2002). Our finding is rather in accordance with other reports showing that saccade lengths increase during viewing time (e.g. Castelhana & Henderson, 2008; Castelhana et al., 2009; Mills et al., 2011). In these studies, saccades lengths increased when visual information was degraded (Castelhana & Henderson, 2008), or when cognitive task was implemented (Castelhana et al., 2009; Mills et al., 2011). It is possible that the scenes used in the current study were too simple and were not providing enough background features, which might have contributed to lack of differences in saccade lengths between viewing phases in adults. However, it is difficult to explain why 12-month-olds increased their saccade lengths while younger infants had similar lengths throughout the viewing time. One possibility is that since scene processing is more difficult for children than for adults, 12-month-olds increased both fixation durations and saccade lengths in order to process and extract information as efficiently as possible.

Nevertheless, our findings suggest that by the end of the first year of life, infants change their scanning strategies during the viewing time. Earlier, it has been proposed that spatial orientating (related to ambient mode) and attention to object features (related to focal mode) are emerging around 4 to 6 months of age (for reviews, see Colombo, 2001; Johnson, 2002). Therefore, it is possible that, by the end of the first year of life these

processes are mature enough and allow infants to modify their attentional strategies depending on the viewing time. Additionally, it has been shown that endogenous attention, which might partly determine how engaged infants are to viewing or exploration, improves at the end of the first year of life (Courage, Reynolds, & Richards, 2006). However, in our study, all infant age groups presented similar total looking times across the trials suggesting that they were equally engaged to the task, and the difference in total looking times was only present when compared infants with adults. This result indicates that improved global endogenous attention at the end of first year do not explain our results.

Even if our results indicate that ambient and focal processing modes might emerge by the age of 12 months, the proportion of ambient fixations was lower in all infant groups compared to adults. Moreover, the total scan path length, which measures the gaze distribution within the scene was shorter in all infant age groups compared with adults. We have earlier shown that children from 2- to 8- years of age exhibit lower proportion of ambient fixations and higher proportion of focal fixations than adults up to the age of 8 years (Helo et al., 2014). These results indicate that the focal mode dominates in children. We offer two possible explanations for these findings: first, it is possible that attention to object features matures earlier than spatial orienting and this slower maturation is reflected by a dominance of focal mode. Second, it is possible that more advanced oculo-motor functions (e.g., fixation and saccade programming) that are needed to orientate within scenes (ambient mode) are still immature by the end of the first year, or even by later childhood.

To sum up, our results showed that 12-month-olds but not younger infants use two different attentional strategies while exploring scenes. Longer fixation durations and longer first saccade latencies between 12-month-olds and adults suggest that even if these strategies are functional, infants devote more cognitive effort to scene processing. In addition, infants explore the scenes in a less scattered manner as indicated by shorter scan path lengths and lower proportion of ambient fixations. Altogether, our results suggest that attention mechanisms associated with scene viewing are developing in early infancy but adult-like strategies might be reached later during childhood.

## *Acknowledgments*

We express our gratitude to the infants and parents who participated to the study, we thank for their kindness and cooperation. We also thank the *Laboratoire de Psychologie et NeuroCognition* of Grenoble for testing the participants. A.H. was supported by doctoral fellowship from CONICYT, Chile.

## References

- Aslin, R. N., & Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Perception & Psychophysics*, *17*(3), 293–302. doi:10.3758/BF03203214
- Bloch, H., & Carchon, I. (1992). On the onset of eye-head coordination in infants. *Behavioural Brain Research*, *49*(1), 85–90. doi:10.1016/S0166-4328(05)80197-4
- Bronson, G. (1990). Changes in infants' visual scanning across the 2- to 14-week age period. *Journal of Experimental Child Psychology*, *49*(1), 101–125. doi:10.1016/0022-0965(90)90051-9
- Bronson, G. (1991). Infant differences in rate of visual encoding. *Child Development*, *62*(1), 44–54. doi:10.2307/1130703
- Bronson, G. (1994). Infants' transitions toward adult-like scanning. *Child Development*, *65*(5), 1243–1261. doi:10.2307/1131497
- 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. doi:10.1037/0096-1523.33.4.753
- Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal Experimental Psychology*, *62*(1), 1–14. doi:10.1037/1196-1961.62.1.1
- 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. doi:10.1167/9.3.6.Introduction
- Chandna, A. (1991). Natural history of the development of visual acuity in infants. *Eye*, *5*, 20–26. doi:10.1038/eye.1991.4
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*, 337–67.
- Colombo, J., Freeseaman, L. J., Coldren, J. T., & Frick, J. E. (1995). Individual differences in infant fixation duration: Dominance of global versus local stimulus properties. *Cognitive Development*, *10*(2), 271–285. doi:10.1016/0885-2014(95)90012-8
- Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseaman, L. J. (1991). Individual differences in infant visual attention: are short lookers faster processors or feature processors? *Child Development*, *62*(6), 1247–1257. doi:10.1111/j.1467-8624.1991.tb01603.x
- Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' attention to patterned stimuli: Developmental change from 3 to 12 months of age. *Child Development*, *77*(3), 680–695. doi:10.1111/j.1467-8624.2006.00897.x
- Daniel, B. M., & Lee, D. N. (1990). Development of looking with head and eyes. *Journal of Experimental Child Psychology*, *50*(2), 200–216. doi:10.1016/0022-0965(90)90039-B
- Follet, B., Le Meur, O., & Baccino, T. (2011). New insights into ambient and focal visual fixations using an automatic classification algorithm. *I-Perception*, *2*(6), 592–610. doi:10.1068/i0414

- Gilmore, R., & Johnson, M. H. (1995). Working memory in infancy: Six-month-olds' performance on two versions of the oculomotor delayed response task. *Journal of Experimental Child Psychology*, *59*, 397–418.
- 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. doi:10.1016/j.visres.2014.08.006
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219–222. doi:10.1111/j.1467-8721.2007.00507.x
- Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, *16*(4), 405–422. doi:10.1016/0163-6383(93)80001-O
- Ingle, D. (1967). Two visual mechanisms underlying the behavior of fish. *Psychologische Forschung*, *31*(1), 44–51. doi:10.1007/BF00422385
- Johnson, M. H. (1995). The inhibition of automatic saccades in early infancy. *Developmental Psychobiology*, *28*(5), 281–291. doi:10.1002/dev.420280504
- Johnson, M. H. (2002). The development of visual attention: A cognitive neuroscience perspective. In M. H. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain Development and Cognition: A Reader* (Second Edi.). Oxford: Blackwell Publishers Ltd.
- Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1994). Facilitation of saccades toward a covertly attended location in early infancy. *Psychological Science*, *5*(2), 90–92. doi:10.1111/j.1467-9280.1994.tb00636.x
- Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293–308. doi:10.1016/j.bandc.2008.08.019
- 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. doi:10.1016/S0163-6383(97)90035-7
- 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. doi:10.1167/11.8.17.Introduction
- Nakagawa, A., & Sukigara, M. (2013). Variable coordination of eye and head movements during the early development of attention: A longitudinal study of infants aged 12-36 months. *Infant Behavior and Development*, *36*(4), 517–525. doi:10.1016/j.infbeh.2013.04.002
- Pannasch, S., Helmert, J. R., Roth, K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, *2*(2), 1–19.
- Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception & Psychophysics*, *73*(4), 1120–1132. doi:10.3758/s13414-011-0090-1
- Phillips, J. O., Finocchio, D. V., Ong, L., & Fuchs, A. F. (1997). Smooth Pursuit in 1-to 4-month-old Human Infants. *Vision Research*, *37*(21), 3009–3020.
- Regal, D. M., Ashmead, D. H., & Salapatek, P. (1983). The coordination of Eye and head movements during early infancy: A selective review. *Behavioural Brain Research*, *10*,

125–132.

- Rosander, K., & Von Hofsten, C. (2002). Development of gaze tracking of small and large objects. *Experimental Brain Research*, *146*(2), 257–264. doi:10.1007/s00221-002-1161-2
- Roucoux, a., Culee, C., & Roucoux, M. (1983). Development of fixation and pursuit eye movements in human infants. *Behavioural Brain Research*, *10*(1), 133–139. doi:10.1016/0166-4328(83)90159-6
- Salapatek, P., Aslin, R. N., Simonson, J., & Pulos, E. (1980). Infant saccadic eye movements to visible and previously visible targets. *Child Development*, *51*(4), 1090–1094.
- Shea, S. L., & Aslin, R. N. (1990). Oculomotor responses to step-ramp targets by young human infants. *Vision Research*, *30*(7), 1077–1092. doi:10.1016/0042-6989(90)90116-3
- Slater, A. M., & Findlay, J. M. (1975). Binocular Fixation in the newborn baby. *Journal of Experimental Child Psychology*, *20*, 248–273.
- Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *31*(4), 299–337. doi:10.1007/BF00422717
- Unema, P. J. a., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494. doi:10.1080/13506280444000409
- Velichkovsky, B. M., Rothert, A., Dornhoefer, S. M., & Joos, M. (2002). Towards an Express-Diagnostics for Level of Processing and Hazard Perception. *Transportation Research.*, *5*(2), 145–146.
- Velichkovsky, B. M., Rothert, A., Dornhoefer, S. M., Joos, M., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 145–156. doi:10.1016/S1369-8478(02)00013-X
- Von Hofsten, C., & Rosander, K. (1997). Development of Smooth-Pursuit Tracking in Young Infants. *Vision Research*, *37*(13), 1799–1810.
- 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. doi:10.1111/infa.12049



*Study II: Published in Vision Research*

*The maturation of eye movement behavior: Scene viewing characteristics  
in children and adults*

Helo, Andrea<sup>1,2</sup>, Pannasch, Sebastian<sup>3</sup>, Sirri, Louah<sup>1</sup>, & Rämä, Pia<sup>1,4</sup>.

<sup>1</sup>Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France.

<sup>2</sup>Escuela de Fonoaudiología, Universidad de Chile, Santiago, Chile.

<sup>3</sup>Department of Psychology, Engineering Psychology and Applied Cognitive Research,  
Technische Universität Dresden, Germany

<sup>3</sup>Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France.  
Email:

<sup>4</sup>CNRS (UMR 8158), Paris, France.

Corresponding author: Andrea Helo

Address: Laboratoire Psychologie de la Perception (CNRS UMR  
8158), Université Paris Descartes,

45, rue des Saints-Pères, 75006 Paris,

France. Fax: +33 (0)1 42 86 33 22

Email: [ahelo@med.uchile.cl](mailto:ahelo@med.uchile.cl)

## *Abstract*

The close link between eye movements and visual attention has often been demonstrated. In addition, ambient and focal attentional modes have recently been associated with specific eye movement patterns. The ambient mode—serving the localization of objects and dominating early scene inspection—is expressed by short fixations and large saccade amplitudes. The focal mode—associated with the identification of object details and dominating later stages of scene exploration—is indicated by longer fixations embedded in short saccades. The relationship between these processing modes and eye movement characteristics has so far only been examined in adults. While studies in children revealed a maturation of oculomotor behavior up to adolescence, the developmental aspects of the processing modes are still unknown. The current study explored these mechanisms by investigating eye movements of children of four different age groups (2, 4-6, 6-8, 8-10 years old) and adults while inspecting naturalistic scenes. In general we found that, with growing age fixation durations decrease and saccades amplitude increase. In relation to the time course of scene inspection, we found for all age groups shorter fixations and longer saccade amplitudes at the beginning. In the later phase, fixations were longer and saccades were shorter. At the beginning of scene inspection, a strong saliency influence on eye guidance was found for all participants, but for the two youngest groups this influence prevailed the overall time course. These results indicate that ambient and focal processing strategies can be observed as early as two years of age.

**Keywords:** Scene perception; Visual development; Ambient and focal processing; Eye-movements

## ***1. Introduction***

Vision is an active process through which the visual system extracts information from the environment. The highest visual quality is achieved in a small region at center of gaze, covering the foveal and parafoveal range of about 5 degrees of visual angle. Therefore, saccades, rapid eye movements, are necessary to center new visual information to the foveal region. Acquiring information from the environment takes place during fixations—periods of relative gaze stability—but is suppressed during saccades (e.g., Findlay & Walker, 1999; Thiele, Henning, Kubischik, & Hoffmann, 2002; Volkman, 1986). To extract relevant fragments of a scene, eye movements are performed to allow fixating these regions. Thus, the spatial location of an eye-fixation is understood as an indicator of where visual attention is allocated (Henderson, 2007; Tatler & Vincent, 2008).

Eye movement patterns during scene perception have been investigated extensively in adults (for reviews, see Martinez-Conde, Macknik, & Hubel, 2004; Rayner, 2009; Schütz, Braun, & Gegenfurtner, 2011). Much of the research has been concerned with processes of spatial target selection. Principally, where viewers tend to fixate within a scene and which factors modulate this selection. It has been proposed that information processing and scan patterns are influenced by the interaction of bottom-up and top-down factors (Henderson, 2003; Oliva, 2005; Torralba, Oliva, Castelhana, & Henderson, 2006). Bottom-up processing is guided by low-level features of images, such as saliency (Itti & Koch, 2000; Tatler, Baddeley, & Gilchrist, 2005; Tatler & Vincent, 2008) whereas top-down processing is based on endogenous control, such as semantic schema knowledge, working memory and behavioral task demands (Castelhana, Mack, & Henderson, 2009; Fischer, Graupner, Velichkovsky, & Pannasch, 2013; Mills, Hollingworth, & Dodd, 2011; Tatler & Vincent, 2008).

Examining the time course of scene exploration revealed that early in scene inspection mainly saliency aspects influence viewing behavior while later in time top-down control becomes more dominant (Castelhana et al., 2009; Mills et al., 2011; Tatler & Vincent, 2008). These processing characteristics have recently been related to the patterns of eye movements, demonstrating a relationship between saccade amplitude, fixation duration and the respective attention mode (Pannasch, Helmert, Roth, & Walter, 2008; Pannasch & Velichkovsky, 2009; Tatler & Vincent, 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005; Velichkovsky, Rothert, Kopf, Dornhöfer, & Joos, 2002). Attentional processing has often been distinguished into ambient and focal modes

(Trevarthen, 1968). The ambient mode is expressed by short fixations (<180 ms) followed by large amplitude saccades (>5°). The focal mode, in contrast, is characterized by long fixations (>180 ms) surrounded by saccades of relatively short amplitudes (<5 dg) (Tatler & Vincent, 2008; Unema et al., 2005; Velichkovsky et al., 2005, 2002). It has been suggested that ambient processing is related to the overall spatial orientation in a scene, whereas focal processing serves the identification of objects (Pannasch & Velichkovsky, 2009; Velichkovsky et al., 2005, 2002). For example, Velichkovsky et al., (2002) demonstrated that ambient fixations are related to the detection of hazardous events in a dynamic driving task, whereas focal fixations are more related to the identification of these events. A subsequent experiment using static images revealed better recognition performance for fragments inspected in focal mode in contrast to those that were explored by ambient fixations (Velichkovsky et al., 2005). Based on these findings, it was suggested that the ambient and focal processing modes may recruit two distinct visual pathways (Pannasch & Velichkovsky, 2009; Velichkovsky et al., 2005, 2002). Ambient processing was related to dorsal pathway activity, including areas located in the dorsal occipital and parietal lobes, by rapidly transferring visual information of low spatial resolution. Therefore, this pathway has been associated with the processing of spatial relations and motion direction thereby enabling the spatial exploration of the environment (Corbetta, Patel, & Shulman, 2008; Milner & Goodale, 2008). In contrast, the ventral pathway, including the temporo-parietal junction and the ventral frontal cortex, was related to the detection of salient and behaviorally significant stimuli as well as to the processing of object representations (Corbetta et al., 2008; Milner & Goodale, 2008). Therefore, it has been linked to the focal mode. So far, behavioral and neuroanatomical aspects of the two processing modes have been examined in adults. However, it still remains unknown if this dichotomy is already present early in life or if not, when does it evolve throughout development.

The oculomotor behavior develops in several steps during early childhood and adolescence. The capability to fixate a target is acquired during the first few months of life (Chandna, 1991; Roucoux, Culee, & Roucoux, 1983) but more complex aspects of the fixation system, such as steadiness of fixations and cognitive control continues to develop until adolescence (Luna, Velanova, & Geier, 2008). For instance, when maintaining a fixation on a target, durations increase while the number of reflexive saccades decrease between 4 to 15 years of age (Aring, Grönlund, Hellström, & Ygge, 2007; Ygge, Aring, Han, Bolzani, & Hellström, 2005). Furthermore, 8-years-olds are less proficient than 10-years-olds in maintaining fixation when peripheral distractors appear, suggesting that cognitive control of

fixation develops until the age of 10 (Paus, Babenko, & Radil, 1990).

Regarding the development of saccade control, parameters such as peak velocity, latency and accuracy have been investigated (reviewed in Luna et al., 2008). While some studies found no age specific differences for saccade velocity (e.g. Luna, Garver, Urban, Lazar, & Sweeney, 2004; Munoz, Broughton, Goldring, & Armstrong, 1998), others reported an increase in saccade velocity from 3 to 14 years of age (Irving, Steinbach, Lillakas, Babu, & Hutchings, 2006). Amplitudes of saccades were found to be shorter and spatially less precise when comparing children with adults. However, these parameters tend to stabilize around the age of 10 years (Fioravanti, Inchingolo, Pensiero, & Spanio, 1995; Irving et al., 2006; Munoz et al., 1998). Likewise, saccade latencies have been found to decrease until the age of 15 years (Fukushima, Hatta, & Fukushima, 2000; Irving et al., 2006; Klein & Foerster, 2001; Luna & Sweeney, 2004; Munoz et al., 1998). Furthermore, cognitive control of saccade execution, operationalized by the performance in pro- and anti-saccades tasks, reaches an adult-like performance level at around 10 to 12 years of age (Fukushima et al., 2000; Irving, Tajik-Parvinchi, Lillakas, González, & Steinbach, 2009; Klein & Foerster, 2001). Taken together, similarly to the development of the fixation system, most aspects of saccade control continue developing through childhood.

Developmental aspects of eye movement control have been investigated mainly with the use of artificial stimuli and tasks (for review, see Luna et al., 2008). Only few studies employed naturalistic stimuli (e.g., Egami et al., 2009; Vurpillot, 1968). In two studies, using visual comparative search, it was shown that children from 3 to 6 years-olds exhibit less exploratory eye movements than older children (6 to 14 years-olds); indicated by fewer and scarcely distributed fixations (Egami et al., 2009; Vurpillot, 1968). Results of studies on global and local processing of hierarchical visual forms have suggested a local processing bias in infants and younger children (e. g. Dukette & Stiles, 1996, 2001; Poirel, Mellet, Houdé, & Pineau, 2008; Vinter, Puspitawati, & Witt, 2010). These studies have shown that the processing of local stimulus properties dominates until the age of 4 years whereas processing in a more global manner becomes more influential around 6 to 9 years of age (Dukette & Stiles, 1996, 2001; Poirel et al., 2008; Vinter et al., 2010). Moreover, the local processing dominance corresponds with the fewer exploratory eye movements in younger children (Poirel et al., 2008) and might be related to the maturation of parietal and visual cortical areas (Poirel et al., 2011). The few studies examining eye-movement patterns during scene viewing have tested atypically developed children (e.g. autistic and Williams syndrome) focusing mainly on fixation landing and scene comprehension (Fletcher-Watson, Leekam,

Benson, Frank, & Findlay, 2009; Riby & Hancock, 2008, 2009).

Recent reports revealed that the functionality of the dorsal and ventral pathways is developing during childhood (Braddick & Atkinson, 2011; Dekker, Mareschal, Sereno, & Johnson, 2011; Gordon & McCulloch, 1999; Parrish, Giaschi, Boden, & Dougherty, 2005). There seem to be characteristic patterns (Braddick, Atkinson, & Wattam-Bell, 2003; Braddick & Atkinson, 2011; Gunn et al., 2002), suggesting that the ventral pathway matures earlier than the dorsal pathway (Armstrong, Maurer, & Lewis, 2009; Gunn et al., 2002; Klaver et al., 2008). This is indicated, for instance, by the performance in the form coherence task—which is associated with the ventral pathway—children reach a similar level as adults with of 6-7 years of age. In contrast, performance in the motion coherence task—which is associated to the dorsal pathway—becomes comparable to the level of adults by the age of 10-11 years (Armstrong et al., 2009; Gunn et al., 2002; Klaver et al., 2008). Moreover, atypically developed children (e.g. developmental hemiplegic, dyslexia and autism) show greater impairment in spatial tasks, suggesting that dorsal network is more vulnerable during development (Gunn et al., 2002; Koldewyn, Whitney, & Rivera, 2011; Ridder, Borsting, & Banton, 2001). These findings propose differences in the development of the two distinct visual pathways, but so far, these aspects were not linked to ambient and focal processing modes.

In summary, empirical evidence revealed several developmental steps in the maturation of the visual system from childhood to adolescence (reviewed in Luna et al., 2008). These findings are based mostly on studies employing simple and artificial stimuli, while only few studies examined age-related differences in eye movement patterns with the use of complex or naturalistic stimuli (Egami et al., 2009; Vurpillot, 1968). In particular, there is a gap in knowledge regarding eye-movement patterns of children in free visual exploration. Therefore, in the current study, we investigated eye-movement patterns during scene perception in typically developing children in large range of age, covering the developmental period from 2 to 10 years. Our first aim was to examine to what extent eye-movement pattern of children and adults are similar. In particular we were interested in the time-course specific characteristics, that is, ambient processing dominance early in scene exploration followed by the prevailing focal processing during later phases (Pannasch et al., 2008; Tatler & Vincent, 2008; Unema et al., 2005). Considering that the ventral pathway might mature before the dorsal pathway, we expect the initial ambient processing being less prominent in younger children compared with older children and adults. For focal processing, in contrast, we expect stronger indications in younger children since previous findings revealed a dominance of local

processing below the age of 6 years (Dukette & Stiles, 1996, 2001; Poirel et al., 2008; Vinter et al., 2010). Our second aim was to determine whether the influence of bottom-up saliency on the guidance of eye-movements is similar for all age groups. Since it is known that cognitive resources, which are directly related to top-down control, increase by age (Gathercole, Pickering, Ambridge, & Wearing, 2004; Gathercole, 1999; Hitch, Halliday, Schaafstal, & Schraagen, 1988; Klenberg, Korkman, & Lahti-Nuuttila, 2010; Pearson & Lane, 1991; Pickering, 2001; Sanders, Stevens, Coch, & Neville, 2006), we anticipate a stronger impact of saliency on eye-movement behavior for younger children. We furthermore expect that saliency is mainly determining eye-movement behavior during the early phase of scene viewing as shown before (Castelhano et al., 2009; Mills et al., 2011; Tatler & Vincent, 2008). To answer these questions, our participants freely explored scenes while their eye movements were recorded. We analysed fixation durations and saccade amplitudes during the early phase and late phase of scene viewing. Furthermore we examined the amount of focal fixations in all tested age groups to explore whether focal processing is more evident in younger than in older children and adults. Finally, to study influences of low-level image features (i.e. saliency) on eye movement behavior, observed fixation positions were compared with the predictions of a saliency model between groups and phases.

## ***2. Materials and Methods***

### ***2.1. Subjects***

A total of 101 subjects participated in the experiment including 23 adult volunteers (14 females and 9 males, mean age 29 years, range 25-39 years) and 78 children. The children were assigned to four groups according to their age: 2-years-olds (9 girls and 9 boys, mean age 24,8 months, range 24-26 months), 4 to 6-years-olds (14 girls and 8 boys, mean age 4,6 years, range 4,0-5,8 years), 6 to 8-years-olds (14 girls and 4 boys, mean age 7,1 years, range 6,3-7,11 years), and 8 to 10-years-olds (8 girls and 12 boys, mean age 9,1 years, range 8,2-10,1 years). In order to facilitate reading, we will henceforth use the following labels for the groups of different ages: 2y-group, 4-6y-group, 6-8y-group, 8-10y-group and adults-group.

The children were recruited from a database of parents who volunteered to participate in child development studies, and came from diverse socio-economic backgrounds in the Parisian region. All subjects had normal or corrected to normal vision with no hearing or

language impairment. All children were born full-term and presented a typical development. The study was conducted in conformity with the declaration of Helsinki and approved by the Ethics Committee of the University of Paris Descartes. Four additional children were recruited but their data were rejected due to calibration problem.

## ***2.2. Apparatus***

Eye movements were sampled monocularly at 500 Hz using the EyeLink 1000 Remote eye trackers system (SR Research, Ontario, Canada) with a spatial resolution below  $0.01^\circ$  and a spatial accuracy of better than  $0.5^\circ$ . 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). Saccades and fixations were defined using the saccade detection algorithm supplied by SR Research: Saccades were identified by deflections in eye position in excess of  $0.1^\circ$ , with a minimum velocity of  $30^\circ\text{s}^{-1}$  and a minimum acceleration of  $8000^\circ\text{s}^{-2}$ , maintained for at least 4ms. Pictures were displayed using a GeForce 7300 GT card and a CRT display (Sony GDM F520) at 1024x728 pixels at a refresh rate of 100 Hz viewed from a distance of 60 cm.

## ***2.3. Stimuli***

Thirty digitized scenes with a resolution of 1024 x 768 pixels served as stimuli. These scenes were color pictures taken from children books or movies with several central objects or characters and an eventful background. From half of these images, we took significant image segments with a size of 200 x 200 pixels. Another fifteen image segments were taken from similar images, serving as catch-trial image segments.

## ***2.4. Procedure***

Children participants and their parents as well as the adult participants were informed of the purpose of the study before signing the consent. Participants seated in a sound attenuated, dimly lit room; only the 2-years-olds were sitting on the laps of their parents.

Participants were instructed to explore the images. Before the start of the experiment and before each block (see below) a 5-point calibration and validation was performed. Each



trial started with a drift check accompanied by the presentation of a full-screen scene image for 10 seconds. For 4 to 10-years-olds and the adults the scene was subsequently replaced by an image segment presented at the center of the screen; in 50% of the cases the image segment was valid. Participants had to determine whether the segment was part of the previous seen picture or not and respond by pressing the respective button. The segment was showed for 5 seconds or until a keyboard button was pressed (see Fig. 1A for an example trial). The segment recognition test was only included in the experiment to maintain the motivation of our participants. Participants older than 4 years of age completed 3 blocks, each consisted of 10 trials including the image segment recognition. The 2y-group performed only the free-viewing task and completed only 2 blocks of 10 trials (see Fig. 1B for an example trial). Scene images were shown in randomized order. Between the blocks, children were allowed to have a small break and watch a short animation film (duration of 2 minutes). Eye-movements were recorded during all trials. The total duration of the experiment, including the preparation for recording and calibration, was approximately 30 minutes per participant.

## **2.5. Data Analysis**

The data analyses were carried out using SPSS 21.0 and MATLAB 8.0. Only fixations and saccades starting after the image appeared and finishing before the image disappeared were included in the analysis. All fixations outside the presentation screen, fixations shorter than 90 ms and fixations around eyeblinks were discarded. The percentage of missing data per trial was calculated and trials with more than 40% of missing data were also discarded.

According to earlier findings (e.g. Velichkovsky, Dornhoefer, Pannasch, & Unema, 2000), fixations were anticipated to show a right skewed distribution where the median represents a more reliable value than the mean. Therefore, the median of each subject was calculated and used for the further analysis. To compare different groups the mean of medians was used.

The influence of stimuli saliency on fixation landings was also analyzed. The Graph-Based Visual Saliency (GBVS) algorithm was used to determine a saliency map for each image (Harel, Koch, & Perona, 2007). To determine the level of correct prediction of fixation landings, a receiver operating characteristics (ROC) analysis between eye-movements (provided as x,y fixation points) and the saliency maps for each image was performed in all subjects (e. g. Follet, Fontaine, & Meur, 2011; Hou, Harel, & Koch, 2011; Zhao & Koch,

2011). Then the mean of area under the curve (AUC) for each subject was calculated. A value of 0.5 indicates random performance, whereas 1.0 denotes a perfect prediction.

For the statistical testing the respective values were subjected to one-way analysis of variance (ANOVA), repeated measures ANOVA, and repeated measures multivariate analysis of variance (MANOVA). Bonferroni correction was used in the post-hoc analyses. Nonparametric Spearman correlation was used to determine effect of age in fixation duration and saccade amplitude.

### 3. Results

#### 3.1. Task performance

Task performance was analyzed for all participants (except for the 2y-group). The percentages of correct responses were entered into a one-way ANOVA and showed an effect of age group,  $F(3,78) = 22.15$ ,  $p < .001$ . Further post-hoc analyses revealed inferior performance for the 4-6y-group (71.5% of correct responses) compared with all other groups (6-8y-group,  $M = 83.3\%$ ; 8-10y-group,  $M = 89.5\%$ ; and adults-group,  $M = 93.9\%$ ), all  $p$ -values  $< .01$ . Furthermore, performance was significantly different between 6-8y-group and adults-group,  $p < .01$ .

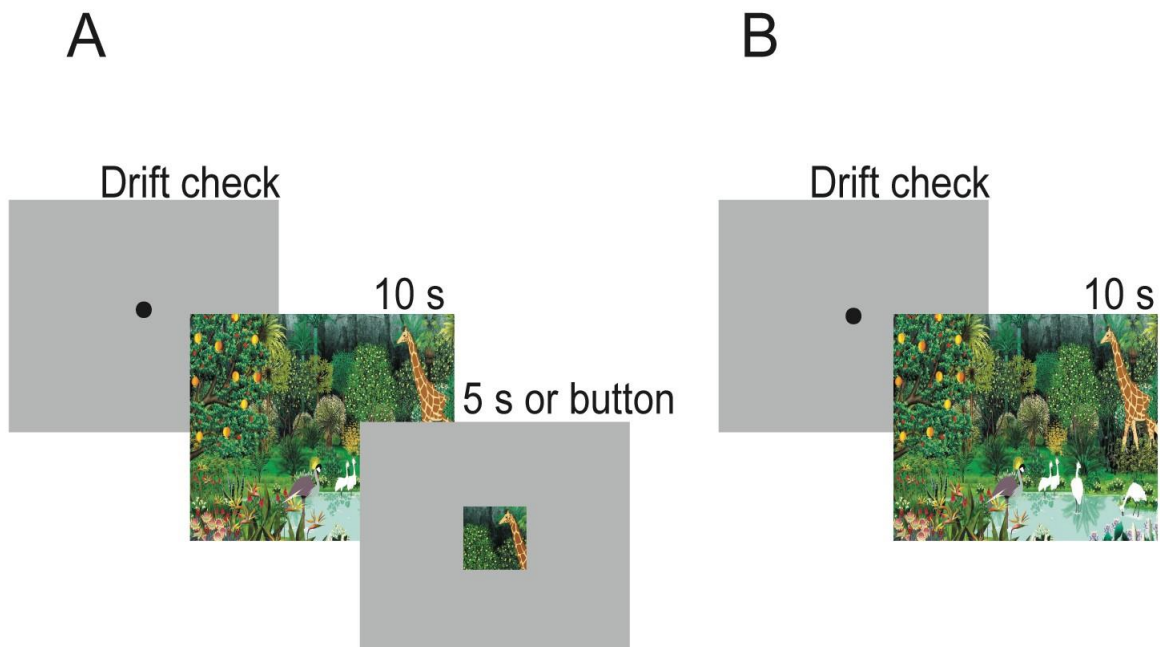


Figure 1. Example trial as it was presented to (A) the 4 to 10 years old participants and the adults and (B) 2 years old group.

### ***3.2. Fixation duration and saccade amplitude***

In order to determine whether general eye movement patterns changed with age, correlations between fixation duration, saccade amplitude and age were calculated. There was a negative correlation between fixation duration and age revealing shorter fixations with increasing age,  $r(78) = -0.74$ ,  $p < .001$ , whereas saccade amplitude was positively correlated with age,  $r(78) = 0.59$ ,  $p < .010$ . Furthermore, to determine with which age adult-like values were achieved a repeated measure MANOVA was conducted to compare fixation durations and saccadic amplitudes between groups. A significant multivariate effect was found for the main effect of group on both fixation durations and saccade amplitudes,  $F(2,8) = 23.65$ ,  $p < .001$ . Follow-up ANOVAs confirmed significant differences between the groups for fixation durations  $F(4,101) = 48.02$ ,  $p < .001$ , and saccade amplitudes,  $F(4,96) = 14.07$ ,  $p < .001$ . Post-hoc analyses indicated that fixations were longer in the 2y-group ( $M = 332$ ) and 4–6y-group ( $M = 277$ ) compared with the other age groups (6-8y-group,  $M = 252$ ; 8-10y-group,  $M = 247$ ; and adults,  $M = 233$ , all  $p$ -values  $< .05$  (see Figure 2A). Fixations of the 2y-group were also longer than those of 4-6y-group,  $p < .001$ . Saccade amplitudes were significantly shorter in the 2y-group,  $M = 3.4$  compared with other age groups (4-6y-group,  $M = 4.35$ ; 6-8y-group,  $M = 4.88$ ; 8-10y-group,  $M = 4.81$ ; and adults  $M = 4.46$ ), all  $p$ -values  $< .01$  (see Figure 2B).

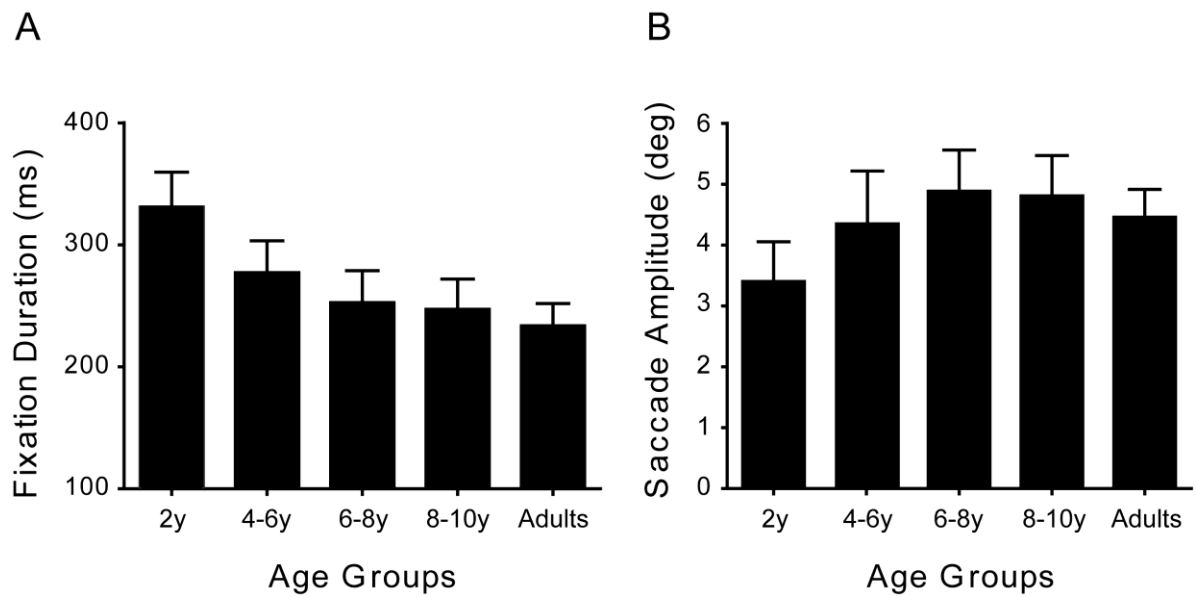


Figure 2. Fixation durations (A) and saccade amplitudes (B) for the different age groups. Error bars depict the standard deviation.

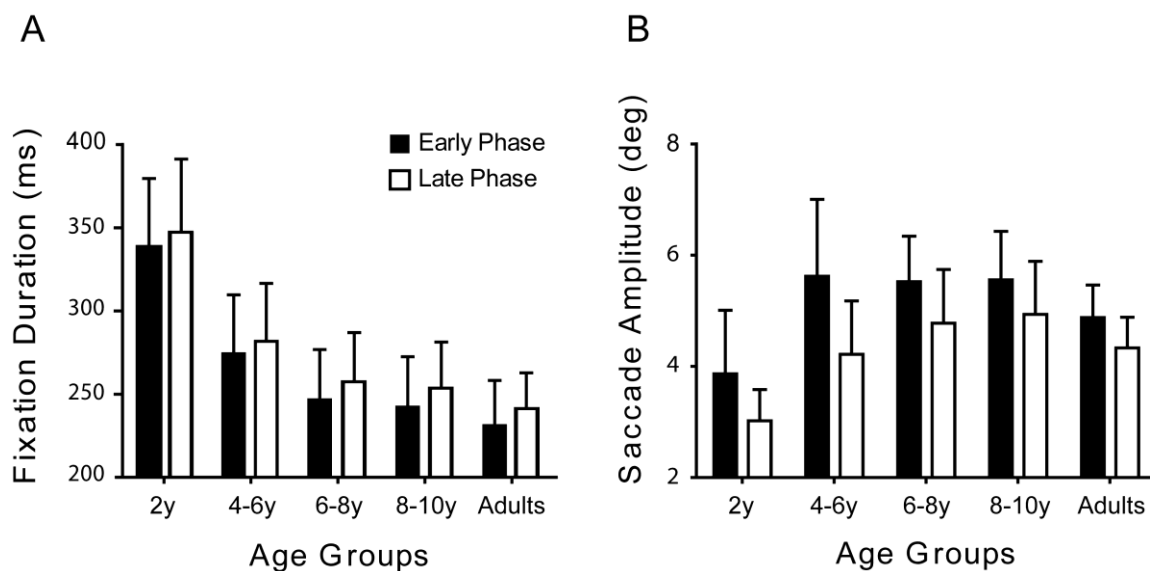
### 3.3. Time course of fixation durations and saccade amplitudes

Possible influences of time course on the gaze behavior were analyzed by comparing two different viewing phases. The early viewing phase comprised the time interval 0-2 seconds while the late viewing phase covered the 4-6 seconds time interval. For both viewing phases, we calculated the median for fixation duration and for saccade amplitude per subject; these individual medians were used for statistical testing. Fixation durations and saccade amplitudes were examined for differences along the time course by conducting a 5 (group) x

2 (viewing phase) repeated measures MANOVA. Significant effects were found for processing phase,  $F(2,8) = 24.4$ ,  $p < .001$ , and group,  $F(2,8) = 37.7$ ,  $p < .001$ , with no interaction,  $F < 1$ . A follow-up ANOVA for fixation durations revealed main effects for group,  $F(4,96) = 65.2$ ,  $p < .01$ , and for viewing phase  $F(1,96) = 4.7$ ,  $p < .05$ , with no interaction,  $F < 1$ . As expected, fixation durations were shorter in the early viewing phase (early:  $M = 264.4$ , late:  $M = 274.3$ ). Since there was no interaction between group and viewing phase, the increase in fixation durations from the early to the late phase can be considered as a stable effect across different age groups (2y-group,  $M_s = 337.7$  vs.  $347.3$ ; 4-6y-group,  $M_s = 274.3$  vs.  $281.8$ ; 6-8y-group,  $M_s = 246.6$  vs.  $257.4$ ; 8-10y-group,  $M_s = 242.2$  vs.  $253.6$ ; and adults  $M_s = 231.0$  vs.  $241.3$ ; see Figure 3A). However, post-hoc pairwise  $t$ -tests revealed

significant differences between the phases only for the 8-10y-group and in adults,  $p < .01$ .

The follow-up ANOVA for saccade amplitudes revealed main effects for group,  $F(4,96) = 23.5$ ,  $p < .001$ , and for processing phase,  $F(1,96) = 41.4$ ,  $p < .001$ , with no interaction,  $F = 1.5$ . Regarding the factor group, saccades were shortest for the 2y-group. Saccade amplitudes were longer during the early viewing phase (early:  $M = 5.1$ , late:  $M = 4.3$ ). Due to the absence of an interaction, the decrease in the length of saccades from the early to the late phase can be assumed as valid for all groups (2y-group,  $M_s = 3.9$  vs.  $3.0$ ; 4-6y-group,  $M_s = 5.6$  vs.  $4.2$ ; 6-8y-group,  $M_s = 5.5$  vs.  $4.8$ ; 8-10y-group,  $M_s = 5.6$  vs.  $4.9$ ; and adults  $M_s = 4.8$  vs.  $4.3$ ; Figure 3B). This was proved to be significant by conducting post-hoc pairwise  $t$ -tests (all  $p < .01$ , except 2-y-group,  $p = .01$ ).



*Figure 3.* Fixation durations (A) and saccade amplitudes (B) during early and late viewing phases for the different age groups. Error bars depict the standard deviation.

### **3.4. Proportion of “ambient” and “focal” fixations in each age group**

In order to know whether the proportion of ambient and focal fixations was different between the age groups, we categorized the fixations based on fixation duration and following saccade amplitude. Short fixations ( $< 180$  ms) followed by long saccades ( $> 5$  deg) were considered as ambient fixations, whereas long fixations ( $> 180$  ms) followed by short saccades ( $< 5$  deg) were considered as focal fixations (Unema et al., 2005). We considered the total of these two types of fixations as the 100% of the fixations for the analysis. To compare

the proportion of ambient and focal fixation between groups, a 5 (group) x 2 (fixation type) repeated measures ANOVA was conducted. Statistical testing revealed a significant main effect of fixation type,  $F(1,96) = 2039.98$ ,  $p < .001$ , but no effect of group,  $F < 1$ . Furthermore, there was a significant interaction between group and fixation type,  $F(4,96) = 29.54$ ,  $p < .001$ . All groups showed higher proportion of focal than ambient fixations. Post-hoc analyses revealed the largest proportion of focal fixations for the 2y-group (94.8); with increasing age, this proportion decreased (4-6y-group: 0.85, 6-8y-group: 0.81; 8-10 y-group: 0.78 and adults-group: 0.72; all  $p$ -values  $< .05$ ). The proportions furthermore differed between the 4-6y-group, 8-10y-group and the adults-group,  $p < .01$ , as well as between the 6- 8y-group and the adults-group,  $p < .01$ . These results show that proportion of ambient fixation increase and proportion of focal fixation decrease until age of 6 to 8 years old.

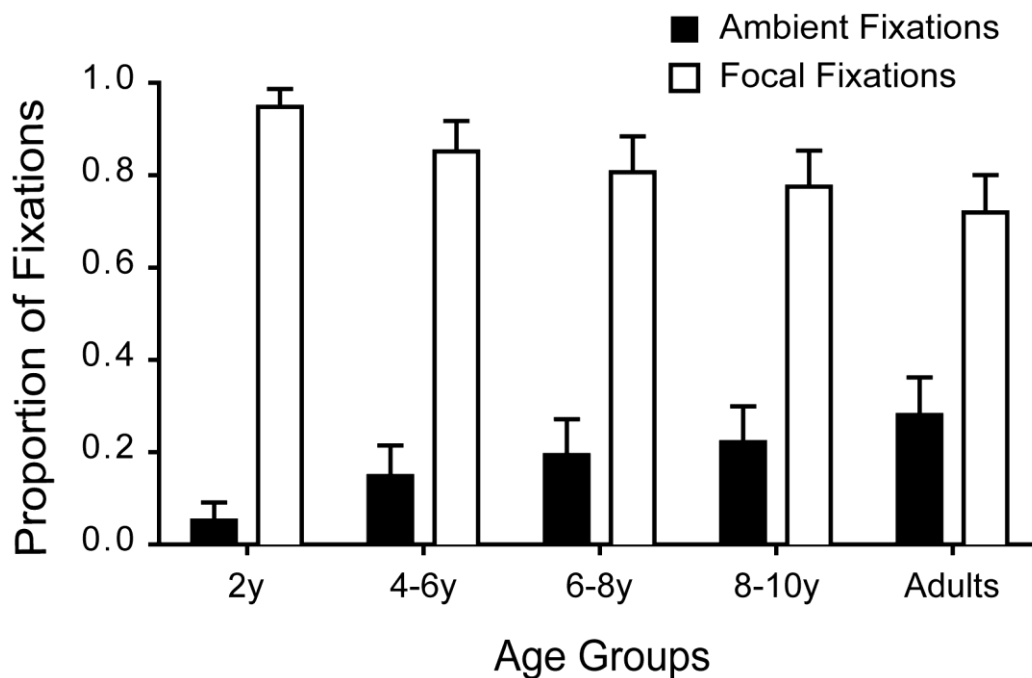
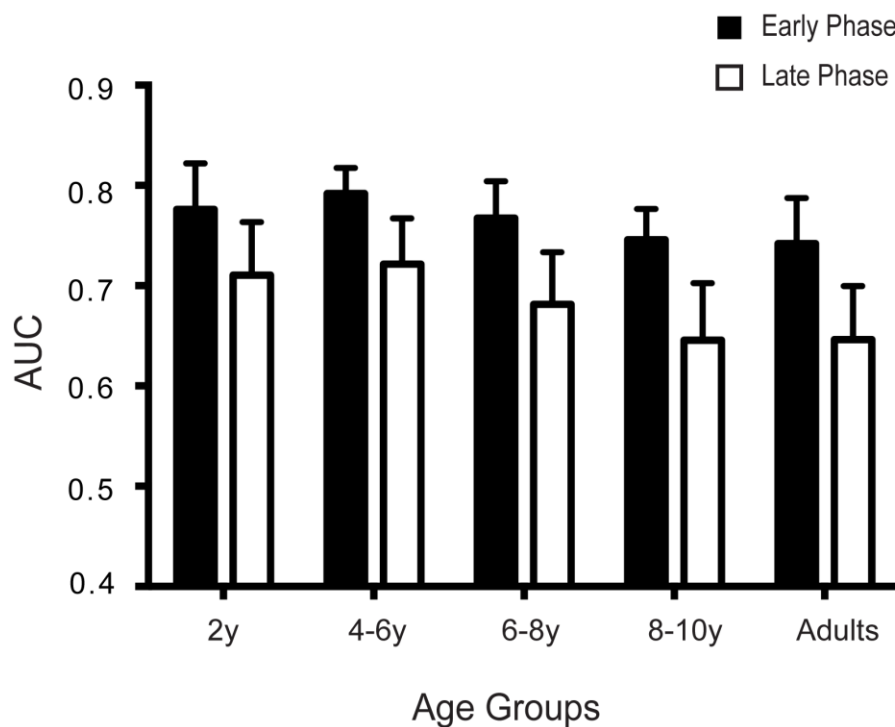


Figure 4. The proportion of ambient and focal fixations for the different age groups.

### 3.5. Effect of saliency on fixation landings

In order to determine the influence of low-level features on eye movement guidance in different age groups we compared predicted and recorded fixation distributions. Previous studies suggested that during the early viewing phase—dominated by the ambient mode— saliency aspects of images mainly influence eye movement behaviour

while during the late viewing phase—dominated by focal mode—top-down control becomes more important (Castelhano et al., 2009; Mills et al., 2011; Tatler & Vincent, 2008). Thus, we also tested whether there were differences between early and late viewing phases in the influences of saliency on fixations landings. Predicted fixation distributions were obtained by calculating the saliency map for each image using GBVS (Harel et al., 2007). We estimated the ROC to determine the level of correct predictions of fixation landings in each age group and each phase and we compared the prediction level between the different age groups and phases. To compare the fixation landing a 5 (group) x 2 (viewing phase) repeated measures ANOVA was conducted. Main effects of group,  $F(4,96) = 6.64$ ,  $p < .001$ , and viewing phase,  $F(4,96) = 247.67$ ,  $p < .001$ , were significant, with no interaction,  $F > 1$ . The post-hoc analyses indicated a better match between the predicted and observed data for the 2y-group and the 4-6y-group compared with the 8-10y-group and the adults-group  $p < .05$ . Therefore, the match between predicted and observed behaviour differed between group but this match is for all groups better for the early phase.



*Figure 5.* Mean of area under the curve (AUC) indicating the difference between computational saliency maps and eye movement behaviour by groups in early and late phases. Error bars depict the standard deviation.



#### *4. Discussion*

In the current study we compared eye-movement patterns of typically developing children of different ages (range 2-10 years) and healthy adults during the free visual exploration of naturalistic scenes. Our main aims were to ascertain whether ambient and focal modes are present during scene-viewing across this age range and to determine whether the influence of bottom-up factors such as salience in eye movement behaviour are similar in different age groups.

Our results show that fixation durations decrease while saccade lengths increase with age. Adult-like values were reached earlier for saccade amplitudes (4 to 6 years old) than for fixation durations (6 to 8 years old). Furthermore, we obtained shorter fixation durations and longer saccade amplitudes during the early viewing phase (first 2 seconds) compared with the late viewing phase (4 to 6 seconds). On a more detailed level, our data suggest that the ambient and focal processing modes are present already at 2 years of age. However, when the proportion of fixations was analysed, the number of focal fixations was significantly higher in younger children (2y-group and 4-6y-group) compared with all other age groups; adults-like values were achieved near to 8-10 years of age. Analyzing the influence of bottom-up factors on eye guidance at different ages revealed a stronger impact of saliency for younger children (2y-group and 4-6y-group) than for older children (8-10y-group) and adults. The further comparison of fixation landings between early and late phases revealed higher influence of salience during the early than the late phase in all age groups.

Saccade amplitudes were shorter in the 2y-group compared with the older groups. Saccade lengths that were comparable to adults were achieved in the 4-6y-group. Our results agree with previous findings where shorter saccade amplitudes and lower accuracy in pro- and anti-saccade tasks was observed for younger children in contrast to adults. However, in these tasks, it was shown that adult-like values are achieved only at 10 years of age (Fioravanti et al., 1995; Irving et al., 2006, 2009; Munoz et al., 1998). Conversely, we found longer fixations in younger children, up to 6 years of age. Unlike our results, fixation durations have been found to increase from 4 to 15 years of age when maintaining fixation. This suggests that the ability to maintain attention and to inhibit reflexive saccades continues to develop through childhood until adolescence (Aring et al., 2007; Ygge et al., 2005). In another study, where children from 4 to 16 years of age had to compare faces in a search task, it was found that gaze duration increased with age (Egami et al., 2009). However, these

tasks are clearly different from the procedure used in our study where subjects were free to explore the shown images; inhibiting saccades, maintaining fixations as well as comparative eye-movements were not required. Besides, it has been shown that in free viewing when 7 to 8-month old babies had to explore images of faces, their fixation durations were even longer compared to what we observed in our 2y-group (700 vs. 330 ms; Papageorgiou et al., 2013). These data support our results, suggesting that fixation durations in free viewing decrease progressively during development. There are at least two possible explanations for longer fixations in younger children. Firstly, it has been proposed that fixation durations are directly related to the difficulty of cognitive processes like information extraction (Chen, Meier, Blair, Watson, & Wood, 2013; Groner & Groner, 1989; Just & Carpenter, 1980) and recall in memory tasks (Velichkovsky et al., 2005). Additionally, it has been shown that infants who have shorter looking times are faster in the processing of visual stimuli when compared with same-aged infants with longer looking times (Colombo, Mitchell, Coldren, & Freeseaman, 1991; Sigman, Cohen, Beckwith, Asarnow, & Parmelee, 1991). Thus, younger children might have longer fixations because they require more time to process visual information due to their less developed cognitive resources. Additionally, saccade latencies are shown to decrease with increasing age (Fukushima et al., 2000; Irving et al., 2006; Klein & Foerster, 2001; Luna & Sweeney, 2004; Munoz et al., 1998). Accordingly, a second possible explanation is that younger children have longer fixations because it takes more time to them to program the next saccade.

It has been reported that eye-movement patterns are affected by task requirements (Castelhano et al., 2009; Fischer et al., 2013; Mills et al., 2011; Tatler & Vincent, 2008). Thus, differences in fixation durations and saccade amplitudes between groups might be also explained by differences in task demands. In our study, the 2y-group explored the images without any task, whereas the older participants performed an additional recognition task. With regard to the recognition performance, we found an increase in accuracy with age. Mills et al. (2011) investigated eye-movement parameters in adults during free viewing, memory and search tasks, and observed shortest saccades for the free viewing condition compared to other conditions of their experiment. Longer fixation durations were obtained for the comparison of free-viewing and search, while there were no differences between free-viewing and the memory task. According to these results, the shorter saccades and longer fixations for our youngest group might be also explained by differences in the task demands and not necessarily refer to developmental aspects of

eye movement control. However, due to the design of the present investigation, the results do not allow answering this question. Further work, varying systematically the difficulty, might provide detailed insights to test this assumption.

We furthermore explored fixation durations and saccade amplitudes within early and late phases of scene viewing and tested for age-related relationships. Although the differences were not statistically significant for all groups, we observed a robust trend with regard to changes related to the viewing-time: saccadic amplitudes decreased while fixation durations increased over time. The trend in this behavior was obtained already with the age of 2 years (see Fig. 3). This observation is in accordance with earlier findings for adult participants, suggesting a direct link between eye movement behavior and distinct modes of processing (Follet et al., 2011; Tatler & Vincent, 2008; Unema et al., 2005). Ambient mode has been related to preattentive scanning associated with the exploration of the spatial layout dominating the early phase of viewing time, whereas the focal mode has been related to attentive processing to objects identification, dominating later phases of scene exploration (Unema et al., 2005; Velichkovsky et al., 2005; Velichkovsky et al., 2002). Previous investigations in adults demonstrated the robust relationship between fixation duration, saccade amplitude and viewing phases for different visual stimuli (paintings, computer rendered scenes, photographs and emotional stimuli; Pannasch et al., 2008) as well as for different tasks (localization vs. identification tasks, Unema et al., 2005). The present observations are in line with these previous findings. We obtained an increase in fixation durations and a decrease in saccade amplitudes when there was no task given to children below 4 years of age as well as when a recognition task was provided for children older than 4 years. One critical issue here might be the definition of early and late viewing periods. We applied the same time windows to all of our participants. There is a possibility that across the different age groups the boundaries for early and late phases should rather be adjusted more individually. Future work should compare and analyse the time course on a more finely grained resolution, in order to confirm or reject the classification for the time course that was applied here. Nevertheless, we think that the present findings provide further evidence for the existence of the two distinct processing mechanisms, thereby indicating the relevance of this fundamental feature of our visual system.

The focal and ambient modes have been associated to the activity of ventral and dorsal visual pathways, respectively, and it has been suggested that some functions associated to the

ventral pathway might mature around the age of 6 years whereas those associated the dorsal pathway mature above 8 years of age (Braddick et al., 2003; Braddick & Atkinson, 2011; Gunn et al., 2002). Also, several previous studies have shown that until the age of 4 years, children exhibit a local bias in the processing of hierarchical visual forms, whereas by the age of 6-9 years, a global preference becomes visible (e. g. Dukette & Stiles, 1996, 2001; Poirel et al., 2008; Vinter et al., 2010). It has been reported that the shift from local to global preference is related to the anatomical maturation of the brain areas associated to the dorsal pathway (Poirel et al., 2011). Based on this information, it is reasonable to expect a focal bias in the ocular behavior of children younger than 8 years of age. Therefore, in order to find out whether these two attention modes are similarly present in all age groups, the fixations were classified as focal and ambient, based on fixation durations and the amplitude of the subsequent saccade. The results showed that two youngest groups had a higher proportion of focal fixations and lower proportion of ambient fixations compared with the older groups. Adult-like values were achieved near to 8-10 years old. This higher number of focal fixation type in younger children suggests a dominance of focal mode in children younger than 6 to 8 years of age. We suggest that the focal mode dominance in our younger groups (i.e. below 6 years of age) might be related to both the earlier maturation of ventral visual pathway (Braddick et al., 2003; Braddick & Atkinson, 2011; Gunn et al., 2002) and the tendency to process stimuli based on their local characteristics (e. g. Dukette & Stiles, 1996, 2001; Poirel et al., 2008; Vinter et al., 2010).

Our results furthermore show a better match between recorded gaze distribution and the distribution predicted by the saliency model for children at the age of 2 to 6 years than for older participants. These results correspond with previous reports (for reviews, see Henderson & Ferreira, 2004; Henderson, 2003) showing that eye guidance during scene perception is influenced by the combination of low-level features of the images and top-down control. Since the cognitive resources—that allow top-down control—are developing during childhood, (e.g. Gathercole, 1999; Gathercole, Pickering, Ambridge, & Wearing, 2004; Hitch, Halliday, Schaafstal, & Schraagen, 1988; Klenberg, Korkman, & Lahti-Nuuttila, 2010; Pearson & Lane, 1991; Pickering, 2001; Sanders, Stevens, Coch, & Neville, 2006) gaze behavior should be stronger influenced by low-level features at younger ages. Moreover, previous studies demonstrated that saliency aspects dominate the spatial distribution of fixations landing early in scene viewing. This form of control seems stronger associated with the ambient mode while top-down control becomes more dominant during

late phases (Castelhano et al., 2009; Mills et al., 2011; Tatler & Vincent, 2008). Our results also agree with these previous findings; saliency predictions revealed a better match with the actual landing positions of fixations during the early viewing phase and were less accurate during the late viewing phase. This was a global observation across all tested groups.

In summary, our results indicate distinct viewing patterns in scene perception before and after the age of 6-8 years. This suggests that eye movement control during scene viewing develops until age of 6-8 years, expressed by larger proportions of focal fixations and stronger low-level saliency influences. However, the relationship between saccade amplitude and fixation duration seems age-independent since indications for ambient and focal processing modes are already present at the age of 2 years. The present study provides a first overview about the relationship between gaze parameters in children of different ages, thereby revealing developmental differences. Additionally, this work provides a first approach to better understand the relationship between gaze behavior and attentional mechanisms from a developmental perspective. Future work should better control for certain aspects such as task difficulty and the precise time course in order to allow for a more detailed and complete picture about the developmental process.

### *Acknowledgments*

We express our gratitude to the infants and parents who participated to the study, we thank for their kindness and cooperation. We also thank Patrick Cavanagh, who supported this work by providing his EyeLink eye tracking system for this study. A.H. was supported by doctoral fellowship from CONICYT, Chile.

## *Annexe Study II*

In addition to previous analyses (published version) three additional parameters were examined in children. These new analyses will be discussed in the general discussion of the thesis.

*The durations of the first fixations to the scene and the scan path lengths were compared between age groups. The main effect of age group was found for both first fixations  $F(4,96) = 25.77, p < .001$  and scan path lengths  $F(4,96) = 37.25, p < .001$ . The 2y-group presented the longest first fixation durations ( $M = 321$  ms) compared with the other age groups, 4-6y-group also had longer fixation durations ( $M = 283$  ms) compared with older groups (6-8y-group,  $M = 248$  ms; 8-10y-group,  $M = 247$  ms; and adults  $M = 229$  ms). The 2y-group also presented the shortest scan path lengths ( $M = 66$  deg) compared with the other age groups and 4-6y-group had shorter scan path lengths ( $M = 122$  deg) compared with older groups (6-8y-group,  $M = 161$  deg; 8-10y-group,  $M = 151$  deg; and adults  $M = 154$  deg).*

*In addition, the proportion of fixation type was calculated including all fixation types (in the previous analysis only fixations that could be classified as ambient or focal were included). Statistical testing revealed a significant main effect of fixation type,  $F(1,96) = 894.51, p < .001$ , but no effect of group,  $F < 1$ . Furthermore, there was a significant interaction between age group and fixation type,  $F(4,96) = 10.17, p < .001$ . All groups showed higher proportion of focal than ambient fixations. Post-hoc analyses between groups revealed the lowest proportion of ambient fixations for the 2y-group (0.01); with increasing age, this proportion increased (4-6y-group: 0.04, 6-8y-group: 0.05; 8-10 y-group: 0.06 and adults-group: 0.07; all  $p$ -values  $< .05$ ). The proportions of 4-6y-group differed from 8-10y-group and the adults-group,  $p < .05$ . No differences between groups were observed for focal fixations  $p > .05$ . These results showed that when the whole population of fixations were considered, only ambient fixations increased by age, stabilizing by 6 to 8 years of age.*

## References

- Aring, E., Grönlund, M. A., Hellström, A., & Ygge, J. (2007). Visual fixation development in children. *Graefe's archive for clinical and experimental ophthalmology*, 245(11), 1659–1665. doi:10.1007/s00417-007-0585-6
- Armstrong, V., Maurer, D., & Lewis, T. L. (2009). Sensitivity to first- and second-order motion and form in children and adults. *Vision Research*, 49(23), 2774–2781. doi:10.1016/j.visres.2009.08.016
- Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research*, 51(13), 1588–1609. doi:10.1016/j.visres.2011.02.018
- Braddick, O., Atkinson, J., & Wattam-Bell, J. (2003). Normal and anomalous development of visual motion processing: motion coherence and “dorsal-stream vulnerability.” *Neuropsychologia*, 41(13), 1769–1784. doi:10.1016/S0028-3932(03)00178-7
- 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. doi:10.1167/9.3.6.
- Chandna, a. (1991). Natural history of the development of visual acuity in infants. *Eye*, 5, 20–26. doi:10.1038/eye.1991.4
- Chen, L., Meier, K. M., Blair, M. R., Watson, M. R., & Wood, M. J. (2013). Temporal characteristics of overt attentional behavior during category learning. *Attention, Perception & Psychophysics*, 75(2), 244–256. doi:10.3758/s13414-012-0395-8
- Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseaman, L. J. (1991). Individual Differences in Infant Visual Attention: Are Short Lookers Faster Processors or Feature Processors? *Child Development*, 62(6), 1247–1257. doi:10.1111/j.1467-8624.1991.tb01603.x
- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, 58(3), 306–324. doi:10.1016/j.neuron.2008.04.017
- Dekker, T., Mareschal, D., Sereno, M. I., & Johnson, M. H. (2011). Dorsal and ventral stream activation and object recognition performance in school-age children. *NeuroImage*, 57(3), 659–670. doi:10.1016/j.neuroimage.2010.11.005
- Dukette, D., & Stiles, J. (1996). Children's analysis of hierarchical patterns: Evidence from a similarity judgment task. *Journal of Experimental Child Psychology*, 63(1), 103–140. doi:10.1006/jecp.1996.0044



- Dukette, D., & Stiles, J. (2001). The effects of stimulus density on children's analysis of hierarchical patterns. *Developmental Science*, 4(2), 233–251. doi:10.1111/1467-7687.00168
- Egami, C., Morita, K., Ohya, T., Ishii, Y., Yamashita, Y., & Matsuishi, T. (2009). Developmental characteristics of visual cognitive function during childhood according to exploratory eye movements. *Brain & Development*, 31(10), 750–757. doi:10.1016/j.braindev.2008.12.002
- Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, 22(4), 661–674. doi:10.1017/S0140525X99002150
- Fioravanti, F., Inchingolo, P., Pensiero, S., & Spanio, M. (1995). Saccadic eye movement conjugation in children. *Vision Research*, 35(23-24), 3217–3228. doi:10.1016/0042-6989(95)00152-5
- 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. doi:10.3389/fnsys.2013.00017
- Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia*, 47(1), 248–257. doi:10.1016/j.neuropsychologia.2008.07.016
- Follet, B., Le Meur, O., & Baccino, T. (2011). New insights into ambient and focal visual fixations using an automatic classification algorithm. *i-Perception*, 2(6), 592–610. doi:10.1068/i0414
- Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements. *Brain and Development*, 22(3), 173–180. doi:10.1016/S0387-7604(00)00101-7
- Gathercole, S. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3(11), 410–419. doi:10.1016/S1364-6613(99)01388-1
- Gathercole, S., Pickering, S., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, 40(2), 177–190. doi:10.1037/0012-1649.40.2.177
- Gordon, G. E., & McCulloch, D. L. (1999). A VEP investigation of parallel visual pathway development in primary school age children. *Documenta Ophthalmologica*, 99(1), 1–10.
- Groner, R., & Groner, M. T. (1989). Attention and eye movement control: An overview.

*European Archives of Psychiatry and Neurological Sciences*, 239(1), 9–16.  
doi:10.1007/BF01739737

- Gunn, A., Cory, E., Atkinson, J., Braddick, O., Wattam-Bell, J., Guzzetta, A., & Cioni, G. (2002). Dorsal and ventral stream sensitivity in normal development and hemiplegia. *Neuroreport*, 13(6), 843–847.
- Harel, J., Koch, C., & Perona, P. (2007). Graph-Based Visual Saliency. In B. ChoLkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems* (pp. 545–552). Cambridge, MA: MIT Press.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504. doi:10.1016/j.tics.2003.09.006
- Henderson, J. M. (2007). Regarding Scenes. *Current Directions in Psychological Science*, 16(4), 219–222. doi:10.1111/j.1467-8721.2007.00507.x
- Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (Psychology., pp. 1–58). New York, NY: Psychology Press.
- 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. doi:10.3758/BF03213479
- Hou, X., Harel, J., & Koch, C. (2011). Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1), 194–201. doi:10.1109/TPAMI.2011.146
- Irving, E. L., Steinbach, M. J., Lillakas, L., Babu, R. J., & Hutchings, N. (2006). Horizontal saccade dynamics across the human life span. *Investigative Ophthalmology & Visual Science*, 47(6), 2478–2484. doi:10.1167/iovs.05-1311
- Irving, E. L., Tajik-Parvinchi, D. J., Lillakas, L., González, E. G., & Steinbach, M. J. (2009). Mixed pro and antisaccade performance in children and adults. *Brain Research*, 1255, 67–74. doi:10.1016/j.brainres.2008.12.006
- 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. doi:10.1016/S0042-6989(99)00163-7
- Just, M. a, & Carpenter, P. a. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. doi:10.1037/0033-295X.87.4.329
- Klaver, P., Lichtensteiger, J., Bucher, K., Dietrich, T., Loenneker, T., & Martin, E. (2008).

- Dorsal stream development in motion and structure-from-motion perception. *NeuroImage*, 39(4), 1815–1823. doi:10.1016/j.neuroimage.2007.11.009
- Klein, C., & Foerster, F. (2001). Development of prosaccade and antisaccade task performance in participants aged 6 to 26 years. *Psychophysiology*, 38(2), 179–189. doi:10.1017/S0048577201981399
- Klenberg, L., Korkman, M., & Lahti-Nuutila, P. (2010). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology*, 20(1), 407–428. doi:10.1207/S15326942DN2001
- Koldewyn, K., Whitney, D., & Rivera, S. M. (2011). Neural correlates of coherent and biological motion perception in autism. *Developmental Science*, 14(5), 1075–1088. doi:10.1111/j.1467-7687.2011.01058.x
- Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, 75(5), 1357–1372. doi:10.1111/j.1467-8624.2004.00745.x
- Luna, B., & Sweeney, J. A. (2004). The emergence of collaborative brain function: FMRI studies of the development of response inhibition. *Annals of the New York Academy of Sciences*, 1021, 296–309. doi:10.1196/annals.1308.035
- Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, 68(3), 293–308. doi:10.1016/j.bandc.2008.08.019
- Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3), 229–240. doi:10.1038/nrn1348
- 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. doi:10.1167/11.8.17.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3), 774–785. doi:10.1016/j.neuropsychologia.2007.10.005
- 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. doi:10.1007/s002210050473
- Oliva, A. (2005). Gist of the Scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (Vol. 696, pp. 251–257). San Diego, CA: Elsevier.
- Pannasch, S., Helmert, J. R., Roth, K., & Walter, H. (2008). Visual fixation durations and

- saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2), 1–19.
- Pannasch, S., & Velichkovsky, B. M. (2009). Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. *Visual Cognition*, 17(6-7), 1109–1131. doi:10.1080/13506280902764422
- Papageorgiou, K., Smith, T., Wu, R., Kirkham, N., Johnson, M., & Ronald, A. (2013). An investigation of the relationship between individual differences in infant fixation durations and later temperament and behaviour in childhood. *17th EUROPEAN CONFERENCE ON EYE MOVEMENTS Book of abstracts*.
- Parrish, E. E., Giaschi, D. E., Boden, C., & Dougherty, R. (2005). The maturation of form and motion perception in school age children. *Vision Research*, 45(7), 827–837. doi:10.1016/j.visres.2004.10.005
- Paus, T., Babenko, V., & Radil, T. (1990). Development of an ability to maintain verbally instructed central gaze fixation studied in 8- to 10-year-old children. *International Journal of Psychophysiology*, 10(1), 53–61. doi:10.1016/0167-8760(90)90045-F
- Pearson, D. a, & Lane, D. M. (1991). Auditory attention switching: A developmental study. *Journal of Experimental Child Psychology*, 51(2), 320–334. doi:10.1016/0022-0965(91)90039-U
- Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory*, 9(4), 423–432. doi:10.1080/09658210143000182
- Poirel, N., Mellet, E., Houdé, O., & Pineau, A. (2008). First came the trees, then the forest: developmental changes during childhood in the processing of visual local-global patterns according to the meaningfulness of the stimuli. *Developmental psychology*, 44(1), 245–253. doi:10.1037/0012-1649.44.1.245
- Poirel, N., Simon, G., Cassotti, M., Leroux, G., Perchey, G., Lanoë, C., ... Houdé, O. (2011). The shift from local to global visual processing in 6-year-old children is associated with grey matter loss. *Plos one*, 6(6), e20879. doi:10.1371/journal.pone.0020879
- Rayner, K. (2009). *Eye movements and attention in reading, scene perception, and visual search. Quarterly journal of experimental psychology (2006)* (Vol. 62, pp. 1457–1506). doi:10.1080/17470210902816461
- Riby, D. M., & Hancock, P. J. B. (2008). Viewing it differently: social scene perception in Williams syndrome and autism. *Neuropsychologia*, 46(11), 2855–2860. doi:10.1016/j.neuropsychologia.2008.05.003
- Riby, D. M., & Hancock, P. J. B. (2009). Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. *Journal of intellectual disability*

research : *JIDR*, 53(2), 169–181. doi:10.1111/j.1365-2788.2008.01142.x

Ridder, W. H., Borsting, E., & Banton, T. (2001). All Developmental Dyslexic Subtypes Display an Elevated Motion Coherence Threshold. *Optometry and Vision Science*, 78(7), 510–517. doi:10.1097/00006324-200107000-00014

Roucoux, A., Culee, C., & Roucoux, M. (1983). Development of fixation and pursuit eye movements in human infants. *Behavioural Brain Research*, 10(1), 133–139. doi:10.1016/0166-4328(83)90159-6

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. doi:10.1016/j.neuropsychologia.2005.10.007

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception : A selective review, 11, 1–30. doi:10.1167/11.5.9.

Sigman, M., Cohen, S. E., Beckwith, L., Asarnow, R., & Parmelee, A. H. (1991). Continuity in cognitive abilities from infancy to 12 years of age. *Cognitive Development*, 6(1), 47– 57. doi:10.1016/0885-2014(91)90005-X

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, 45(5), 643–659. doi:10.1016/j.visres.2004.09.017

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2), 1–18.

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K.-P. (2002). Neural mechanisms of saccadic suppression. *Science*, 295(5564), 2460–2462. doi:10.1126/science.1068788

Torralba, A., Oliva, A., Castelano, 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. doi:10.1037/0033-295X.113.4.766

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, 31(4), 299–337. doi:10.1007/BF00422717

Unema, P., Pannasch, S., Joos, M., & Velichkovsky, B. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473–494. doi:10.1080/13506280444000409

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements : Evidence from static and dynamic scene

- perception. In M. Bara, B. G. Barsalou, L. & Bucciarelli (Ed.), *Proceedings of the XXVII Conference of the Cognitive Science Society* (pp. 2283–2288). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. (M. Bara, B. G. Barsalou, L. & Bucciarelli, Ed.) *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 145–156. doi:10.1016/S1369-8478(02)00013-X
- Vinter, A., Puspitawati, I., & Witt, A. (2010). Children's spatial analysis of hierarchical patterns: construction and perception. *Developmental Psychology*, *46*(6), 1621–1631. doi:10.1037/a0020615
- Volkman, F. C. (1986). Human visual suppression. *Vision Research*, *26*(9), 1401–1416. doi:10.1016/0042-6989(86)90164-1
- Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology*, *6*(4), 632–650. doi:10.1016/0022-0965(68)90108-2
- Ygge, J., Aring, E., Han, Y., Bolzani, R., & Hellström, A. (2005). Fixation stability in normal children. *Annals of the New York Academy of Sciences*, *1039*, 480–483. doi:10.1196/annals.1325.049
- Zhao, Q., & Koch, C. (2011). Learning a saliency map using fixated locations in natural scenes. *Journal of Vision*, *11*, 1–15. doi:10.1167/11.3.9.

*Study III: In preparation*

*Influence of semantic consistency and perceptual saliency on visual attention during scene viewing in 2-year-olds*

Helo, Andrea<sup>1,2</sup>, Danteney-Dordoigne, Luchile<sup>1</sup>, Pannasch, Sebastian<sup>3</sup>, & Rämä, Pia<sup>4</sup>

<sup>1</sup>Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France.

<sup>2</sup>Departamento de Fonoaudiología, Universidad de Chile, Santiago, Chile.

<sup>3</sup>Department of Psychology, Engineering Psychology and Applied Cognitive Research, Technische Universität Dresden, Germany

<sup>4</sup>CNRS (UMR 8158), Paris, France.

Corresponding author: Andrea Helo

Address: Laboratoire Psychologie de la Perception (CNRS UMR 8158),  
Université Paris Descartes, 45, rue des Saints-Pères, 75006 Paris, France.

Fax: +33 (0)1 42 86 33 22

Email: ahelo@med.uchile.cl

## *Abstract*

Conceptual representations of everyday scenes are built in interaction with visual environment and these representations guide our visual attention. Earlier, it has been shown that objects that are either visually salient or inconsistent with the scene context attract our attention. The aim of the present study was to examine to what extent semantic violations of different saliency attract visual attention in 24-month-old toddlers. We analyzed eye movements in 24-month-old toddlers and adults during the exploration of everyday scenes containing an object that was either inconsistent (e.g., soap on a breakfast table) or consistent (e.g., soap in a bathroom) with the; these objects could either be of high or low saliency. We found that toddlers as well as adults looked longer at inconsistent and at salient objects. However, high-salient objects attracted the gaze earlier than low-salient objects only in the toddler group. Additionally, toddlers with lower vocabulary skills presented a larger consistency effect than toddlers with higher vocabulary skills whereas both groups were equally attracted by saliency. Our results indicate that 24-month-old children use scene context to guide visual attention when exploring visual environment. However, perceptual features have a stronger influence in the guidance of attention 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.

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



## *1. Introduction*

Our everyday visual environment is predictable even if particular aspects may vary from one situation to another. For instance, objects are most likely to appear in certain contexts (e.g., an oven in a kitchen) but their position varies depending on the layout of the scenes (e.g., two different kitchens). In this regard, a natural scene can be defined as a human- scaled view of the real world with a coherent semantic configuration (Henderson & Ferreira, 2004; Hollingworth & Henderson, 1999). The semantic configuration refers to the probability of an object to appear in a scene as well as to the relationship between the objects of a scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Henderson & Ferreira, 2004; Henderson, Weeks, & Hollingworth, 1999). Early literature proposed that based on experience with visual environment configurations, a “scene knowledge” is built and stored in the long-term memory (Barlett, 1932; Hock, Romanski, Galie, & Williams, 1978; Mandler & Johnson, 1976; Potter, 1975). This scene knowledge allows viewers to quickly extract the global meaning of a scene, i.e. the so-called gist. After the extraction of the gist, viewers generate a set of expectations about the type of objects that could be contained and also where they should be located within a scene (Biederman et al., 1982; Hock, Romanski, Galie, & Williams, 1978; Mandler & Johnson, 1976; Oliva, 2005; Potter, 1976). These expectations guide further visual exploration. Previous research indicates that attentional allocation during scene viewing – as reflected by fixation locations (Yarbus, 1967) – is determined by both informativeness and perceptual image features such as color, contrast, and intensity (Oliva, 2005; Tatler & Vincent, 2008). Therefore, eye fixations during scene processing often are directed to objects that are either semantically relevant and/or visually salient within the scene. Furthermore, when viewers’ expectations are violated by an inconsistent object within a scene fixations are directed to this object (Henderson et al., 1999; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Vö & Henderson, 2009).

To date, there is no complete consensus on whether semantic inconsistencies guide the eye movements before the object is fixated and how perceptual (e.g., saliency) and cognitive (e.g. scene knowledge) processes interact during scene viewing. While some studies argue that semantic inconsistencies are detected rapidly, i.e. within the first 200 ms, influencing early eye movements before the fixation of the incongruent object (Becker, Pashler, & Lubin, 2007; Loftus & Mackworth, 1978; Underwood, Humphreys, & Cross, 2007; Geoffrey Underwood & Foulsham, 2006; Geoffrey Underwood, Templeman, Lamming, & Foulsham,

2008); others claim that in order to be detected, incongruent objects need to be fixated (De Graef, Christiaens, & D'Ydewalle, 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Vö & Henderson, 2009; Vö & Henderson, 2011). Regarding perceptual and cognitive interactions, it has been shown that visual attention might be more perceptual or more semantically guided depending on the task goals (Einhäuser, Rutishauser, & Koch, 2008; Hayhoe & Ballard, 2005). Particularly, saliency has shown to be a good predictor of early gaze distribution – predicting up to the first five fixations better than chance – during free viewing (Parkhurst, Law, & Niebur, 2002) or memory tasks (Underwood & Foulsham, 2006). Conversely, during visual searching tasks, where visual attention is cued to a specific target, saliency has shown to be a poor predictor (Henderson et al., 2007). Based on these and other findings it has been proposed that the strength of bottom-up saliency contribution is modulated by top-down control (Einhäuser et al., 2008; Parkhurst et al., 2002; Theeuwes, 2010; Treue, 2003). However, the amount of contribution of each component on gaze guidance is still under debate. One line of evidence claims a stronger effect of saliency on gaze allocation (Itti & Koch, 2001; Zelinsky, 2008) while the other argues that gaze distribution is mainly directed by cognitive factors (Hwang, Wang, & Pomplun, 2011; Torralba, Oliva, Castelhana, & Henderson, 2006).

Cognitive resources (e.g. Gathercole, 1999; Gathercole, Pickering, Ambridge, & Wearing, 2004; Hitch, Halliday, Schaafstal, & Schraagen, 1988; Klenberg, Korkman, & Lahti-Nuutila, 2010; Pearson & Lane, 1991; Pickering, 2001; Sanders, Stevens, Coch, & Neville, 2006) as well as eye movement control (Luna, 2008; Karatekin 2007; Açik, Sarwary, Schultze-Kraft, Onat, & König, 2010; Helo, Pannasch, Sirri, & Rämä, 2014) are developing during childhood. Therefore, it is plausible that the interaction between cognitive (e.g., semantic) and perceptual (e.g., saliency) gaze guidance is different in young children and adults. Semantic scene knowledge is built through visual experience (Barlett, 1932; Hock et al., 1978; Mandler & Johnson, 1976), and thus, it is likely that semantic guidance increases with age. However, to date there is little evidence about semantic scene processing development (Bornstein, Arterberry, & Mash, 2010; Bornstein, Mash, & Arterberry, 2011a, 2011b; Duh & Wang, 2014; Hock et al., 1978). In contrast, there is more evidence about development of eye movement control during natural scene exploration. Studies in children have shown that they tend to be less explorative (i.e. to present less scattered distribution of eye fixations within scenes) than adults (Açik, Sarwary, Schultze-Kraft, Onat, & König, 2010). Likewise, saliency has been shown to guide eye movements to a larger extent in younger than older children and adults (Açik et al., 2010;

Helo et al., 2014). Particularly, we showed that children younger than 6 years of age were more attracted to salient areas of images than eight to ten-year-olds and adults (Helo et al., 2014). These findings indicate that gaze allocation is preferentially perceptual-guided in young children. To our knowledge, however, only one study conducted on 15-month-olds has investigated how perceptual and semantic factors interact during natural scene viewing in infants. Duh and Wang (2014) reported that children looked longer to the screen after a short exposure (500ms) to a scene when a perceptually salient change that preserved the scene gist occurred than when no change was presented. However, they missed the salient change and looked equally long the screen during a change and a no-change trial in a longer exposure (3000ms). In a second experiment (using only the longer exposure time), children looked longer at an image illustrating a change that disrupted the scene meaning (e.g., a beach umbrella replaced by a table) than an image presenting a perceptually salient change that preserved the gist (e.g., a beach umbrella replaced by a colorful beach umbrella). These results suggest that infants take in account both low-level image features and semantic properties when processing visual scenes by the age of 15 months. Low-level features were processed faster than semantic properties, but semantic properties overwrote perceptual features when children were given more time to extract the gist of the scene (Duh & Wang, 2014). Further empirical evidence using free exploration tasks is, however, needed to better understand interactions between eye movement guidance based on perceptual features and semantic properties in young children.

It is well known that language and visual processes are closely related (Ferreira & Tanenhaus, 2007; Henderson & Ferreira, 2004). Eye movements are often directed towards objects that have been referred to in speech (Eberhard, Spivey-Kowlton, Sedivy, & Tanenhaus, 1995; Tanenhaus, Spivey-Kowlton, Eberhard, & Sedivy, 1995) and linguistic inputs can affect visual perception in searching and motion-detection task (Meteyard, Bahrami, & Vigliocco, 2007; Spivey, Tyler, Eberhard, & Tanenhaus, 2001). Correspondingly, visual context affects online comprehension of spoken sentences and visual information facilitates the resolution of ambiguities in spoken language (Allopenna, Magnuson, & Tanenhaus, 1998; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Kowlton, Eberhard, & Sedivy, 1995, Trueswell, Tanenhaus, & Garnsey, 1994). In addition, when participants are asked to name seen objects during their scene exploration, these two processes, naming and viewing, influence each other (Clarke, Coco, & Keller, 2013; Coco, Malcolm, & Keller, 2014). Particularly, fixation landings and perceptual properties of an object have shown to enhance its probability of naming. At the same time,

the fixation distribution is affected by linguistic parameters 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).

It is also known that language processing is automatically activated during exploration of visual displays and implicit naming can influence gaze allocation even when linguistic information is not provided (Matlock & Richardson, 2004; Richardson & Matlock, 2007, Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus et al., 1995, Görge et al., 2013; Meyer et al., 2007; Chabal & Marian, 2015). There is recent evidence indicating that also children use internal verbalization already during early language acquisition (Khan, 2013; Mani & Plunkett, 2010, 2011). In studies by Mani and Plunkett (2010, 2011), children at 18 and 24 months of age were presented with unlabelled images (e.g., cup) following by a pair of images containing a phonologically similar target (e.g., cat) and unrelated distractor (e.g. shoe). The result showed that looking times of the named target was affected by the prime image even it was never named. These findings indicate that children implicitly activated the label of the prime image. Spontaneous labelling of objects has been also shown in 2 year-old children when playing freely with a set of novel objects (Samuelson & Smith, 2005). Moreover, there is evidence that children at the age of 24-month-olds are more likely to silently name objects than adults (Khan, 2013). Interestingly, toddlers seem to label images even when linguistic input is not provided (Khan, 2013) while adults label images only in linguistic tasks (Meyer & Damian, 2007; Morsella & Miozzo, 2002; Telling, 2008; see also Zelinsky & Murphy, 2000, Kahn et al., 2013) tasks.

In the current study, our first aim was to examine the influence of saliency and semantic inconsistencies on attention allocation in 24-month-old children during a free exploration of visual scenes. Our secondary aim was to investigate whether linguistic skills (i.e., as measured by productive vocabulary size) contribute to attention allocation during scene exploration. A group of 24-month-old children and a group of adults were presented with familiar visual scenes (e.g., kitchens, bedrooms) while their eye movements were tracked. Half of the scenes displayed an inconsistent target object with the scene context (e.g., soap on a breakfast table), while the other half contained consistent target object with the scene context. (e.g., soap in a bathroom). To examine the effect of saliency and semantic inconsistencies on eye movement guidance we compared extrafoveal (e.g., saccade latency and index of the first fixation to the target) and foveal (e.g., dwell time, amount of fixations and gaze durations) measures for consistent and inconsistent objects having

high or low saliencies within the scenes. Based on previous findings on the development of visual attention (Helo, Rämä, Pannasch, & Meary, in press) and semantic scene processing (Duh & Wang, 2014) during early infancy and childhood, we expected that 24-month-olds are able to use scene context for their eye movement guidance during scene exploration, and thus, their visual attention is allocated to semantic inconsistencies. However, since semantic processing is still in development during early childhood, and earlier findings have shown a stronger effect of saliency on scene viewing in younger than in older children (Açık et al., 2010; Helo et al., 2014), we expected a dominance of saliency on gaze allocation in toddlers compared with adults. To investigate whether linguistic skills contribute to the allocation of attention during scene exploration, we compared gaze distribution in children with higher productive vocabulary skills to those of lower vocabulary skills. We hypothesized that vocabulary skills, previously associated with implicit naming (Samuelson & Smith, 2005) would influence the allocation of visual attention. More particularly, we expected that children with higher vocabulary skills use linguistic, besides perceptual and semantic cues, in the selection of attention location, thus increasing looking times on consistent objects compared with low producers.

## ***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 insufficient number of trials, and thus, the result of 42 toddlers (21 girls, mean age 24 months, range 23-25 months) were analyzed for this study. Children were recruited from a database of parents who agreed to volunteer in child development studies, and came from diverse socio-economic 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 the set 2. Ages and number of boys and girls were the same in both sets (9 girls set 1, 12 girls set 2; mean age 24 months, range 23-25 in both sets). Likewise, there was no difference in the number of produced words between children viewing the images in set 1 ( $n=19$ , 241 words) and set 2 ( $n = 15$ , 270 words),  $t(31)=-0.59$ ,  $p>.05$ . We also compared the number of targets

object the participants were able to produce (based on the questionnaire filled by the parents,  $n=40$ , images in set 1 ( $n=19,5$ , 6 words) and set 2 ( $n=15$ , 5,9 words), and no differences were found between sets,  $t(38) = -1.47$ ,  $p > .05$ . All adult participants had normal or corrected to normal vision, and they lacked any 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  $0.1^\circ$ , with a minimum velocity of  $30^\circ\text{s}^{-1}$  and a minimum acceleration of  $8000^\circ\text{s}^{-2}$ , maintained for at least 4ms. 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  $1024 \times 728$  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 had a resolution of  $1024 \times 768$  pixels and 24-bit color depth. Scenes were photographed using a Nikon D5100 camera. Four scene types illustrated home interior scenes (kitchens, bathrooms, bedrooms or living rooms). The images represented typical examples of Parisian indoor sceneries taking from five different homes. Each scene was presented in an either semantically consistent or semantically inconsistent version. Semantically inconsistent scenes included a target object inconsistent with the context of the scene (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

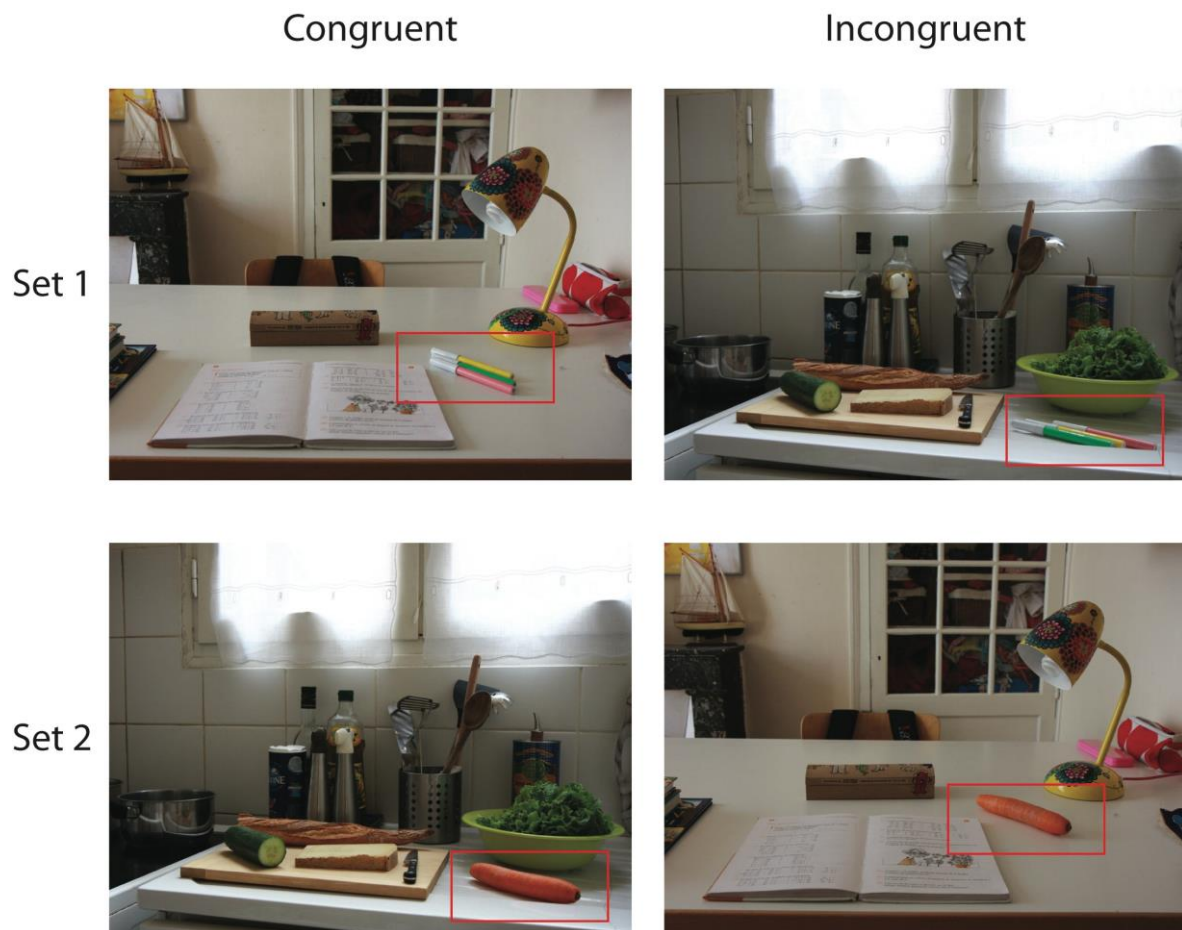
twice, both in a consistent and in an inconsistent scene context (see, Fig. 1). In a particular scene, the objects in the consistent and inconsistent scenes were of similar size. Objects locations were proportionately distributed across the four quadrants of a scene. An area of interest (AOI) was defined for each inconsistent object (Inconsistent AOI) and its control object (Consistent AOI). The mean area of AOIs was 64 cm<sup>2</sup>. 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<sup>2</sup>. 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.

The semantic congruency was tested in a pilot study with 16 adult participants who viewed the scenes but did not participate in the main experiment. In the pilot study, each participant saw all 36 images (i.e. both sets) and they were asked to rate the coherence of each scene using the scale from 1 to 9 (1 was highly coherent and 9 was highly incoherent). The ratings differed significantly between coherent and incoherent scenes,  $F(1,16) = 352.07$ ,  $p < .001$ , suggesting that there was no ambiguity between coherent and incoherent scenes (set 1:  $M = 1.1$  for consistent images and  $M = 5.96$  for inconsistent images; set 2:  $M = 1.11$  for consistent images and  $M = 5.96$  for inconsistent images). Ratings did not differ between the images sets,  $F < 1$ .

Saliency was controlled for semantically consistent and inconsistent scenes and also for sets. Using a MATLAB Saliency Toolbox (Walther & Koch, 2006). This tool creates a saliency map of the image combining features such as intensity, color and orientation of the visual images and establishes a saliency level for each region in an image. To study whether object saliency within the scenes influenced the looking times, we classified the scenes based on saliency of the AOIs. Based on this saliency map consistent and inconsistent AOIs for each image were ranked from 1 to 7 depending on its saliency level within the scene (1 being the most salient). An AOI that obtained first to third rank was considered as representing a “high salient AOI” whereas an AOI with a fourth or higher rank were considered as a “low salient AOI”. Using this classification, twelve scenes were categorized as containing high salient consistent objects, ten scenes containing high salient inconsistent objects, six scenes low salient consistent objects, and eight scenes low salient inconsistent objects.

To compare saliency levels between consistent and inconsistent AOIs and sets, we conducted a 2 (consistency) x 2 (image set) repeated measures ANOVA. The saliency of consistent AOIs (consistent condition, Set 1:  $M = 2.67$ , Set 2,  $M = 4.22$ ) were not

different from those of inconsistent AOIs (inconsistent condition, Set1:  $M= 3.67$ , Set 2:  $M=4$ )  $F<1$ . Also no difference between sets or interaction between semantically consistent and inconsistent scene type and set was found  $F>1$ .



*Figure 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 Procedures***

All participants completed the experiment in a sound attenuated, dimly lit room. Children participants were seated on the laps of their parents. All participants were presented with a set of 18 scenes. Half of the participants inspected the images of set 1, while the other half explored the images of set 2. Adult participants were also informed that they would perform a memory test that 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, each including six trials presented in a randomized order. Between the blocks, children were allowed to have a small break or to watch a short animation film with duration of 2 minutes. The total duration of the task was approximately 15 minutes. At the end of the experiment, adult participants were presented with images of 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 recorded responses will not be presented here.

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, Dale, Reznick, Thal, Bates, & Hartung, 1993). The parents were asked to fill the CDI in 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 or produced the labels of target objects used in the experiment. All parents filled the second questionnaire.

## ***2.5 Data Analysis***

The data analysis was carried out using SPSS (version 21.0) and MATLAB (version 8.0). Only fixations and saccades starting after the image appeared and finishing before the image disappeared were included in the analysis. All fixations outside the presentation screen were discarded.

Only the trials where the participants inspected the screen at least during 60% of total viewing time and children participants knew the label of the target object were included in the analyses. Thirteen percent of trials (177 out of 720 trials) were excluded in children while no trials were excluded from adult data. Also participants with more than 50% of rejected trials were excluded from the analyses. The data of three children were rejected.

The pre-processed eye-tracking data was used to generate the following eye movement parameters for our analyses. In the statistical analyses, the main effects and interactions of consistency (consistent versus inconsistent), saliency (high versus low) and age group (24

month-olds versus adults) were subjected to repeated measures analyses of variance (ANOVA). Paired t-tests were used in the post-hoc analyses when interaction was found. Additionally, a repeated measures ANOVA with consistency as within-subjects factor and vocabulary skills (normal-high versus normal-low producers) as between-subjects factor was performed to determine the effect of vocabulary size on inconsistency detection in children.

### ***3. Results***

As a first step, we analyzed extrafoveal and foveal measures to examine whether semantic inconsistency modulates attention allocation in adults and toddlers. Latency of the first saccade and the index of the first fixation AOI served as extrafoveal measures by indicating to what extent eye movements—before the first fixation on the target—are modulated by consistency. Foveal measures reveal the degree of attention allocated to the target. These measures were dwell time (i.e., looking time in the AOI divided by the total looking time of the trial), first-pass gaze duration (i.e., looking time in the AOI immediately after the first fixation in this region), first-pass fixation count (the amount of fixations in the AOI immediately after the first fixation in this region) and first and second fixation durations (the durations of first and second fixation landings in the AOI).

#### ***3.1 Extrafoveal processing of scene consistency and saliency***

To test whether semantic inconsistency influenced eye movements prior to the first fixation, we performed 2 (age group) x 2 (consistency) x 2 (saliency) repeated measures of ANOVAs with the index of the first fixation in the AOI and latency to the first saccade to the AOI as dependent variables.

The main effect of consistency was not significant on the two measures, all  $F < 1$ . The main effect of saliency  $F(1,64) = 5.2$ ,  $p < .05$  and age group was significant on the latency to the first saccade to the AOI  $F(1,64) = 24.5$ ,  $p < .01$ . Adult participants were significantly faster to fixate the AOI than toddlers (Table 1). There was also an interaction between saliency and age group for both parameters. Paired t-tests for the first fixation in the AOI  $t(35) = -2.77$ ,  $p < .05$  and the latency to the first saccade to the AOI  $t(35) = -2.84$ ,  $p < .05$  showed that toddlers reached the salient AOIs before than less salient AOIs. No differences were found in adults  $p > .05$  (Table 1).

Table 1. Means and standard deviations for extrafoveal and foveal parameters for both age groups

	<i>Adults</i>				<i>Children</i>				
	<i>Consistent</i>		<i>Inconsistent</i>		<i>Consistent</i>		<i>Inconsistent</i>		
	<i>High</i>	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>	<i>Low</i>	
Latency to the first saccade to the AOI (ms)	1408 (470)	1295 (407)	1440 (849)	1295 (849)	1224 (708)	1751 (777)	2422 (1563)	1724 (990)	2413 (1438)
Index of the first fixations to the AOI	5.87 (1.72)	6.01 (1.66)	6.02 (2.73)	6.01 (1.66)	5.22 (2.12)	5.18 (1.69)	7.11 (4.46)	5.22 (2.5)	6.48 (3.1)
Dwell time (%)	0.13 (0.04)	0.11 (0.05)	0.23 (0.08)	0.23 (0.08)	0.17 (0.06)	0.17 (0.09)	0.12 (0.1)	0.24 (0.09)	0.15 (0.11)
First-pass gaze duration (ms)	454 (180)	423 (248)	793 (276)	793 (276)	490 (193)	710 (409)	701 (965)	1067 (770)	724 (616)
First-pass fixation count	1.71 (0.49)	1.5 (0.69)	2.38 (0.73)	2.38 (0.73)	1.58 (0.49)	1.71 (0.78)	1.57 (1.06)	2.14 (0.98)	1.68 (0.97)
First fixation duration (ms)	261 (69)	264 (77)	342 (117)	342 (117)	306 (124)	403 (155)	459 (297)	439 (230)	389 (197)

### *3.2 Foveal processing of scene consistency and saliency*

The influence of semantic consistency and saliency on eye movements after the first fixation was examined by analyzing dwell time, first-pass gaze duration, first-pass fixation count, first fixation duration with four separate 2 (age group) x 2 (consistency) repeated measures ANOVAs.

There was a main effect for consistency  $F(1,70) = 37.07$  and saliency  $F(1,70) = 23.32$ ,  $ps < .01$  on dwell time, with longer dwell time for inconsistent (see Fig. 2 and Table 1) and for salient (see Fig. 4 and Table 1) objects compared with consistent and non-salient objects. There was no main effect for age group  $F < 1$  and no interaction between any parameter.

The main effect for consistency  $F(1,70) = 19.55$  and saliency  $F(1,70) = 18.01$ ,  $ps < .01$  were also found on the first pass gaze duration, indicating that more time was spent on inconsistent and on salient objects. The main effect of age was also significant,  $F(1,70) = 7.82$ ,  $p < .01$ , children spent more time than adults in the AOI the first time they reached it. No interactions were found,  $F_s < 1$  (see Table 1).

For the first-pass fixation count, we also obtained a significant main effect for consistency  $F(1,70) = 13.68$ , and saliency,  $F(1,70) = 22.84$ ,  $p < .01$ , indicating a greater number of fixations on inconsistent and salient objects (see Table 1). No further main effects or interactions were found,  $F_s < 1$ .

We also analyzed the effect of consistency on first fixation duration. No significant main effects were found for consistency  $F(1,64) = 1.07$  or saliency,  $F < 1$ ,  $ps > .05$ . The main effect of age was significant,  $F(1,64) = 22.43$ ,  $p < .01$ . Fixation durations were longer in toddlers than adults (see Table 1).

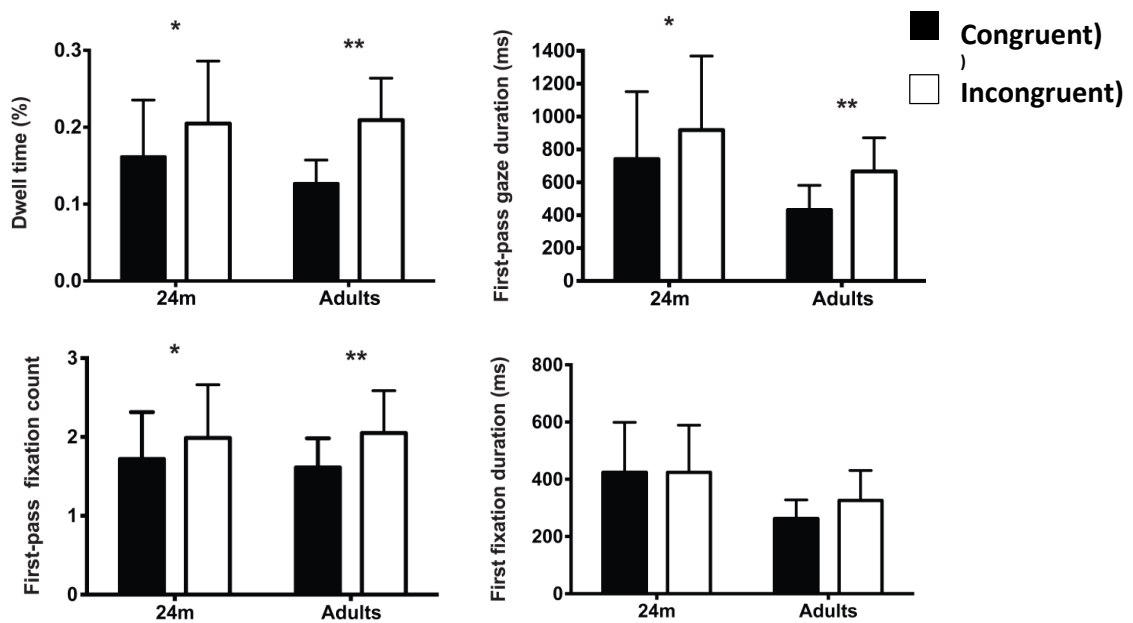


Figure 2. Foveal parameters for age groups. Dwell time (A), first-pass gaze duration (B), first-pass count (C), first fixation duration (D) in toddler and adult groups for consistent and inconsistent scenes. Error bars depict standard deviation. \* < 0.05, \*\* < 0.01

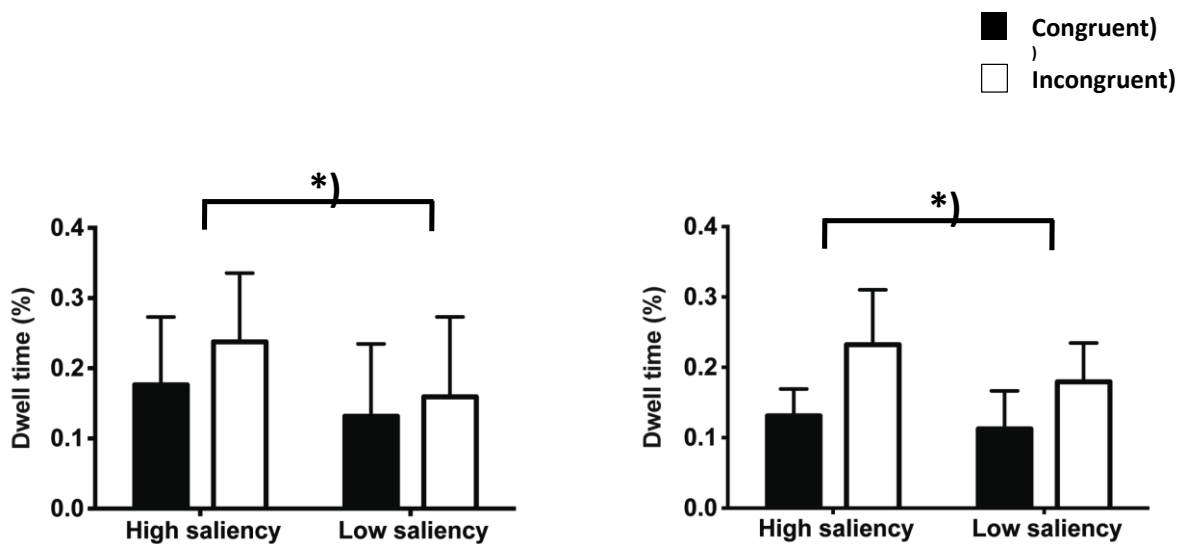


Figure 3. Dwell times for high and low saliency AOIs. Dwell time for low- and high-salient consistent and inconsistent AOIs in toddler (A), and adult (B) groups. Error bars depict standard deviation. \* < 0.05

### 3.3 Vocabulary size effect

Next, we explored whether productive vocabulary skills interact with semantic consistency and saliency effect in toddlers. In this analysis only the 33 toddlers, from whom the vocabulary score was obtained, were included. The mean vocabulary score was calculated for each toddler. From the range of 8-517 words, the median vocabulary score (261) was calculated and used to split the toddlers into two word producer groups. Accordingly, we obtained a normal-to-low producer group ( $n = 16$ , 7 girls) and normal-to-high producer group ( $n = 17$ , 11 girls). In the normal-to-low group, the average score was 136 words ( $SD = 69$ , range 8–248 words) and in the normal-to-high group the average score was 372 words ( $SD = 86$  range, 274–517 words). The scores were significantly different between the producer groups,  $t(31) = 8.89$   $p < .01$ . In addition to the CDI we also examined whether children were familiar with the objects illustrated in the AOIs. On average, based on the evaluation by their parents, toddlers had seen 97% of the target objects (normal-to-high producers 96%, normal-to-low producers 97%). They understood the labels 89% of target objects (normal-to-high producers 90%, normal-to-low producers 87%) and were able to produce 67% (normal-to-high producers 79%, normal-to-low producers 52%) of them.

Possible influences of vocabulary size on semantic consistency and saliency effects were examined. In the following section only interactions between language and consistency or saliency will be reported. As in previous section, index of the first fixation and latency to the first saccade to the AOI was analysed as a representative parameter of extrafoveal oculomotor behaviour. A 2 (language group) x 2 (consistency) x 2 (saliency) repeated measures ANOVAs was conducted. No main effect of language group and no interaction effect were found for any on these two parameters,  $F_s < 1$ ,  $p_s > .05$ .

Possible influences of vocabulary size on semantic consistency and saliency effects were analysed using only the parameters that were shown to be sensitive to manipulations of saliency and semantic previously: dwell time, first-pass gaze duration, and first-pass fixation count were analysed using three separate 2 (language group) x 2 (consistency) x 2 (saliency) repeated measures ANOVAs.

For dwell time to the AOI there was a significant main effect for the language group  $F(1,31) = 6.15$  and an interaction between age group and consistency,  $(1,31) = 6.54$ ,  $p < .05$ . No other interactions were found. Paired t-tests revealed longer dwell time to AOIs in semantically inconsistent than consistent scene types in normal-to-low producers,  $t(15) = -$

3.91,  $p < .01$ , whereas no significant difference was found for normal-to-high producers,  $t(16) = -0.17$ ,  $p > .05$ . Additionally, dwell times to consistent and inconsistent AOIs were compared between language groups. High producer presented longer dwell times to the consistent objects than low producers,  $t(31) = 3.83$ ,  $p > .01$ . No differences were found for inconsistent objects,  $p > .05$ .

The main effect of language group was not significant,  $F(1,31) = 2.32$  on the first pass gaze duration and the interaction between consistency and language group was marginal  $F(1,31) = 3.98$   $p = .055$ . Paired t-tests showed longer looking times to AOIs in inconsistent than consistent scenes in normal-to-low,  $t(15) = -3.41$ ,  $p < .01$ , but not in normal-to-high,  $t(16) = 0.45$   $p > .05$  producers. For the first-pass fixation count, there was an interaction effect, between consistency and language group  $F(1,31) = 6.94$ ,  $p < .05$ . No effect for language group  $F < 1$  and no other interactions were found. Again, paired t-tests showed the larger amount of fixations to AOIs in inconsistent than consistent scenes types in normal-to-low producers,  $t(15) = -4.1$ ,  $p < .01$  but no differences were found in normal-to-high producers  $t(16) = -0.26$ ,  $p > .05$ . Additionally, the first pass gaze durations and the first-pass fixation count to consistent and inconsistent AOIs were compared between language groups. High producer presented longer first pass gaze durations the consistent objects than low producers,  $t(31) = 2.29$ ,  $p > .05$  while no differences were found between language groups for inconsistent objects in this parameter,  $p > .05$ . When first-pass fixation count was analysed, no differences were found between language groups in any consistency.

In addition four parameters previously associated to attentional engagement (total trial time) and visual processing (fixation durations) fast scene context processing (first fixation to the scene) and gaze distribution over the scene (scan path length) were examined. The looking time to the screen during the whole trial (total trial time) was analysed to find out whether children in both language groups were equally engaged to the task. Also fixation durations during the whole viewing time were analysed to compare visual speed processing between language groups. The total trial time,  $t(31) = .04$ , the fixation durations,  $t(31) = -.18$ , first fixation to the scene,  $t(31) = .2$ , and scan path length,  $t(31) = -1.72$ , all  $ps > .05$  were not different between normal-to-high (The total trial time = 5934 ms, SD = 387, fixation durations = 338 ms, SD = 59 ms, first fixation to the scene = 409, SD = 109, scan path length 83, SD = 20) and normal-to-low (The total trial time = 5929 ms, SD = 386, fixation durations = 342 ms, SD = 59, first fixation to the scene 402, SD = 101, scan path length = 94, SD = 16) producers.



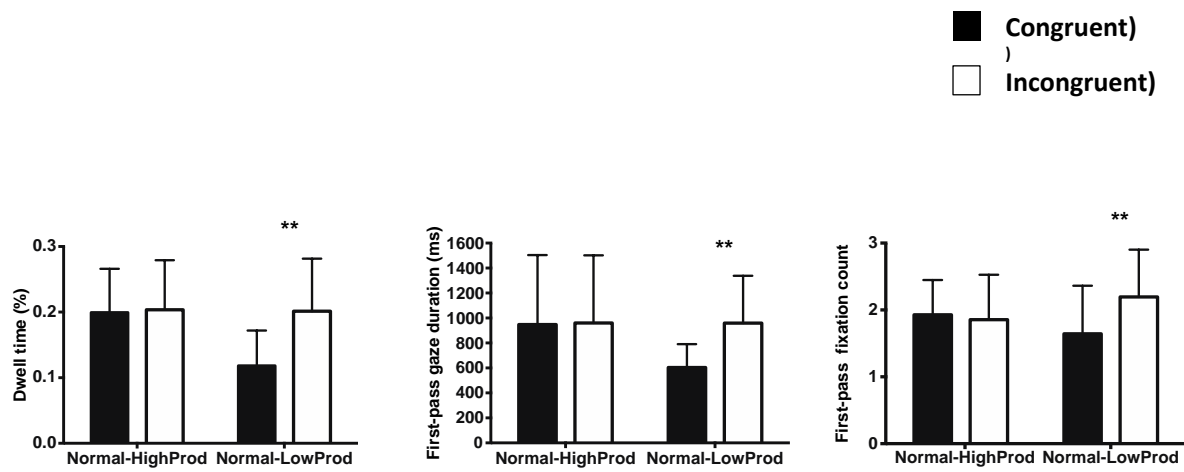


Figure 4. Foveal parameters for language groups. Dwell time (A), first-pass gaze duration (B), first-pass count (C) in normal-high and normal-low producer groups for consistent and inconsistent AOIs. Error bars depict standard deviation. \* < 0.05, \*\* < 0.01

#### *4. Discussion*

In the present study, our aims were the following: first, we investigated the interaction of saliency and semantic inconsistencies on attention allocation in 24-month-old children during a free exploration of visual scenes. Second, we assessed the effect of vocabulary skills on the guidance of eye movements. We compared looking times to semantically inconsistent and consistent high-salient and low-salient objects during visual exploration of everyday scenes in toddlers and in adults. Our results showed that both children and adults looked longer to high salient than to low-salient objects. However, toddlers were attracted faster to high-salient than low-salient objects while saliency did not affect early eye movement guidance in adults before the objects were fixated. Both age groups presented a similar consistency effect (i.e. they looked longer at inconsistent than consistent objects). However, even if semantically inconsistent objects attracted equally the gaze of children from both vocabulary groups; the consistent objects attracted significantly more the attention of toddlers with higher than lower vocabulary skills. As a result, toddlers with lower vocabulary skills presented a larger consistency effect than toddlers with higher vocabulary skills.

Consistency effects on foveal parameters were found in both age groups. This effect has been previously demonstrated in adults (Friedman, 1979; Henderson et al., 1999; Vö & Henderson, 2009), however, our findings provide new developmental evidence that 24-month-old children are capable to generate semantic expectations of their visual environment and to use scene context to guide their eye movements. Earlier developmental studies have shown that infants of 4- to 6-months take in account the context in which objects are embedded when they explore single objects or faces (Bornstein et al., 2011a, 2011b; Richmond & Nelson, 2009). Also, it has been previously reported that at 15 months of age infants are able to detect a semantically inconsistent object in a change blindness paradigm (Duh & Wang, 2014). However, in these previous studies, the objects were placed near to the center of the image and the density of objects was low or limited to a single object. Central bias (Tatler, 2007) and facilitation of object processing in low density scenes have already been reported in adults (Pannasch, Helmert, Roth, & Walter, 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005). Thus, these factors may have helped young infants to allocate their gaze to inconsistent objects and consequently to perceive inconsistencies between objects and contexts. To our knowledge, our study is the first to show that toddlers are able to use scene context to guide their attention during a free exploration of real scenes containing a high

density of objects.

Consistency effects were found on foveal but not on extrafoveal measures both in toddlers and in adults, providing further evidence to the debate on whether semantic inconsistent objects can be processed or not in visual periphery. Our results support the idea that foveal processing is necessary to detect semantic object–scene inconsistencies during scene viewing (De Graef et al., 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Vö & Henderson, 2009; Vö & Henderson, 2011). The latencies of the first saccades to both consistent and inconsistent objects as well as the time spent looking at these objects immediately after they were fixated were longer in toddlers than in adults. However, both age groups exhibited equal amount of fixations. These results suggest that both age groups had similar exploration strategies even execution of eye movements were slower in toddlers. Longer fixation durations during scene perception have already been reported in children compared with adults (Helo et al., 2014; Helo, Rämä, Pannasch, & Meary, in press). Also, longer fixations have been associated with higher cognitive effort during visual processing (Colombo, Mitchell, Coldren, & Freesean, 1991; Wass & Smith, 2014) and with longer time to program the following saccade (Fukushima, Hatta, & Fukushima, 2000; Gredebäck, Örnkloo, & von Hofsten, 2006; Irving, Steinbach, Lillakas, Babu, & Hutchings, 2006; Klein & Foerster, 2001; Luna, Velanova, & Geier, 2008; Matsuzawa & Shimojo, 1997; Munoz, Broughton, Goldring, & Armstrong, 1998). Altogether, our findings and previous evidence suggest that differences between age groups in our study are explained by speed of visual processing and eye movement maturity rather than by delayed skills in processing of semantic scene context.

Toddlers allocated their gaze earlier to high salient than to low salient objects while adults did not exhibit the same effect. This finding is in line with previous developmental studies showing a stronger influence of saliency on eye movement guidance in children than in adults (Açık et al., 2010; Helo et al., 2014; Kooiker, Steen, & Pel, 2016). However, after the objects were fixated, the differences disappeared and both age groups looked longer to high salient than to low salient objects. Our results agree with previous reports showing that saliency attracts the gaze in adults (Clarke et al., 2013; Coco et al., 2014; Itti & Koch, 2000). Furthermore, our study provides additional evidence that even though perceptual features have stronger effect on the early eye movements in toddlers than in adults the influence of saliency on object processing is similar between these two age groups. There was no interaction between saliency and semantic processing after the objects were fixated in either of the two age groups. Both toddlers and adults looked longer at inconsistent than consistent

objects in both saliency conditions. This finding provides additional evidence to the previous literature in adults showing that saliency does not affect semantic inconsistency detection (Coco et al., 2014; Foulsham & Underwood, 2008; Underwood & Foulsham, 2006) and provides new evidence about the contribution of saliency and semantic features to gaze allocation in young children. Altogether, our results indicate a stronger saliency bias on the early eye movements in toddlers than in adults but similar contribution of perceptual and semantic guidance during object processing in both age groups.

In addition, we hypothesized that vocabulary skills influence visual attention in toddlers. Earlier, it has been shown that 24-month-old children are more likely than adults to activate linguistic representations spontaneously in non-communicative contexts (Khan, 2013). Also, there is a previous study showing that at two years of age, the amount of spontaneous naming during manipulation of novel objects is associated with productive vocabulary size (Samuelson & Smith, 2005). Based on these previous findings, we expected that children with higher vocabulary size were more likely to name objects during visual exploration. Longer looking times to consistent objects has been previously observed in adults when they had been asked to name objects during and after scene exploration (Clarke et al., 2013), and thus, we also expected that high producers look longer to the consistent objects than low producers. Our results showed that both language groups were equally attracted to the inconsistent objects (dwell time 20% versus 19%, respectively, see Fig. 4) indicating that both language groups detected semantic inconsistencies. In contrast, normal-to-high producers exhibited longer looking times to consistent objects than normal-to-low producers (20% versus 12%, respectively). According to parental reports, normal-to-high producers were able to name the majority of objects illustrated in the AOIs while normal-to-low producers were able to name only about half of these objects (both language groups understood equal number of object labels). Thus, it is possible that children with higher vocabulary skills silently named the objects more often than children with lower vocabulary skills and additionally directed their gaze more often to objects in general. No significant differences were found in other parameter such as attentional engagement (measured by total trial dwell time) or processing speed (measured by fixation durations) that could have reflected different cognitive abilities between language groups. Based on previous and our findings, we propose that high producers are more likely to silently name during scene exploration, and thus, in our study they behaved like adults performing an active naming task.

## *Conclusion*

Our study provided new evidence that 24-month-old children direct their attention to semantic scene violations during viewing of real visual scenes. We propose that by the age of two years, toddlers are able to use conceptual schemas of their everyday environment to guide further exploration of the complex visual scenes. Saliency had stronger influence on early eye movements in toddlers than in adults. However, after the objects were fixated, both semantic and salient features had similar effect on both age groups. Based on these findings, we propose that developmental differences in gaze allocation during scene exploration are related to stronger perceptual bias on gaze guidance rather than less top-down control in children. Regarding language skills, toddlers with higher vocabulary skills looked longer to consistent objects than those with lower production levels. However, saliency and semantic inconsistencies attracted the gaze equally in both language groups. Altogether, these findings indicate that language skills influence more semantic than perceptual guidance during scene perception.

### *Acknowledgments*

We express our gratitude to the infants and parents who participated to the study, we thank for their kindness and cooperation. We also thank Patrick Cavanagh, who supported this work by providing his EyeLink eyetracking system for this study. A.H. was supported by doctoral fellowship from CONICYT, Chile.

## References

- Açık, 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. doi:10.3389/fpsyg.2010.00207
- Barlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK. Cambridge University Press.
- 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. doi: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. doi: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. doi:10.1016/j.infbeh.2009.10.003
- Bornstein, M. H., Mash, C., & Arterberry, M. E. (2011). Perception of object-context relations: Eye-movement analyses in infants and adults. *Developmental Psychology, 47*(2), 364–375. doi:10.1037/a0021059.Perception
- Colombo, J., Mitchell, D. W., Coldren, J. T., & Freesean, L. J. (1991). Individual differences in infant visual attention: are short lookers faster processors or feature processors? *Child Development, 62*(6), 1247–1257. doi:10.1111/j.1467-8624.1991.tb01603.x
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564. doi:10.1111/j.0956-7976.2004.00719.x
- De Graef, P., Christiaens, D., & D’Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*, 317–329.
- Duh, S., & Wang, S.-H. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology, 72*, 142–61. doi:10.1016/j.cogpsych.2014.03.001
- Egami, C., Morita, K., Ohya, T., Ishii, Y., Yamashita, Y., & Matsuishi, T. (2009). Developmental characteristics of visual cognitive function during childhood according to exploratory eye movements. *Brain and Development, 31*(10), 750–757. doi:10.1016/j.braindev.2008.12.002
- Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology. General, 88*, 181–191. doi:10.1037/0096-3445.108.3.316
- 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. doi:10.1016/S0387-7604(00)00101-7

- Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. van Gompel (Ed.), *Eye movements: A window on mind and brain* (pp. 617–637). Amsterdam: Elsevier.
- Gopnik, A., & Meltzoff, A. (1987a). Early semantic developments and their relationship to object permanence, means-ends understanding, and categorization. *Children's Language*, 6, 191–212.
- Gopnik, A., & Meltzoff, A. N. (1987b). The Development of Categorization in the Second Year and Its Relation to Other Cognitive and Linguistic Developments. *Child Development*. doi:10.2307/1130692
- Gredebäck, G., Örnkloo, H., & von Hofsten, C. (2006). The development of reactive saccade latencies. *Experimental Brain Research*, 173(1), 159–164. doi: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. doi: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*, Submitted.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., Mack, M., Gompel, R. Van, Fischer, M., ... Eye, E. (2007). Visual saliency does not account for eye movements during visual search in real world scenes. In *Eye Movements: A Window on Mind and Brain* (pp. 537–562). doi:10.1167/9.3.6.
- 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. doi:10.4324/9780203488430
- 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. doi:10.1037/0096-1523.25.1.210
- 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. doi: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. doi: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. doi:10.1167/iovs.05-1311
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203. doi:10.1038/35058500
- Kail, R. (2000). Speed of information processing. *Journal of School Psychology*, 38(1), 51–61. doi:10.1016/S0022-4405(99)00036-9



- Kail, R., & Salthouse, T. (1994). Processing speed as a mental capacity. *Acta Psychologica*, 86(2-3), 199–225. doi:Cited By (since 1996) 227\rExport Date 21 February 2012
- Klein, C., & Foerster, F. (2001). Development of prosaccade and antisaccade task performance in participants aged 6 to 26 years. *Psychophysiology*, 38(2), 179–189. doi:10.1017/S0048577201981399
- Kollmorgen, S., Nortmann, N., Schröder, S., & König, P. (2010). Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS Computational Biology*, 6(5), 1–20. doi:10.1371/journal.pcbi.1000791
- Kooiker, M. J. G., Steen, J. Van Der, & Pel, J. J. M. (2016). Development of salience-driven and visually-guided eye movement responses. *Journal of Vision*, 16(5), 1–11. doi:10.1167/16.5.18.doi
- Kowler, E., & Martins, a. (1982). Eye movements of preschool children. *Science*. doi:10.1126/science.7156979
- Krishna, O., Yamasaki, T., Aizawa, K., Helo, A., & Rämä, P. (2016). Gaze Distribution Analysis and Saliency Prediction Across Different Age Groups. *European Conference on Computer Vision, Submitted*.
- 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. doi: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. doi: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. doi:10.1177/0956797610373371
- 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. doi:10.1016/S0163- 6383(97)90035-7
- Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, 14(4), 710–716. doi:10.3758/BF03196826
- 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. doi:10.1007/s002210050473
- Ngon, C., & Peperkamp, S. (2016). What infants know about the unsaid: Phonological categorization in the absence of auditory input. *Cognition*, 152, 53–60. doi:10.1016/j.cognition.2016.03.014
- 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. doi:10.1016/B978-012375731-9/50045-8
- Pannasch, S., Helmert, J. R., Roth, K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes : Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2), 1–19.
- 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.
- Poulin-Dubois, D., Graham, S., & Sippola, L. (1995). Early lexical development: The contribution of parental labeling and infants categorization abilities. *Journal of Child Language*, 22, 325–343.
- Pringle, H. L., Irwin, D. E., Kramer, a F., & Atchley, P. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin & Review*, 8(1), 89–95. doi:10.3758/BF03196143
- Rämä, P., Sirri, L., & Serres, J. (2013). Development of lexical–semantic language system: N400 priming effect for spoken words in 18- and 24-month old children. *Brain and Language*, 125(1), 1–10. doi:10.1016/j.bandl.2013.01.009
- Spotorno, S., & Faure, S. (2011). Change detection in complex scenes: Hemispheric contribution and the role of perceptual and semantic factors. *Perception*, 40(1), 5–22. doi:10.1068/p6524
- 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. doi: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.
- Torkildsen, J. V. K., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H. G., Moen, I., Lindgren, M. (2006). Semantic organization of basic-level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics*, 19(6), 431–454. doi:10.1016/j.jneuroling.2006.01.002
- Torralba, A., Oliva, A., Castelhamo, 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. doi:10.1037/0033-295X.113.4.766
- 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. doi: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.  
doi: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. doi: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. doi:10.1016/j.concog.2006.11.008
- Unema, P. J. a., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473–494. doi:10.1080/13506280444000409
- Võ, M. L., & 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. doi:10.3758/s13414-011-0150-6
- Võ, M. L., & 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. doi:10.1167/9.3.24.Introduction
- Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology*, 6(4), 632–650. doi:10.1016/0022-0965(68)90108-2
- Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19(9), 1395–1407. doi: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. doi:10.1111/infa.12049
- Yarbus, A. L. (1968). Eye movements and vision. *Neuropsychologia*, 6(4), 389–390. doi:10.1016/0028-3932(68)90012-2
- Zelinsky, G. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835. doi:10.1037/a0013118.A

*Study IV: In preparation*

*Word processing in semantic scene context: an ERP study in young children*

Helo, A.<sup>1,2</sup>, Azaiez, N.<sup>1,3</sup>, & Rämä, P.<sup>1,4</sup>

<sup>1</sup> Laboratoire Psychologie de la Perception, Université Paris Descartes, France

<sup>2</sup> School of Speech and Hearing Sciences, University of Chile, Santiago, Chile <sup>3</sup>

Department of Psychology, University of Jyväskylä, Finland

<sup>4</sup> CNRS (UMR 8242), Paris, France

Corresponding author: Andrea Helo

Address: Laboratoire Psychologie de la Perception (CNRS UMR

8158), Université Paris Descartes,

45, rue des Saints-Pères, 75006 Paris,

France. Fax: +33 (0)1 42 86 33 22

Email: [ahelo@med.uchile.cl](mailto:ahelo@med.uchile.cl)

### *Abstract*

Semantic scene context is known to prime object recognition. Developmental studies have shown that semantic priming occurs in object or word contexts already before the second year of life while less is known about semantic processing in scene context in young children. To study this question, 24-month old children were presented with visual scenes (e.g., kitchen) following by spoken words that were semantically either consistent (e.g., spoon) or inconsistent (e.g., bed) with the scene context. Event-related potentials (ERPs) were recorded in response to the words. The results showed that inconsistent words exhibited a larger N400 component over the frontal recording sites. However, individual vocabulary skills contributed to the timing and the distribution of the N400 component. A more negative N400 for inconsistent words was found over the right frontal recording sites in children with lower vocabulary skills while in children with higher vocabulary skills, a more negative N400 was observed over the left frontal sites. The N400 component appeared also earlier in high than in low producers. Our results suggest that 24-month-olds have acquired enough knowledge about semantic regularities of their visual world to integrate contextual scene knowledge to subsequent word processing. However, our results demonstrate that children with different linguistic skills might activate different neural structures in contextual priming.

**Keywords:** Scene semantics, contextual priming, vocabulary development, ERPs, N400 effect

## *1. Introduction*

Everyday visual environment typically contains multiple objects organized in predictable semantic and spatial configurations. A natural scene refers to a semantically coherent (and often nameable) human-scaled view of the real world comprising both background elements and objects (e.g., Henderson & Ferreira, 2004; Hollingworth & Henderson, 1999). During the course of visual experience, conceptual representations of visual scenes are built and stored in long-term memory (Hock, Romanski, Galie, & Williams, 1978; Mandler & Johnson, 1976; Oliva & Torralba, 2007; Potter, 1975). Stored scene knowledge is activated during the process of scene interpretation and it allows viewers to extract rapidly the global meaning of a particular scene (Henderson & Hollingworth, 1999; Oliva, 2005; Potter, 1975, 1976). Scene context facilitates also the processing of objects that typically belong to a particular scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport & Potter, 2004; Davenport, 2007; Friedman, 1979; Heise & Ansorge, 2014; Palmer, 1975). Studies using priming paradigms have shown that a short preview of a scene facilitates recognition, search and memorization of objects that are consistent with the scene context (Hillstrom, Scholey, Liversedge, & Benson, 2012; Josephs, Draschkow, Wolfe, & Võ, 2016; Palmer, 1975; Võ & Henderson, 2010). Likewise, semantically consistent objects are detected faster and more accurately than inconsistent objects during scene exploration (Biederman et al., 1982; Davenport & Potter, 2004; Davenport, 2007; Heise & Ansorge, 2014). Also event-related potential (ERP) technique has been used to investigate mechanisms of scene-object priming (Ganis & Kutas, 2003; Mudrik, Lamy, & Deouell, 2010; Võ & Wolfe, 2013). These studies demonstrated that a more negative N400 component was elicited in response to inconsistent than to consistent visual objects illustrated within visual scenes (Ganis & Kutas, 2003; Mudrik et al., 2010; Võ & Wolfe, 2013). The enhanced N400 component was suggested to be analogous to the N400 that has been shown to reflect semantic integration in linguistic modality (for review, Kutas & Federmeier, 2011). In the linguistic domain the N400 response typically appears over the central-parietal recording sites (for review, Kutas & Federmeier, 2011), while in visual studies, the N400 is usually more frontally distributed (Ganis, Kutas, & Sereno, 1996; Kutas & Federmeier, 2011; Kutas & Van Petten, 1994).

There is some recent evidence showing that semantic scene context or scene background affects object processing also in young children (Bornstein, Mash, & Arterberry,

2011a, 2011b; Duh & Wang, 2014; Richmond & Nelson, 2009). Bornstein and co-workers (Bornstein et al., 2011a, 2011b) investigated object-context relations in 4 and 6 month of age infants during a free viewing task. Infants were presented with images of vehicles or animals (e.g., car or tiger) in different contexts: in a white background, in an incongruent background (a car in a grass or a tiger in a street) or in a congruent background. The authors found that infants presented different eye movement patterns depending on the object-context relation. In another study, 9-month-olds were tested in a scene-face paradigm in which they were first presented with three faces that were superimposed on three different arbitrary scenes. After encoding phase, children were presented with all three faces together with one of the scenes seen previously, and the results showed that infants preferred to look at the face that had been presented with the test background earlier (Richmond & Nelson, 2009). These studies suggest that infants take into account the context in which the objects are presented already in their first year of life. It has been further demonstrated that by the age of 15 months infants have already acquired, at least a preliminary, knowledge about the semantic rules or regularities of their visual environment and they look longer a visual display illustrating an object that violates these regularities (Duh & Wang, 2014). Using a free exploration task, it was further demonstrated that 2-years-old children looked longer at semantically inconsistent than consistent objects (Helo et al., in preparation).

The ERP technique has not yet been applied to study detection of semantic violations in scene context in young children: one of the reasons might be that controlling the eye-movements of young children is challenging. In contrast, the ERP technique has been widely used in developing populations to study detection of semantic violations in an object context. In a typical cross-modal picture-word matching task, the children are presented with common objects that are either consistent or inconsistent to the object labels (e.g., the picture dog together with a word of a dog or a car). Several studies have found the N400 component to be larger for inconsistent than for consistent object labels (e.g. Friedrich & Friederici, 2004, 2005a, 2005c, 2010; Torkildsen et al., 2006, 2008; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007). The N400 effect, a larger amplitude for inconsistent words, has been observed in 14- to 24-month-old children, and also already in 12-month-olds with a relatively high (more than 4 words) productive vocabulary, suggesting that children are capable of integrating visual and auditory information based on their semantic content during early development (Friederici, 2006; Friedrich & Friederici, 2004, 2005a, 2005b, 2010). In the study by Torkildsen et al. (2006), the incongruity between the picture and the word was either between-category (e.g., a picture of a dog with the word “car”) or within-category

(e.g., a picture of a dog with the word “cat”) violation, and the ERPs were compared with those elicited by a congruent (e.g., a picture of dog with the word “dog”) condition. The N400-like effect was earlier and larger in 20-month-olds for between- than for within-category violations, suggesting that by the end of second year of life, children have developed a semantically graded lexicon where basic-level words for objects from the same superordinate category have a closer relationship than basic-level words from different superordinate categories (Torkildsen et al., 2006).

The distribution and the lateralization of the N400 component have been shown to vary across developmental studies (e.g. Friedrich & Friederici, 2004, 2005a, 2005c, 2010; Torkildsen et al., 2006, 2008, 2007) and also vocabulary skills have been shown to contribute to the occurrence, latency and the distribution of the N400. In 20-month-olds, the N400 effect was found for between category violations in children with both low and high vocabulary skills, while the effect for within category violation was found only in children with a high vocabulary level (Torkildsen et al., 2006). The distribution of the N400 effect was found to be more focally distributed in 24-month-olds compared with 19-month-olds (Friedrich & Friederici, 2005c) as well as for newly learned than for familiar words at 20-month-olds (Torkildsen et al., 2008). Another ERP study, using a semantic priming task for spoken words, showed that the N400 priming effect was obtained in 24-month-olds, but only in the subgroup of 18-month children with normal-to-high productive skills (Rämä, Sirri, & Serres, 2013). The N400 has also shown to appear earlier in children with higher than in lower word productive skills (Friedrich & Friederici, 2004; Torkildsen et al., 2006, 2008).

Children learn words in everyday contexts in which there are many words, many possible referents, or sometimes even no referents, and the cues to guide the acquisition of word-to-world mappings might be limited. Children also learn words in a variety of action and discourse contexts (Tomasello, Strosberg, & Akhtar, 2008) and they can even acquire novel words from overheard speech (Akhtar, Jipson, & Callanan, 2014). Visual context might contribute to the acquisition of word-to-object mappings since certain words are more frequently expressed in specific contexts (e.g., word “knife” in kitchen) than in some others. Our aim in the current study was to investigate the effect of semantic scene context on word processing in 24-months-old children. The age from 20 to 22 months is associated with a burst in vocabulary size for both comprehension and production (Ganger & Brent, 2004; Nazzi & Bertoncini, 2003), and at the end of the second year of life, the mental lexicon is shown to be organized by semantic taxonomic



categories in linguistic modality (Arias-Trejo & Plunkett, 2013; Delle Luche, Durrant, Floccia, & Plunkett, 2014; Rämä et al., 2013; Sirri & Rämä, 2015; Styles & Plunkett, 2009; Torkildsen et al., 2006). Here, we created a contextual priming task involving both a visual and a linguistic component. Children were presented with consistent (e.g., kitchen scene and word ‘knife’) and inconsistent (e.g., kitchen scene and word ‘bus’) scene-words pairs and the ERPs were recorded in response to target words. Our aim was to investigate whether contextual information facilitates word processing as manifested by more enhanced N400 component in response to incongruent than to congruent target words. Based on previous literature on language-related ERPs showing that language skills contribute to the occurrence of the N400 response (Friedrich & Friederici, 2004, 2010; Torkildsen et al., 2006, 2008), we expected that the N400 effect in children with higher vocabulary skills is more pronounced and occurs earlier than in children with lower vocabulary skills.

## ***2. Material and methods***

### ***2.1. Participants***

Thirty-one (15 girls and 16 boys; age range: 23-24 months, mean age: 24 months) participated in the study. All children were monolingual French learners. Children were recruited from a database of parents who agreed to volunteer in child development studies, and came from middle class socio-economic backgrounds in the Parisian region. The mean number of maternal education years was 16 (range from 11 to 19 years). All children were born full-term and none of them suffered from hearing or language impairment. The parents were informed about the purpose of the study and gave informed consent before participating. The French translation of the MacArthur Communicative Development Inventory for Words and Sentences (CDI; Fenson, Dale, Reznick, Thal, Bates, & Hartung, 1993) was used to measure productive vocabulary size. The parents were asked to fill the inventory at home during the two weeks following the study. Based on vocabulary inventory scores and medial split, the participants were divided into two language groups: normal-to-high and normal-to- low producers. The mean number of maternal education years was 16 (range from 12 to 19 years) for children in normal-to-high and 15 (range from 11 to 19 years) in normal-to-low producer groups. An additional twenty children were tested but their data were rejected due to not being able to complete the experiment ( $n = 14$ ), insufficient number of trials in each experimental condition ( $n = 4$ ) or parents did not return CDI ( $n = 2$ ). The

study was conducted in conformity with the declaration of Helsinki and approved by the Ethics Committee of the University of Paris Descartes.

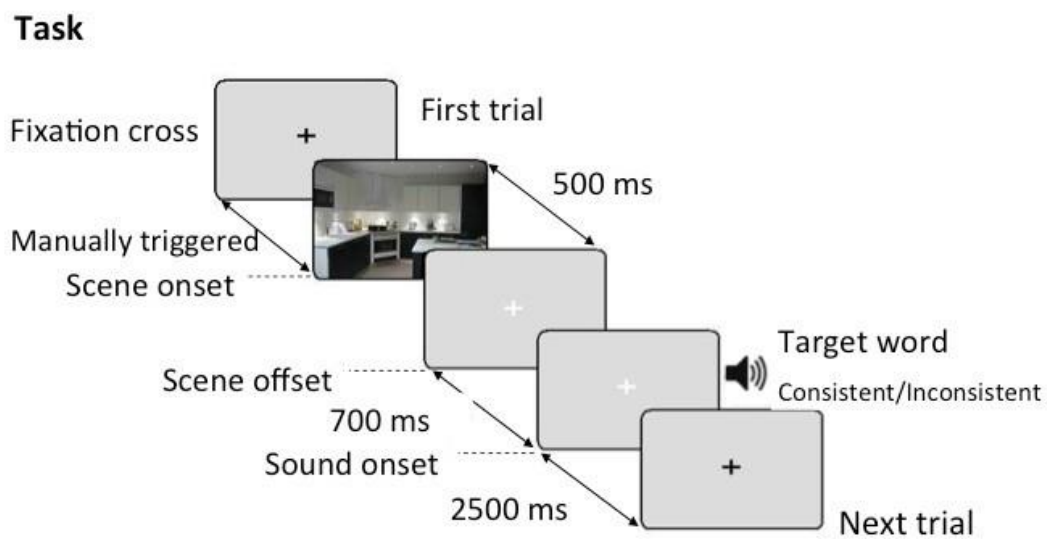
## ***2.2. Stimuli***

A total of seventy-two colored real-world photographs (1024 x 768 pixels) illustrating six different scene types were used as stimuli (Fig. 1). Four scene types illustrated home interior scenes (kitchens, bathrooms, bedrooms or living rooms) and two of them outdoors scenes (parks or streets views). Each scene type contained equal number of scene items. The images were selected from Internet and represented examples of indoor or outdoor sceneries of Parisian homes or neighborhoods. Spoken words were thirty-six basic level familiar words (Table 1). Word durations varied between 431 and 1057 ms (mean duration 694 ms). The mean number of syllables was 1.9 (range: from 1 to 3). The words were recorded and edited with Audacity V 2.0.5 and Praat V 5.2.2 softwares. The speaker was a native French female who was asked to pronounce the words slowly in a neutral voice. The sound levels were normalized among the words. Visual scenes and spoken words were arranged into thirty-six semantically consistent (e.g., kitchen scene and word “spoon”) and thirty-six semantically inconsistent (e.g., kitchen scene and word “bus”) prime-target scene-word pairs (Table 1). Each scene type was presented equally both in consistent or inconsistent conditions. Each word was also presented twice; once in each condition. Each scene-word pair was presented twice resulting in a total of 144 trials.

## ***2.3. Experimental procedure***

Children were seated on their parents’ laps in a dimly lit room at  $\approx$ 80 to 100 cm from a 17” LCD monitor (with a 1024 x 768 pixels resolution, 32 bits color quality and 100-Hz refresh rate) and two loudspeakers from which the sounds were delivered. Parents were instructed not to communicate verbally with their child during the actual experiment. Each trial began with the presentation of a black cross on a light grey background followed by the presentation of a scene image. Children’s gaze was monitored through a video camera placed in front of them and the scene image onset was triggered manually by the experimenter once the child was looking at the screen. The scene image remained visible for 500 ms following by a spoken word, which was delivered 700 ms after the offset of the image. After 2500

ms, the next trial started. (Fig.1). If the child lost her/his interest in the task, or did not direct her/his gaze on the screen, a short animation movie was shown to motivate the child again. The trials (72 congruent and 72 incongruent) were presented in an intermixed and randomized order and the whole experiment lasted about 15 minutes. Short breaks were taken if needed.



*Figure 1.* Illustration and timing of a trial in the scene-word priming task.

#### ***2.4. EEG acquisition and pre-processing***

Continuous electroencephalogram (EEG) was recorded (bandpass = 0.1–100 Hz, sampling rate = 250 Hz) from 128 electrodes using a Geodesic Sensor Net (GSN, Netstation EGIS V4.5.6) referenced to the vertex during the acquisition. Impedances were kept below 50 k $\Omega$ . EEG data analysis was performed using EEGLAB 13.4, a freely available open source toolbox (<http://www.sccn.ucsd.edu/eeglab>) running under Matlab 8.4 (The Mathworks). First,

data were filtered (0.3–30 Hz) and bad channels were replaced using spherical spline interpolation. The data were re-referenced to the average of all electrodes. Epochs were extracted based of the word onset (-200 – 1000 ms) and baseline corrected (-200 – 0 ms).

Only the epochs where the participants knew the word were included in the analyses. On average, children knew 78% of the words presented in the task (normal-to-high producers 86%, and normal-to-low producers 72%). The average of rejected epochs based on word knowledge was 8 (normal-to-high producers = 5, range 0-15, and normal-to-low-producers = 10, range 0-23). The epochs including artifacts (eye-movements, blinks, motion artifacts exceeding  $\pm 150 \mu\text{V}$  in any channel) were also excluded. The epochs were averaged separately for each subject and trial type (consistent or inconsistent). The epochs were grand-averaged across all participants for each trial type. A minimum of 10 artifact-free trials per trial type was included in further analyses. The mean number of accepted trials was 35 (SD = 18) for consistent and 36 (SD = 17) for inconsistent trial type. The mean number of trials in normal-to-high producers was 41 for consistent (SD = 20) and 42 for inconsistent (SD = 18) trials. The mean number of trials for normal-to-low producers was 30 for consistent (SD = 14) and 30 for inconsistent (SD = 14) trials. The number of accepted trials between language groups,  $F = 3.83$ ,  $p > .05$  or trial types,  $F < 1$ ,  $p > .05$  were not significantly different. No interaction between these factors was found  $F < 1$ ,  $p = > .05$ .

### ***2.5. Data and Statistical analysis***

The N400 component was measured as mean amplitude at three time windows: from 400ms to 550ms, 550ms to 700ms, and from 700ms to 850ms. The mean amplitudes of each time window were calculated separately for each electrode and the means, extracted from the nine frontal and nine parietal electrodes, were averaged. The midline electrodes were excluded from the statistical analyses, resulting in 36 channels in four regions of interest. The 36 electrodes with their approximate equivalents according to the 10-20 international system of electrodes sites (when appropriate/possible) are as follow: 26 (AF7), 23 (AF3), 18 (AF1), 33 (F7), 27 (F5), 24 (F3), 19 (F1), 7 (FFC1), and 22 (FP1) in left frontal, 2 (AF8), 3 (AF4), 10 (AF2), 122 (F8), 123 (F6), 124 (F4), 4 (F2), 5 (FFC2), and 9 (FP2) in right frontal, 41 (C5), 36 (C3), 47 (CP5), 42 (CP3), 37 (CP1), 52 (P3), 53 (CPP3), 54 (CPP1), 61 (P1), in left central-posterior, 104 (C6), 103 (C4), 98 (CP6), 93 (CP4), 87 (CP2), 92 (P4), 86 (CPP4), 79 (CPP2), 78 (P2), in right central-posterior (see Figs. 1A and 1B).

The statistical analyses were conducted with SPSS (IBM SPP statistics, version 20). A repeated-measures analysis of variance (ANOVA) included trial type (consistent versus inconsistent), recording site (frontal versus parietal), and hemisphere (left versus right) as within-subject factors, and productive vocabulary skills group (high versus low) as a between-subject factor was conducted separately for each time interval of interest. When the interaction between trial type and language group was significant, ANOVAs were conducted separately for each language groups. The Greenhouse-Geisser correction was applied for non-sphericity when appropriate. Only the main effects of the factors productive vocabulary skills group, the trial type, significant interactions with these factors, and significant post hoc comparisons are reported

### ***3. Results***

#### ***3.1. Vocabulary skills***

In total, children produced between 8 and 517 words, and the median vocabulary score (219) was used to group the children into two language groups, hereafter labeled as normal-to-low ( $n = 15$ , 8 girls) and normal-to-high ( $n = 16$ , 7 girls) groups. Children in the normal-to-low group produced on average 118 words ( $SD = 69$ , range 17 – 219 words) while children in the normal-to-high group produced on average 326 words ( $SD = 120$  range, 222 – 579 words). The scores were significantly different between the groups,  $t(29) = 5.84$   $p < .01$ .

#### ***3.2. ERPs in response to target words***

##### ***N400: first time window***

A significant interaction among trial type, recording site, hemisphere, and language group was found for the amplitudes of N400 during the first time window,  $F(1,29) = 11.7$ ,  $p < .01$ . Follow up ANOVAs were conducted separately for each language group. The normal-to-high producers presented a significant interaction among trial type, recording site, and hemisphere,  $F(1,14) = 10.8$ ,  $p < .01$ . In the normal-to-low producers, the interaction among trial type, recording site, hemisphere, was close to significant,  $F(1,15) = 4.35$ ,  $p = .054$ . A paired t-test confirmed that the mean amplitudes were more negative for inconsistent ( $2.71$   $\mu V$ ,  $SD = 6$ ) than for consistent ( $4.21$   $\mu V$ ,  $SD = 5.84$ ) scene-word pairs over the left frontal

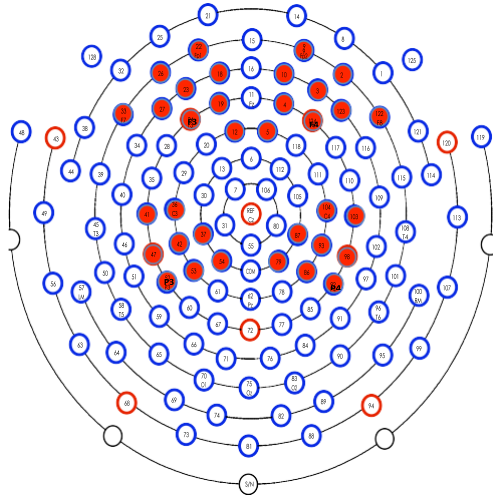
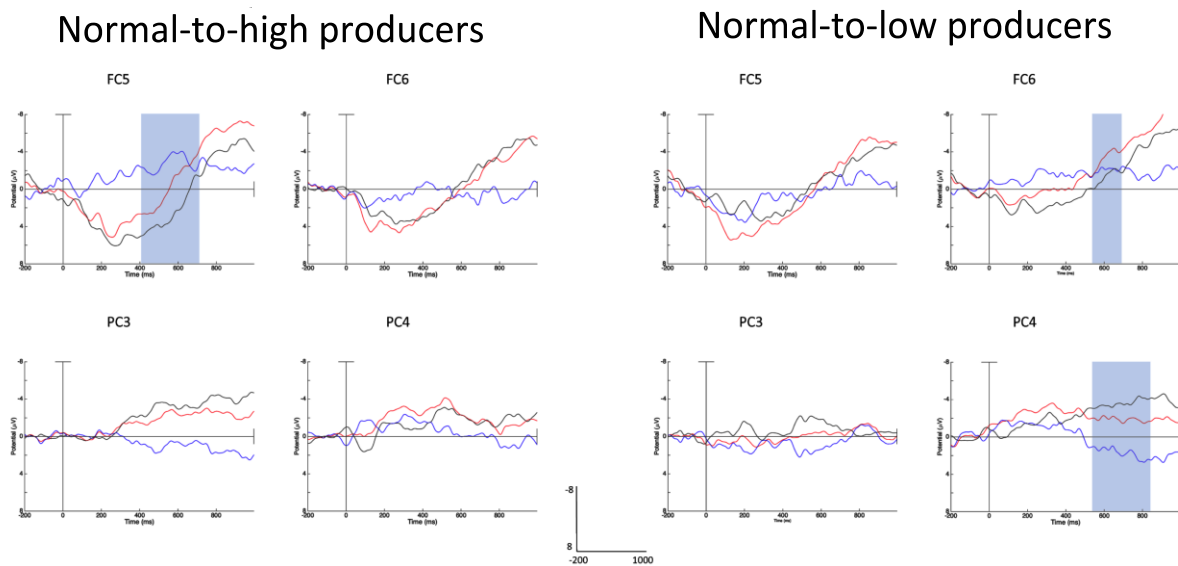
recording sites in normal-to-high producers,  $t(14) = 2.69$ ,  $p < .05$ . Significant differences in amplitudes between trial types were not found in normal-to-low producers at any recording site, all  $ps > .05$ .

#### ***N400: second time window***

Again a significant interaction among trial type, recording site, hemisphere, and language group,  $F(1,29) = 7.64$ ,  $p = .01$  was found for the amplitudes of N400 during the second time window. In both language groups, the interaction among trial type, recording site and hemisphere was close to significance (normal-to-high group:  $F(1,14) = 4.57$ ,  $p = .051$ , and normal-to-low group:  $F(1,15) = 3.9$ ,  $p = .066$ ). A paired t-test confirmed a significant difference between trial types over the left frontal recording sites in normal-to-high producers,  $t(14) = 2.24$ ,  $p < .05$ . Mean amplitudes were more negative for inconsistent ( $-1.38 \mu\text{V}$ ,  $SD = 6.46$ ) than for consistent trials ( $0.99 \mu\text{V}$ ,  $SD = 6.37$ ). In the normal-to-low group, a paired t-test indicated a significant difference over the right frontal,  $t(15) = 2.21$ ,  $p < .05$ , and the right parietal,  $t(15) = -2.16$ ,  $p < .05$ , recording sites. Over the frontal recording sites, the mean amplitudes were more negative for the inconsistent ( $-4.48 \mu\text{V}$ ,  $SD = 8.05$ ) than for the consistent ( $-2.71 \mu\text{V}$ ,  $SD = 2.81$ ) trial types while over the parietal recording site, the amplitudes were more negative for the consistent ( $-2.71 \mu\text{V}$ ,  $SD = 2.88$ ) than for the inconsistent ( $-0.91 \mu\text{V}$ ,  $SD = 3.84$ ) trial type.

#### ***N400: third time window***

A significant interaction among trial type, recording site, hemisphere, and language group,  $F(1,29) = 5.85$ ,  $p < .05$  was found for the amplitudes of N400 during the third time window. Follow up ANOVAs were conducted separately for each group. The normal-to-low producers group presented an interaction among trial type, region site, hemisphere,  $F(1,14) = 4.5$ ,  $p = .051$  which was close to significant. Post-hoc t-tests confirmed a significant difference between trial types over the right parietal recording site,  $t(15) = 2.98$ ,  $p < .01$ . Amplitudes were more negative for the consistent ( $-2.93 \mu\text{V}$ ,  $SD = 3.27$ ) than for the inconsistent ( $-0.48 \mu\text{V}$ ,  $SD = 4.06$ ) trial type.

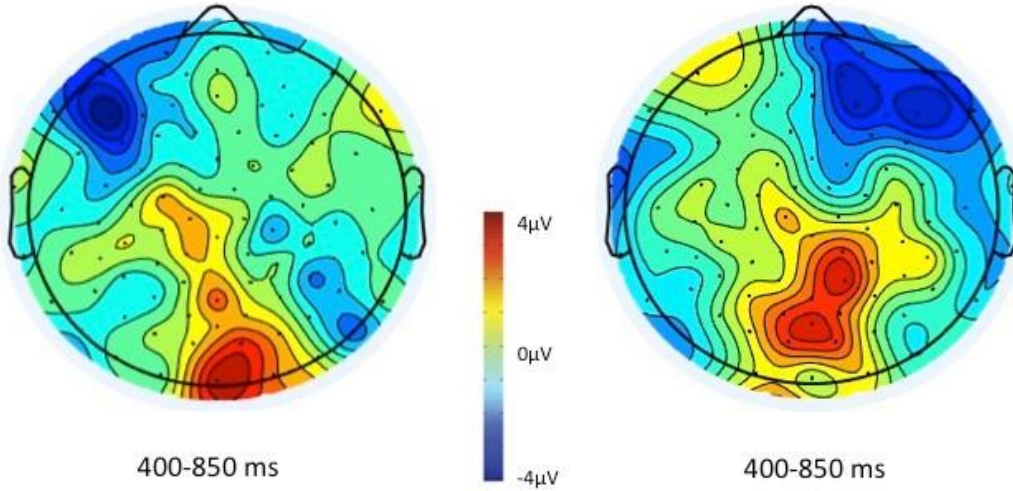
**A****B**

*Figure 2.* (A) The frontal and centro-parietal recording areas included in the statistical analyses over the left- and right hemispheres. Each area included 8 electrodes (illustrated in red). (B) Grand-averaged waveforms for consistent (black lines) and inconsistent (red lines) scene-word pairs in normal-to-high (left) and normal-to-low (right) producers. The difference waves between conditions are illustrated in blue lines. One example of recording site over both hemispheres and areas are illustrated. The approximate electrode placements according to the 10-20 international electrode position system are indicated (FC5, FC6, CP3 and CP4). The vertical line illustrates the target word onset. The light blue box indicates the significant difference between conditions.

### Scalp distributions of inconsistency effect (Inconsistent – Consistent picture-word pairs)

Normal-to-high producers

Normal-to-low producers



*Figure 3.* Topographical maps illustrating the distribution of N400 effect (average amplitudes measured from difference waves over the time window of 400 to 850 ms) in normal-to-high (left) and normal-to-low (right) producers



#### *4. Discussion*

The aim of the present study was to investigate whether visual semantic scene context primes word processing in 24-month-old-children and whether the priming effect is affected by individual vocabulary skills. Children were exposed to consistent and inconsistent scene- word pairs and their ERPs in response to target words were measured. The results showed that the amplitudes of the N400 component were more pronounced for inconsistent than for consistent scene-word pairs. However, both latencies and hemispheric distribution of the N400 component were modulated by language skills. Children with lower vocabulary skills exhibited a later N400 effect over the right frontal recording sites whereas in the group of children with higher vocabulary skills, the N400 effect was observed earlier over the left frontal sites. Our findings indicate that 2-year-olds have acquired enough knowledge about semantic regularities of their visual world to integrate contextual scene knowledge to subsequent word processing but language skills affect the occurrence and distribution of the N400 during contextual priming.

Previous studies regarding lexical-semantic development have shown that both associative (e.g., bunny and carrot) and taxonomic (e.g., dog and chicken) organizations of lexical-semantic system are established by the end of the second year of life (Arias-Trejo & Plunkett, 2009, 2013; Styles & Plunkett, 2009). In these studies, children looked longer the named target picture of “bunny” after first hearing a word “carrot” (associative relation) or a “chicken” (taxonomic relation). Equally, using the ERP technique, it was shown that children at the age of 20-months had developed a graded lexicon where basic-level words for objects from the same superordinate category have a closer relationship than basic-level words from different superordinate categories (Torkildsen et al., 2006). Additionally, there is evidence that by the age of 20 months, children are sensitive not only to basic-level categories but also to contextual categories (kitchen items versus bathroom items) (Mandler & Bauer, 1988). In our study, the target words (e.g., spoon, table, bed, pillow, tree...) were basic-level elements of scene categories (e.g., kitchen, bedroom, park...). These elements were chosen to represent typical, but not exclusive, items of a particular scene and they were never illustrated in a given scene. Thus, our results indicate that the priming effects were driven by semantic contextual knowledge associated with a particular scene rather than by item-specific taxonomic knowledge. This further suggests that at the age of 2-years children have acquired knowledge about scene categories and this abstract knowledge

contributes to their subsequent word recognition. In addition, our study demonstrated that children are able to extract semantic gist very quickly: a preview of a scene for 500 ms was enough to activate the representations of items semantically associated with the scene context. Unfortunately, the preview time was constant and thus, our results do not provide evidence how fast children are capable of extracting scene gist. In adults, the gist is recognized after a very short exposure (less than 100 ms) of visual scenes (Friedman, 1979; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Potter, 1975, 1976; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001).

The N400 effect in our study was found over the frontal recording sites while in previous developmental studies the N400 effect in object or word contexts has been predominantly found over the central-parietal recording sites (e.g. Friedrich & Friederici, 2005a, 2005b, 2005c, 2010; Rämä et al., 2013). There are, however, some studies reporting that both anterior and posterior regions are involved in processing of semantic incongruences (e.g., Friedrich & Friederici, 2004; Torkildsen et al., 2006) or that frontal rather than posterior regions are activated (Torkildsen et al., 2007) in children. It has been proposed that the frontally distributed activation is related to increased attentive demands (Torkildsen et al., 2006) or image-specific semantic processing (Friedrich & Friederici, 2004) in children. Obviously, in our study, children may have use more image-specific than linguistic processes, which probably explains the frontal distribution. Concerning the hemispheric distribution, the N400 effect has been found to be equally large over the left and the right hemispheres (e.g., Friedrich & Friederici, 2004), more pronounced over the right than the left hemisphere (Rämä et al., 2013) or temporally more extended over the left hemisphere (Friedrich & Friederici, 2004; Torkildsen et al., 2006). However, in children with higher vocabulary skills, the N400 effect has been found often over the right hemisphere (Friedrich & Friederici, 2004; 2010; Rämä et al., 2013; cf. Torkildsen et al., 2008) similarly to adults. Our current results, however, showed that the N400 effect in children with higher vocabulary skills was distributed over the left frontal recording sites while in children with lower vocabulary skills the effect was found over the right frontal recording sites. There is a possibility that the vocabulary effects in picture-word matching tasks are different from those in scene tasks. In picture-word tasks, children are presented with single objects while in our task children were presented with complex visual sceneries with multiple objects in a colorful background. Additionally, in picture-word tasks, the images are usually presented for several seconds and (at least partly) simultaneously with

spoken words. In contrast, in our study the preview of a scene was followed by a short break before the spoken word was delivered involving a higher requirement of working memory. Using working memory paradigms and functional brain imaging techniques, it has been suggested that the left hemisphere might be more involved in verbal and analytical processing while the right hemisphere is involved in more image-based rehearsal strategies (Courtney, Petit, Haxby, & Ungerleider, 1998; Haxby, Ungerleider, Horwitz, Rapoport, & Grady, 1995; Rämä, Sala, Gillen, Pekar, & Courtney, 2001). Thus, it is possible that children with different vocabulary skills used different rehearsal strategies in their task performance, that is, children with higher vocabulary skills were more engaged to verbal while children with lower vocabulary skills more to image-based strategies resulting in different hemispheric distribution in two language groups. Additionally, children with lower vocabulary skills exhibited a more negative late response for consistent than for inconsistent words over the right parietal recording sites. A similar congruence response, found earlier in 19-month-olds, tended to be more prominent in children with low- than with high-comprehension skills, suggesting that it might reflect a larger effort in children with lower language skills in accessing the meaning of words (Friedrich & Friederici, 2004).

To conclude, our results provided evidence that 24-month-old children have acquired enough semantic contextual knowledge associated with particular everyday scenes and they are capable of creating linguistic expectations based on this contextual knowledge. However, children with different linguistic skills might activate different neural structures underlying contextual priming in scene context.

### *Acknowledgements*

We thank all families for their participation and contribution to this research. This research was funded by the Sorbonne Paris Cite Grant, PME DIM Cerveau et Pensée 2013, and LABEX EFL (ANR-10-LABX-0083). A.H. was supported by doctoral fellowship from CONICYT, Chile.

## References

- Akhtar, N., Jipson, J., & Callanan, M. A. (2014). Learning words through overhearing. *Child Development, 72*(2), 416–430.
- Arias-Trejo, N., & Plunkett, K. (2009). Lexical-semantic priming effects during infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3633–3647. doi:10.1098/rstb.2009.0146
- Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition, 128*, 214–227. doi:10.1016/j.cognition.2013.03.008
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177. doi:10.1016/0010-0285(82)90007-X
- 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. doi: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. doi:10.1016/j.infbeh.2010.12.010
- Courtney, S. M., Petit, L., Haxby, J. V, & Ungerleider, L. G. (1998). The role of prefrontal cortex in working memory: examining the contents of consciousness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 353*, 1819–1828. doi:10.1098/rstb.1998.0334
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition, 35*(3), 393–401. doi:10.3758/BF03193280
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564. doi:10.1111/j.0956-7976.2004.00719.x
- Delle Luche, C., Durrant, S., Floccia, C., & Plunkett, K. (2014). Implicit meaning in 18-month-old toddlers. *Developmental Science, 17*(6), 948–955. doi:10.1111/desc.12164
- Duh, S., & Wang, S.-H. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology, 72*, 142–61. doi:10.1016/j.cogpsych.2014.03.001
- Friederici, A. D. (2006). The Neural Basis of Language Development and Its

Impairment. *Neuron*, 52(6), 941–952. doi:10.1016/j.neuron.2006.12.002

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology. General*. doi:10.1037/0096-3445.108.3.316

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, 16(8), 1465–77. doi:10.1162/0898929042304705

Friedrich, M., & Friederici, A. D. (2005a). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neuroreport*, 16(6), 653–656. doi:10.1097/00001756-200504250-00028

Friedrich, M., & Friederici, A. D. (2005b). Phonotactic knowledge and lexical-semantic processing in one-year-olds: brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, 17(11), 1785–1802. doi:10.1162/089892905774589172

Friedrich, M., & Friederici, A. D. (2005c). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. *Neuroreport*, 16(16), 1801–1804. doi:10.1097/01.wnr.0000185013.98821.62

Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, 114(2), 66–71. doi:10.1016/j.bandl.2009.07.004

Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, 40(4), 621–632. doi:10.1037/0012-1649.40.4.621

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, 16(2), 123–144. doi:10.1016/S0926-6410(02)00244-6

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for common sense<sup>2</sup>: an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8(2), 89–106. doi:10.1162/jocn.1996.8.2.89

Haxby, J. V., Ungerleider, L. G., Horwitz, B., Rapoport, S. I., & Grady, C. L. (1995). Hemispheric differences in neural systems for face working memory: A PET-rCBF study. *Human Brain Mapping*, 3, 68–82. doi:10.1002/hbm.460030204

Heise, N., & Ansorge, U. (2014). The roles of scene priming and location priming in object-scene consistency effects. *Frontiers in Psychology*, 5, 1–11. doi:10.3389/fpsyg.2014.00520

- 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. doi:10.4324/9780203488430
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243–271. doi:10.1146/annurev.psych.50.1.243
- Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, 19(2), 204–10. doi:10.3758/s13423-011-0205-7
- 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. doi:10.3758/BF03197475
- Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychologica*, 102(2-3), 319–343. doi:10.1016/S0001-6918(98)00053-5
- Josephs, E. L., Draschkow, D., Wolfe, J. M., & Vö, M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica*, 169, 100–108. doi:10.1016/j.actpsy.2016.05.013
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, 47(26), 3286–3297. doi:10.1016/j.visres.2007.09.013
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62(August), 621. doi:10.1146/annurev.psych.093008.131123
- Kutas, M., & Van Petten, C. (1994). ERP Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA. doi:10.1016/B978-012369374-7/50018-3
- Mandler, J. M., & Bauer, P. J. (1988). The cradle of categorization: Is the basic-level basic? *Cognitive Development*, 3, 247–264.
- 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. doi:10.1037//0278-7393.2.5.529
- Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, 48(2), 507–517. doi:10.1016/j.neuropsychologia.2009.10.011

- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science*, *6*(2), 136–142. doi:10.1111/1467-7687.00263
- Oliva, A. (2005). Gist of the scene. In *Neurobiology of Attention* (pp. 251–256). doi:10.1016/B978-012375731-9/50045-8
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. doi:10.1016/j.tics.2007.09.009
- Palmer, S. (1975). Visual perception and world knowledge: notes on a model of sensory-cognitive interaction. In D. Norman & D. Rumelhart (Eds.), *Explorations in Cognition* (LNRRes. Gr., pp. 279–307). San Francisco: Freeman.
- 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.
- Rämä, P., Sala, J. B., Gillen, J. S., Pekar, J. J., & Courtney, S. M. (2001). Dissociation of the neural systems for working memory maintenance of verbal and nonspatial visual information. *Cognitive, Affective & Behavioral Neuroscience*, *1*(2), 161–71. doi:10.3758/CABN.1.2.161
- Rämä, P., Sirri, L., & Serres, J. (2013). Development of lexical–semantic language system: N400 priming effect for spoken words in 18- and 24-month old children. *Brain and Language*, *125*(1), 1–10. doi:10.1016/j.bandl.2013.01.009
- Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: Evidence from eye tracking. *Developmental Science*, *12*(4), 549–556. doi:10.1111/j.1467-7687.2009.00795.x
- Sirri, L., & Rämä, P. (2015). Cognitive and neural mechanisms underlying semantic priming during language acquisition. *Journal of Neurolinguistics*, *35*, 1–12. doi:10.1016/j.jneuroling.2015.01.003
- Styles, S. J., & Plunkett, K. (2009). How do infants build a semantic system? *Language and Cognition*, *1*(01), 1–24. doi:10.1515/LANGCOG.2009.001
- Thorpe, S., Fize, D., & Marlot, C. (1996). Thorpe1996.pdf. *Letters to Nature*, *381*, 520–522.
- Tomasello, M., Strosberg, R., & Akhtar, N. (2008). Eighteen-month-old children learn words in non-ostensive contexts\*. *Journal of Child Language*, *23*(1), 157–176. doi:10.1017/S0305000900010138
- Torkildsen, J. V. K., Sannerud, T., Syversen, G., Thormodsén, R., Simonsen, H. G., Moen, I., Lindgren, M. (2006). Semantic organization of basic-level words in 20-month-olds:



An ERP study. *Journal of Neurolinguistics*, 19(6), 431–454.  
doi:10.1016/j.jneuroling.2006.01.002

Torkildsen, J. V. K., Svangstu, J. M., Hansen, H. F., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience*, 20(7), 1266–1282. doi:10.1162/jocn.2008.20087

Torkildsen, J. V. K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, 20(4), 332–351. doi:10.1016/j.jneuroling.2007.02.003

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30, 655–668. doi:10.1068/p3029

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, 10(3), 14.1–13. doi:10.1167/10.3.14

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential ERP signatures elicited by semantic and syntactic processing in scenes. *Psychological Science*, 24(9), 1816–1823. doi:10.1177/0956797613476955

## *General Discussion*

## ***Chapter I: Development of eye movement patterns during scene exploration from infancy to childhood***

Eye movement behaviour during scene exploration was examined in infants and children at different ages ranging from 3 months to 10 years of age in Studies I and II. In order to characterize the development of eye movement behaviour, oculomotor parameters (fixation durations and saccade amplitudes) and scanning strategies were compared between age groups. Saccade amplitudes reflect gaze distribution over a scene while fixation durations reflect on-going processing of visual stimuli (Castelhano, Mack, & Henderson, 2009; Henderson & Pierce, 2008; Mills, Hollingworth, & Dodd, 2011; Tatler & Vincent, 2008). Therefore, these two parameters were used as representative measures of oculomotor functions in the current studies. The results from Studies I and II demonstrated that during free exploration of scenes, fixation durations decrease from 12 months to 6- to 8-years of age. Previous evidence from several empirical studies with adults (Henderson & Hollingworth, 1999; Henderson & Pierce, 2008; Mills et al., 2011; van Diepen, Ruelens, & d'Ydewalle, 1999) and with infants (Bronson, 1990, 1991, 1994; Colombo, Freeseaman, Coldren, & Frick, 1995; Colombo, Mitchell, Coldren, & Freeseaman, 1991; Wass & Smith, 2014) have revealed that longer fixation durations are associated with difficulties in visual information retrieval and with higher cognitive effort. Therefore, the present results suggest that visual processing proficiency during scene exploration improves with age. However, fixation durations have also been associated with oculomotor programming of the next saccade in adults (Nuthmann, Smith, Engbert, & Henderson, 2010; Walshe & Nuthmann, 2013, 2015) and there is empirical evidence using artificial stimuli that saccade latencies decrease with age (Fukushima, Hatta, & Fukushima, 2000; Gredebäck, Örnkloo, & von Hofsten, 2006; Irving, Steinbach, Lillakas, Babu, & Hutchings, 2006; Klein & Foerster, 2001; Luna & Seeney, 2004; Munoz, Broughton, Goldring, & Armstrong, 1998). Thus, the decrease of fixation durations with age found in the current studies might have also reflected a decrease of saccade latency with age. Since in the current studies participants performed a free exploration of scenes, saccade latencies could not be measured. However, in previous studies saccade latencies have shown to be stabilized by 14-15 years of age while in the Study II fixation durations reached adult-like values by 6-8 years (Fukushima et al., 2000; Gredebäck et al., 2006; Irving et al., 2006; Klein & Foerster, 2001; Luna & Sweeney, 2004; Luna, Velanova, & Geier, 2008; Munoz et al., 1998). The

differences in the ages of stabilization of saccade latencies in previous studies and fixation duration in the current studies suggest that decreasing fixation durations found in Studies I and II cannot be explained only by maturation of oculomotor programming and that increasing cognitive control might be involved.

Initial fixation duration has been suggested to reflect the time needed for processing the scene context and planning the subsequent eye movement (Castelhano & Henderson, 2007; Chen & Zelinsky, 2006; Mills et al., 2011; Torralba, Oliva, Castelhano, & Henderson, 2006). In infants, first fixation durations decreased already from 3 to 6-month-olds to 9 to 12 months of age and in children, 2 year-olds presented the longest first fixation duration, and the durations decreased with age up to 6 to year-olds. These results are partly different from those obtained for global fixation durations where developmental differences were found later, at 12-months of age. This may suggest that infants are capable of extracting scene context early in development affecting initial fixation durations even the maturation of oculomotor parameter (general fixation durations) start later.

Saccade amplitudes did not differ between infants and adults. However, saccades of 12-month-olds were shorter than those of younger infants (Study I). In children, saccade amplitudes increased with age from 2 to 4-6 years of age (Study II). In addition, the mean saccade amplitudes in 12-month-old infants in Study I were longer than the mean amplitudes in 2-year-old children in Study II while saccade amplitudes in adults were similar in both studies. These results from Studies I and II suggest a U-shaped development from infancy to childhood with saccade amplitudes decreasing from 9 months to 2 years of age to increase again from 2 to 6 years. The results in infants are, however, not in concordance with previous studies showing that infants perform often hypometric saccades when shifting their gaze from central towards peripheral targets (Aslin & Salapatek, 1975; Regal, Ashmead, & Salapatek, 1983; Roucoux, Culee, & Roucoux, 1983). In contrast, the results in children agree with earlier studies showing that young children had shorter saccade amplitudes in pro- and anti- saccade tasks compared with adults and adult-like values were often reached by 10 years of age (Fioravanti, Inchingolo, Pensiero, & Spanio, 1995; Irving et al., 2006; Irving, Tajik- Parvinchi, Lillakas, González, & Steinbach, 2009; Munoz et al., 1998). Based on these earlier findings, it is not straightforward to find an explanation why saccade amplitudes in the current studies showed a U-shaped curve. There is a possibility that the age-associated changes in saccade amplitudes are different when applying free exploration than artificial tasks, further suggesting that maturation of cognitive control of saccades is task-specific, at least at early stages of development.

Further studies addressing the development of saccade control during scene viewing are, however, needed.

To investigate the development of scanning strategies (ambient and focal modes), fixation durations and saccade amplitudes at early and late phases of viewing time were compared between age groups. The results showed that fixation duration increased with viewing time in infants and children from 12 months of age as in adults, but not in younger infants. An increase in fixation durations during viewing time has been previously associated with focal strategies (Norman, 2002; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005; Velichkovsky, Rothert, Dornhoefer, & Joos, 2002). It has been proposed that the increasing fixation durations reflect a transition from a fast exploration (short fixations) to a more detailed processing of the visual environment over time (Norman, 2002; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005; Velichkovsky, Rothert, Dornhoefer, & Joos, 2002). Therefore, the results of Study I suggest that adult-like scanning strategies emerge by the end of the first year of life. However, when the time course of saccade amplitude was analysed in Study I, neither infants nor adults presented the time course associated with ambient and focal modes even this pattern was present in all age groups of children as well as in adults in the Study II. Infants from 3 to 9 months of age and adults exhibited no differences in saccade amplitudes between the two viewing phases while in the 12-month-olds group, saccade amplitudes increased significantly between the two phases. Previously, the transition of saccade amplitudes through the viewing time has shown to be modified by task type and visual information (Casthelano et al., 2009; Mills et al., 2011). For instance, some studies using free viewing tasks have shown that saccade amplitudes remain stable over the viewing time (Casthelano et al., 2009; Mills et al., 2011). This time course of saccade amplitudes has been related to less requirements for visual information processing in free viewing compared with other tasks (Mills et al., 2011). This may also partly explain the lack of significant increase in saccade amplitudes in adults in Study I. To ensure that the experimental design was suitable for young infants, they were presented with only six simple images (only a few objects on a simple background). Also, adults performed a free viewing task. In contrast, in Study II, the experiment consisted of 30 complex images (several objects on a complex background) and adults (as well as older children) performed a simple memory task. Thus, it is possible that the type, the amount of visual stimuli as well as the absence of an explicit goal in Study I made the task too simple for adults. As a result, the increase of saccade amplitudes over time was reduced in this

age group. Contrary to adults, visual processing of visual scenes, even if they are simple, might be a demanding and an interesting task for infants. Previous studies have also shown that saccade amplitudes increase in response to visual processing difficulties, suggesting that longer saccadic amplitudes can be used as a strategy to cover more regions of the scene when visual information retrieval is difficult (Casthelano et al., 2009; Mills et al., 2011). Therefore, 12-month-olds might have adjusted their oculomotor behaviour to improve visual processing by increasing saccade amplitudes over the time.

Altogether, current results indicate that attentional strategies during scene exploration emerge by 12 months of age. Previous evidence showing that endogenous attention develops late in the first year of life (Colombo, 2001). By this age, infants shift from more external (i.e. bottom-up) to more internal (i.e. top-down) control of visual attention allowing them to inhibit attentional distractors, alternate between two stimuli, and maintain vigilance. It is probable that this newly developing ability also allows them to adjust their attentional strategies to the requirements of visual tasks. Even attentional strategies appeared by the end of the first year of life, results of the Studies I and II showed that the development of oculomotor parameters associated to scene exploration such as fixation durations and saccade amplitudes are stabilized much later (around 6 years). These developmental differences suggest that the two modes are of fundamental importance for our survivor since they are present even before accurate oculomotor and cognitive control is achieved. These findings along with previous evidence showing the ambient and focal scanning strategies are present in nonhuman primates, various other animal species (Ingle, Schneider, Trevarthen, & Held, 1967; Trevarthen, 1968) as well as when different types of stimuli (e.g., high versus low density, emotional pictures, landscapes) are presented, support the idea that the existence of ambient and focal modes is a fundamental feature of the visual system.

In order to examine whether scanning strategies were similar between age groups, total scan path lengths and the proportion of ambient and focal fixations were also analysed. The scan path lengths were shorter and the proportion ambient fixations were lower in infants and in children compared with adults up to 6 years of age while no differences between age groups were found for the proportion of focal fixations. These results suggest that developmental changes in scanning behaviour are associated with the emergency and maturation of ambient rather than focal strategies. Furthermore, the results indicate that the ambient mode is associated to a more scattered exploration of images. These findings are in line with studies showing that children switch from a local to a global bias in the

processing of hierarchical visual forms from 4 years of age to 6-9 year-old (e.g. Dukette & Stiles, 1996, 2001; Poirel, Mellet, Houdé, & Pineau, 2008; Vinter, Puspitawati, & Witt, 2010). Therefore, it is plausible that the later representation of ambient processing in children is related to the emergency of global processing (e.g. Dukette & Stiles, 1996, 2001; Poirel et al., 2008; Vinter et al., 2010), suggesting a general shift on visual processing by this age.

## ***Chapter II: Interactions between cognitive and perceptual influences on gaze allocation during scene viewing***

To date, it is recognized that bottom-up guidance of gaze can be modulated by top-down control. For instance, saliency is shown to be a better predictor of gaze allocation in conditions in which top-down modulation is proposed to be lower, such as during early than late stages of viewing time (Castelhano et al., 2009; Mannan, Ruddock, & Wooding, 1995; Parkhurst, Law, & Niebur, 2002; Tatler & Vincent, 2008), or during free viewing or memory compared with searching tasks (Henderson et al., 2007; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006), as well as when images illustrate relatively neutral semantic contents (e.g., nature) compared with images containing semantically relevant elements (e.g. faces or man-made elements) (Nyström & Holmqvist, 2008). In addition, cognitive abilities develop from birth until adulthood (e.g. Gathercole, 1999; Gathercole, Pickering, Ambridge, & Wearing, 2004; Hitch, Halliday, Schaafstal, & Schraagen, 1988; Klenberg, Korkman, & Lahti-Nuuttila, 2010; Pearson & Lane, 1991; Pickering, 2001; Sanders, Stevens, Coch, & Neville, 2006), and thus, it is likely that top-down control also increase with increasing age.

In line with previous findings, the current results in children showed that saliency map predicted better gaze allocation during the early than the late viewing phase in all age groups. This finding supports previous evidence showing that the early stage of viewing time is primarily driven by image properties (Castelhano et al., 2009; Mannan et al., 1995; Parkhurst et al., 2002; Tatler & Vincent, 2008). However, the results showed that the match between gaze distribution and saliency map prediction was better for children from 2 to 6 years than for older participants. This finding is in accordance with the idea that perceptual guidance decreases when cognitive control increases. An earlier study that presented children with various types of visual stimuli (e.g., fractals, pink noise, man-made, and nature scenes) found a stronger general influence of saliency, compared with adults, in even older children (7 to 9 years of age) than in Study II (Ackik et al., 2010). However, in this previous study differences between adults and children were not found for all scene types: for example, adults and children performed similarly when viewing nature scenes but not when viewing scenes containing man-made objects. Together, the results of the current and this previous study suggest that by middle-childhood gaze control has reached a certain level of maturation in comparison with younger children (younger than 6



year-olds) but it has not yet reached an adult-like level since when images with high level of semantic context are presented, the differences with adults remain.

Interactions of top-down and bottom-up factors during scene viewing were more directly examined in 24-month-olds in Study III. In that study, perceptual features (saliency) and semantic content of the scene (scene-object consistency) were manipulated. Participants were presented with real world scenes containing either semantically consistent or inconsistent objects. The results indicated that by 24 month of age, children have acquired sufficient knowledge of semantic scene regularities to retrieve the gist of a scene and use the context in further processing (e.g., scene-object processing). This result is in line with a previous finding obtained in a change blindness paradigm showing that even younger children (15-month-olds) were able to detect semantic scene-object inconsistencies (Duh & Wang, 2014). After a familiarization phase, children looked longer to the screen when an object was replaced by an inconsistent than by a consistent new object. However, the results of the current thesis provide the first evidence that young children can retrieve scene context and use this information to guide their eye movements also during free exploration. Further, in the study by Duh and Wang (2014) children detected semantic inconsistencies only when the scene preview time was long but not when it was shorter. Authors suggested that during shorter exposure times, 15-month-olds were not provided enough time to extract the gist of the scene. In contrast, studies in adults have shown that scene identification occurs as fast as 100 ms (Potter, 1975, 1976; Schyns & Oliva, 1994). Even the design of Study III did not allow to determine how fast the gist was extracted, the results of Study IV showed that an N400 incongruence effect was obtained even after a short preview (500 ms) of a scene in a scene-word priming task. This finding suggests that at this age children are able to retrieve the gist of the scene in a relatively short time.

Semantic inconsistency did not modulate early eye movements either in children or in adults. This finding further supports the idea that foveal processing is necessary to process semantic object–scene inconsistencies during scene viewing (De Graef et al., 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Võ & Henderson, 2009; Võ & Henderson, 2011). Once the object was fixated, both adults and children looked longer at inconsistent objects already during their first gaze pass to the area of interest. This result, previously reported in adults (Friedman, 1979; Henderson et al., 1999; Võ, M & Henderson, 2009), suggests that children are able to integrate scene context and object identity relatively fast.

The present results also showed that both adults and young children looked longer at high salient than at low salient objects. However, children allocated their gaze earlier at high

than at low salient objects while adults did not exhibit the same effect. Adult results corroborate previous findings showing the saliency effect on total gaze durations (Clarke, Coco, & Keller, 2013; Coco, Malcolm, & Keller, 2014; Itti & Koch, 2000) while the results in children agree with previous developmental studies, including Study II, showing a strong effect of saliency in young children (Açik, Sarwary, Schultze-Kraft, Onat, & König, 2010; Helo, Pannasch, Sirri, & Rämä, 2014; Kooiker, Steen, & Pel, 2016). However, the results of Study III showed that developmental differences on the perceptual guidance of gaze were obtained only in extrafoveal parameters suggesting that semantic consistency effect modulated the perceptual influences of saliency on looking times in young children. Current results did not find an interaction between saliency and semantic processing after the objects were fixated in either of the two age groups. Both children and adults looked longer at inconsistent than consistent objects in both saliency conditions. This finding provides additional support to the previous literature in adults showing that saliency does not affect semantic inconsistency detection (Coco et al., 2014; Foulsham & Underwood, 2008; Underwood & Foulsham, 2006) and add new evidence that semantic inconsistency attract gaze allocation irrespective of saliency also in young children. Altogether, the results of Study III provide new evidence that differences in perceptual guidance between age groups is mainly associate to the early guidance of eye movements rather than differences in amount of fixation landings on salient objects.

### ***Chapter III: Effect of linguistic skills on scene processing and eye movement guidance in young children***

There is a large amount of evidence demonstrating that language and visual processing are intimately related, and these two modalities influence each other (Eberhard, Spivey-Kowlton, Sedivy, & Tanenhaus, 1995; Ferreira & Tanenhaus, 2007; Henderson & Ferreira, 2004; Spivey, Tyler, Eberhard, & Tanenhaus, 2001; Tanenhaus, Spivey-Kowlton, Eberhard, & Sedivy, 1995; Tanenhaus & Trueswell, 2006). In addition, during the second year of life there is an extensive improvement in vocabulary skills – so-called vocabulary spurt (Ganger & Brent, 2004; Nazzi & Bertoncini, 2003). However, there is a large individual variation during this period and the vocabulary size can vary from few words to hundreds of words at 24 months of age. These individual differences allow dividing children into language groups, normal-to-low or normal-to-high producers, based on their vocabulary skills to further investigate the effect of linguistic skills on other cognitive functions, such as scene exploration.

The effect of linguistic skills on semantic processing of scenes and eye movement behaviour was investigated in Study III, that is, semantic consistency and saliency influences on gaze allocation were compared between children having either high or low vocabulary skills. The results showed that both language groups looked equally long to inconsistent objects, passing around 20% of their whole viewing time on these objects. Adults also used the same proportion of time to explore inconsistent objects. However, normal-to-high producers exhibited significantly longer looking times to consistent objects than normal-to-low-producers, and thus the difference between conditions were observed only in children with low vocabulary skills. Both groups presented similar initial fixation durations at the scene suggesting that all children exhibited a similar cognitive effort in understanding the scene context. They also had a similar total trial dwell time indicating that they were equally engaged to the task. Likewise, the total path length was similar between groups suggesting that they were similarly likely to distribute their gaze over the scene. The only difference in oculomotor parameters between both language groups was that the high producers looked longer to consistent objects. Empirical evidence in adults has shown that linguistic parameters such as semantic proximity (i.e. co-occurrence of words/objects in a similar context) and word frequency attracted the gaze when participants were asked to name the objects presented in the scene (Clarke et al., 2013; Coco et al., 2014). Furthermore, when

scene exploration was combined with a naming task, looking times to consistent objects increased and they looked even longer than inconsistent objects (Clarke et al., 2013). The results of the current study suggest that high producers might have silently named some of the objects during scene exploration and consequently looked longer the consistent objects than low producers. Implicit naming has shown to be automatically activated during exploration of visual displays in 18 and 24 month-olds (Mani & Plunkett, 2010, 2011) and it has shown to be stronger in 24-month-olds than in adults (Khan, 2013). In addition, there is a study showing that at two years of age children are prone to name novel objects (Samuelson & Smith, 2005) and the amount of spontaneous naming during manipulation of novel objects was associated with productive vocabulary size (Samuelson & Smith, 2005). These findings give further support to the idea that implicit naming increases language-driven exploration of visual scenes in children with higher vocabulary skills.

The ERPs results also give further support to the idea that language processes are activated during scene processing, especially in normal-to-high-producers. Results of this study showed that children with higher vocabulary skills exhibited a N400 effect over the left frontal sites whereas in the group of children with lower vocabulary skills the N400 effect was observed over the right hemisphere. Previously, it has been suggested that the left hemisphere might be more involved in verbal and analytical processing while the right hemisphere is involved in more image-based rehearsal strategies (Courtney, Petit, Haxby, & Ungerleider, 1998; Haxby, Ungerleider, Horwitz, Rapoport, & Grady, 1995; Rämä, Sala, Gillen, Pekar, & Courtney, 2001). Thus, it is possible that children with higher vocabulary skills were more likely to be engaged to verbal processing (naming) while children with lower vocabulary skills were more engaged to image-based strategies. Altogether, the present results indicate that 24-month-olds with high and low vocabulary skills use scene context to further processing of objects and words. However, children with different levels of vocabulary skills might use different visual-driven or linguistic-driven strategies when exploring a scene, which in turns affects their visual attention and the distribution of the ERP responses.

## *Chapter IV: Conclusions and future perspectives*

The present thesis provides novel evidence for developmental aspects of eye movement behaviour during scene viewing and semantic scene processing in young children. Based on the results of the current thesis, two important developmental stages in eye movement behaviour can be identified. The first developmental stage occurs at the end of the first year of life when infants exhibit two key changes in their eye movement control: First, oculomotor parameters start to differentiate from younger infants. Both fixation durations and saccade amplitudes were longer in 12-month-olds compared with younger infants while no differences among younger infants were found. Second, attentional strategies start to emerge by this age; this was reflected by a shift in fixation durations and saccade amplitudes over time. After the first year of life, different oculomotor parameters further develop and the second important stage can be identified around 6 to 8 years of age when fixation durations achieve adult-like values. At this age also ambient scanning strategies (proportion of ambient fixations), scene exploration proficiency (path length) as well as bottom-up influences (saliency map prediction) reach a similar profile than in adults.

The findings of this thesis also indicate that 24-month-old children, independent of their vocabulary skills, are able to retrieve the gist of the scene and also to use scene context in further processing of visual objects or spoken words. Saliency guidance has stronger effects on early eye movements in children than in adults. Regarding these findings, I propose that in children saliency guides early eye movements and they process the semantic scene content once they have reached these salient regions. However, foveal processing of semantic inconsistency occurs independent of saliency. Additionally, I propose that vocabulary skills, previously associated with implicit naming ability, affect eye movement distribution over the scene, possibly attracting the gaze to linguistically relevant (consistent) objects. Likewise, according to the ERP results, linguistic skills seem to affect the speed of visual-linguistic integration as reflected by an earlier N400 response in normal-to-high than normal-to-low producers.

There were, however, some limitations in the current thesis and some important questions remain for future investigations. Visual scenes used in Study II were colour pictures taken from children books or movies, and therefore even they were likely to follow fundamental rules of natural scenes (e.g., probability of appearance, support, or interposition), they were not representing the real world. Following studies would be

needed to investigate whether the results were replicable if real world pictures were used. In addition, scenes used in Study I and II were not the same, and thus, an exhaustive characterization of developmental transition from infancy to early childhood was not possible. It remains to be studied, using exactly the same experimental setting, whether infants and young children follow the same developmental curves demonstrated in the current thesis. Furthermore, ambient and focal fixations were defined based on previous values described in adults, and it remains to be established whether there are specific values for classifying fixation types in infants and children.

Finally, interesting questions such as *when the ability to use scene context to guide visual attention emerge in infants?* or *when children are able to extract the scene gist as fast as adults?* are still opened. Likewise, new experimental paradigms need to be developed to investigate the influence of language on gaze guidance in children. One proposal design to address these questions would be the manipulation of scene preview time and the inclusion of linguistic tasks while children explore the scene.

## References

- Açık, 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://doi.org/10.3389/fpsyg.2010.00207>
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439. <http://doi.org/10.1006/jmla.1997.2558>
- 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://doi.org/10.1016/j.actpsy.2011.01.007>
- Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103*(1), 62–70.
- Aring, E., Grönlund, M. A., Hellström, A., & Ygge, J. (2007). Visual fixation development in children. *Graefe's Archive for Clinical and Experimental Ophthalmology, 245*(11), 1659–1665. <http://doi.org/10.1007/s00417-007-0585-6>
- Armstrong, V., Maurer, D., & Lewis, T. L. (2009). Sensitivity to first- and second-order motion and form in children and adults. *Vision Research, 49*(23), 2774–2781. <http://doi.org/10.1016/j.visres.2009.08.016>
- Aslin, R. N., & Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Perception & Psychophysics, 17*(3), 293–302. <http://doi.org/10.3758/BF03203214>
- Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Report, 14*(6), 468–469.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*, 617–629. <http://doi.org/10.1038/nrn1476>
- Barlett, F. C. (1932). *Remembering: A study in experimental and social psychology* (Cambridge). Cambridge, UK.
- 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://doi.org/10.1037/0096-1523.33.1.20>
- Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception & Psychophysics, 70*(8), 1444–1458. <http://doi.org/10.3758/PP.70.8.1444>
- Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*, 77–80.

- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–233). Hillsdale, N.J.: Lawrence Erlbaum.
- Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology*, *97*(1), 22–27.
- 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://doi.org/10.1016/0010-0285\(82\)90007-X](http://doi.org/10.1016/0010-0285(82)90007-X)
- Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology*. <http://doi.org/10.1037/h0037158>
- Bloch, H., & Carchon, I. (1992). On the onset of eye-head coordination in infants. *Behavioural Brain Research*, *49*(1), 85–90. [http://doi.org/10.1016/S0166-4328\(05\)80197-4](http://doi.org/10.1016/S0166-4328(05)80197-4)
- 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://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://doi.org/10.1016/j.infbeh.2010.12.010>
- Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research*, *51*(13), 1588–1609. <http://doi.org/10.1016/j.visres.2011.02.018>
- Braddick, O., Atkinson, J., & Wattam-Bell, J. (2003). Normal and anomalous development of visual motion processing: Motion coherence and “dorsal-stream vulnerability.” *Neuropsychologia*, *41*(13), 1769–1784. [http://doi.org/10.1016/S0028-3932\(03\)00178-7](http://doi.org/10.1016/S0028-3932(03)00178-7)
- Bronson, G. (1990). Changes in infants’ visual scanning across the 2- to 14-week age period. *Journal of Experimental Child Psychology*, *49*(1), 101–125. [http://doi.org/10.1016/0022-0965\(90\)90051-9](http://doi.org/10.1016/0022-0965(90)90051-9)
- Bronson, G. (1991). Infant differences in rate of visual encoding. *Child Development*, *62*(1), 44–54. <http://doi.org/10.2307/1130703>
- Bronson, G. (1994). Infants’ transitions toward adult-like scanning. *Child Development*, *65*(5), 1243–1261. <http://doi.org/10.2307/1131497>
- Buswell, G. T. (1935). How people look at pictures: A study of the psychology of perception in art. *Chicago: University of Chicago Press*. <http://doi.org/citeulike-article-id:3847289>
- Carpenter, R. H. S. (1988). *Movements of the eyes*. London: Pion.
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin &*



*Review*, 18(5), 890–896. <http://doi.org/10.3758/s13423-011-0107-8>

- 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://doi.org/10.1037/0096-1523.33.4.753>
- Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal Experimental Psychology*, 62(1), 1–14. <http://doi.org/10.1037/1196-1961.62.1.1>
- 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://doi.org/10.1167/9.3.6.Introduction>
- Chandna, A. (1991). Natural history of the development of visual acuity in infants. *Eye*, 5, 20–26. <http://doi.org/10.1038/eye.1991.4>
- Chen, L., Meier, K. M., Blair, M. R., Watson, M. R., & Wood, M. J. (2013). Temporal characteristics of overt attentional behavior during category learning. *Attention, Perception & Psychophysics*, 75(2), 244–56. <http://doi.org/10.3758/s13414-012-0395-8>
- 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://doi.org/10.3389/fpsyg.2013.00927>
- 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), 1096–120. <http://doi.org/10.1080/17470218.2013.844843>
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, 52, 337–67.
- Colombo, J., Freeseaman, L. J., Coldren, J. T., & Frick, J. E. (1995). Individual differences in infant fixation duration: Dominance of global versus local stimulus properties. *Cognitive Development*, 10(2), 271–285. [http://doi.org/10.1016/0885-2014\(95\)90012-8](http://doi.org/10.1016/0885-2014(95)90012-8)
- Colombo, J., Frick, J. E., Ryther, J. S., & Gifford, J. J. (1996). Four-month-olds' recognition of complementary-contour forms. *Infant Behavior and Development*, 19(1), 113–119. [http://doi.org/10.1016/S0163-6383\(96\)90049-1](http://doi.org/10.1016/S0163-6383(96)90049-1)
- Colombo, J., & Horowitz, F. D. (1987). Behavioral state as a lead variable in neonatal research. *Merrill-Palmer Q*, 33(4), 423–437.
- Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseaman, L. J. (1991). Individual differences in infant visual attention: are short lookers faster processors or feature processors? *Child Development*, 62(6), 1247–1257. <http://doi.org/10.1111/j.1467-8624.1991.tb01603.x>
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken Language.

*Cognitive Psychology*, 107(1), 84–107. [http://doi.org/10.1016/0010-0285\(74\)90005-X](http://doi.org/10.1016/0010-0285(74)90005-X)

- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, 58(3), 306–24. <http://doi.org/10.1016/j.neuron.2008.04.017>
- Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' attention to patterned stimuli: Developmental change from 3 to 12 months of age. *Child Development*, 77(3), 680–695. <http://doi.org/10.1111/j.1467-8624.2006.00897.x>
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of Vision*, 10(4), 1–17. <http://doi.org/10.1167/10.4.16>
- Daniel, B. M., & Lee, D. N. (1990). Development of looking with head and eyes. *Journal of Experimental Child Psychology*, 50(2), 200–216. [http://doi.org/10.1016/0022-0965\(90\)90039-B](http://doi.org/10.1016/0022-0965(90)90039-B)
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35(3), 393–401. <http://doi.org/10.3758/BF03193280>
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559–564. <http://doi.org/10.1111/j.0956-7976.2004.00719.x>
- De Graef, P., Christiaens, D., & D'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, 52, 317–329.
- Dekker, T., Mareschal, D., Sereno, M. I., & Johnson, M. H. (2011). Dorsal and ventral stream activation and object recognition performance in school-age children. *NeuroImage*, 57(3), 659–670. <http://doi.org/10.1016/j.neuroimage.2010.11.005>
- Duh, S., & Wang, S.-H. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology*, 72, 142–61. <http://doi.org/10.1016/j.cogpsych.2014.03.001>
- Dukette, D., & Stiles, J. (1996). Children's analysis of hierarchical patterns: Evidence from a similarity judgment task. *Journal of Experimental Child Psychology*, 63(1), 103–140. <http://doi.org/10.1006/jecp.1996.0044>
- Dukette, D., & Stiles, J. (2001). The effects of stimulus density on children's analysis of hierarchical patterns. *Developmental Science*, 4(2), 233–251. <http://doi.org/10.1111/1467-7687.00168>
- Eberhard, K. M., Spivey-Kowlton, J. C., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24(6), 409–436.
- Egami, C., Morita, K., Ohya, T., Ishii, Y., Yamashita, Y., & Matsuishi, T. (2009). Developmental characteristics of visual cognitive function during childhood according to exploratory eye movements. *Brain and Development*, 31(10), 750–757. <http://doi.org/10.1016/j.braindev.2008.12.002>

- 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://doi.org/10.1167/8.2.2>
- Ferreira, F., & Tanenhaus, M. K. (2007). Introduction to the special issue on language-vision interactions. *Journal of Memory and Language*, 57(4), 455–459. <http://doi.org/10.1016/j.jml.2007.08.002>
- Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, 22(4), 661–674. <http://doi.org/10.1017/S0140525X99002150>
- Fioravanti, F., Inchingolo, P., Pensiero, S., & Spanio, M. (1995). Saccadic eye movement conjugation in children. *Vision Research*, 35(23–24), 3217–3228. [http://doi.org/10.1016/0042-6989\(95\)00152-5](http://doi.org/10.1016/0042-6989(95)00152-5)
- 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://doi.org/10.3389/fnsys.2013.00017>
- Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia*, 47(1), 248–257. <http://doi.org/10.1016/j.neuropsychologia.2008.07.016>
- Follet, B., Le Meur, O., & Baccino, T. (2011). New insights into ambient and focal visual fixations using an automatic classification algorithm. *I-Perception*, 2(6), 592–610. <http://doi.org/10.1068/i0414>
- 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://doi.org/10.1167/8.2.6>
- Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology. General*. <http://doi.org/10.1037/0096-3445.108.3.316>
- Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, 16(8), 1465–77. <http://doi.org/10.1162/0898929042304705>
- Friedrich, M., & Friederici, A. D. (2005a). Phonotactic knowledge and lexical-semantic processing in one-year-olds: brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, 17(11), 1785–1802. <http://doi.org/10.1162/089892905774589172>
- Friedrich, M., & Friederici, A. D. (2005b). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. *Neuroreport*, 16(16), 1801–1804. <http://doi.org/10.1097/01.wnr.0000185013.98821.62>

- Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, 114(2), 66–71. <http://doi.org/10.1016/j.bandl.2009.07.004>
- Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements. *Brain and Development*, 22(3), 173–180. [http://doi.org/10.1016/S0387-7604\(00\)00101-7](http://doi.org/10.1016/S0387-7604(00)00101-7)
- 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://doi.org/10.1016/S0387-7604\(00\)00101-7](http://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://doi.org/10.1037/0012-1649.40.4.621>
- Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, 16(2), 123–144. [http://doi.org/10.1016/S0926-6410\(02\)00244-6](http://doi.org/10.1016/S0926-6410(02)00244-6)
- Gathercole, S. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3(11), 410–419. [http://doi.org/10.1016/S1364-6613\(99\)01388-1](http://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://doi.org/10.1037/0012-1649.40.2.177>
- Gilmore, R., & Johnson, M. H. (1995). Working memory in infancy: Six-month-olds' performance on two versions of the oculomotor delayed response task. *Journal of Experimental Child Psychology*, 59, 397–418.
- Gordon, G. E., & McCulloch, D. L. (1999). A VEP investigation of parallel visual pathway development in primary school age children. *Documenta Ophthalmologica. Advances in Ophthalmology*, 99(1), 1–10.
- Gredebäck, G., Örnkloo, H., & von Hofsten, C. (2006). The development of reactive saccade latencies. *Experimental Brain Research*, 173(1), 159–164. <http://doi.org/10.1007/s00221-006-0376-z>
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58(2), 137–176. <http://doi.org/10.1016/j.cogpsych.2008.06.001>
- Groner, R., & Groner, M. T. (1989). Attention and eye movement control: An overview. *European Archives of Psychiatry and Neurological Sciences*, 239(1), 9–16. <http://doi.org/10.1007/BF01739737>
- Gunn, A., Cory, E., Atkinson, J., Braddick, O. J., Wattam-Bell, J., Guzzetta, A., & Cioni, G. (2002). Dorsal and ventral stream sensitivity in normal development and hemiplegia. *Neuroreport*, 13(6), 843–7. <http://doi.org/10.1097/00001756-200205070-00021>

- Hainline, L., Turkel, J., Abramov, I., Lemerise, E., & Harris, C. M. (1984). Characteristics of saccades in human infants. *Vision Research*, 24(12), 1771–1780. [http://doi.org/10.1016/0042-6989\(84\)90008-7](http://doi.org/10.1016/0042-6989(84)90008-7)
- Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 545–552. <http://doi.org/10.1.1.70.2254>
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–194. <http://doi.org/10.1016/j.tics.2005.02.009>
- Heise, N., & Ansorge, U. (2014). The roles of scene priming and location priming in object-scene consistency effects. *Frontiers in Psychology*, 5, 1–11. <http://doi.org/10.3389/fpsyg.2014.00520>
- 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://doi.org/10.1016/j.visres.2014.08.006>
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504. <http://doi.org/10.1016/j.tics.2003.09.006>
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, 16(4), 219–222. <http://doi.org/10.1111/j.1467-8721.2007.00507.x>
- Henderson, J. M., Brockmole, J. R., Castelhana, M. S., Mack, M., Gompel, R. V, Fischer, M., ... Hill, R. (2007). Visual saliency does not account for eye movements during visual search in real world scenes. In *Eye Movements: A Window on Mind and Brain* (pp. 537–562). <http://doi.org/10.1167/9.3.6>
- 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://doi.org/10.4324/9780203488430>
- Henderson, J. M., & Hollingworth, A. (1999a). High-level scene perception. *Annual Review of Psychology*, 50, 243–271. <http://doi.org/10.1146/annurev.psych.50.1.243>
- Henderson, J. M., & Hollingworth, A. (1999b). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5), 438–443. <http://doi.org/10.1111/1467-9280.00183>
- 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://doi.org/10.1037/0096-1523.25.1.210>
- Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, 19(2), 204–10. <http://doi.org/10.3758/s13423-011-0205-7>
- 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://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://doi.org/10.3758/BF03197475>
- Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychologica*, 102(2–3), 319–343. [http://doi.org/10.1016/S0001-6918\(98\)00053-5](http://doi.org/10.1016/S0001-6918(98)00053-5)
- Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: the role of fixation position within the scene. *Memory & Cognition*, 29(2), 296–304. <http://doi.org/10.3758/BF03194923>
- Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, 8(4), 761–768. <http://doi.org/10.3758/BF03196215>
- Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, 16(4), 405–422. [http://doi.org/10.1016/0163-6383\(93\)80001-O](http://doi.org/10.1016/0163-6383(93)80001-O)
- Hou, X., Harel, J., & Koch, C. (2012). Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1), 194–201. <http://doi.org/10.1109/TPAMI.2011.146>
- 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://doi.org/10.1016/j.visres.2011.03.010>
- Ingle, D. (1967). Two visual mechanisms underlying the behavior of fish. *Psychologische Forschung*, 31(1), 44–51. <http://doi.org/10.1007/BF00422385>
- 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://doi.org/10.1167/iovs.05-1311>
- Irving, E. L., Steinbach, M. J., Lillakas, L., Babu, R. J., & Hutchings, N. (2006). Horizontal saccade dynamics across the human life span. *Investigative Ophthalmology & Visual Science*, 47(6), 2478–2484. <http://doi.org/10.1167/iovs.05-1311>
- Irving, E. L., Tajik-Parvinchi, D. J., Lillakas, L., González, E. G., & Steinbach, M. J. (2009). Mixed pro and antisaccade performance in children and adults. *Brain Research*, 1255, 67–74. <http://doi.org/10.1016/j.brainres.2008.12.006>
- 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://doi.org/10.1016/S0042-6989\(99\)00163-7](http://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://doi.org/10.1038/35058500>
- Johnson, M. H. (1994). Visual attention and the control of eye movements in early infancy. In C. Umità & M. Moscovitch (Eds.), *Attention and Performance XV* (pp. 291–310). Cambridge: MIT Press.
- Johnson, M. H. (1995). The inhibition of automatic saccades in early infancy. *Developmental Psychobiology*, 28(5), 281–291. <http://doi.org/10.1002/dev.420280504>
- Johnson, M. H. (2002). The development of visual attention: A cognitive neuroscience perspective. In M. H. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain Development and Cognition: A Reader* (Second Edi). Oxford: Blackwell Publishers Ltd.
- Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1994). Facilitation of saccades toward a covertly attended location in early infancy. *Psychological Science*, 5(2), 90–92. <http://doi.org/10.1111/j.1467-9280.1994.tb00636.x>
- Josephs, E. L., Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica*, 169, 100–108. <http://doi.org/10.1016/j.actpsy.2016.05.013>
- Just, M. a., & Carpenter, P. a. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <http://doi.org/10.1037/0033-295X.87.4.329>
- Kandel, S., Burfin, S., Méary, D., Ruiz-Tada, E., Costa, A., & Pascalis, O. (2016). The Impact of Early Bilingualism on Face Recognition Processes. *Frontiers in Psychology*, 7(July), 1–9. <http://doi.org/10.3389/fpsyg.2016.01080>
- Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review*, 27(3), 283–348. <http://doi.org/10.1016/j.dr.2007.06.006>
- Kayser, C., Nielsen, K. J., & Logothetis, N. K. (2006). Fixations in natural scenes: Interaction of image structure and image content. *Vision Research*, 46(16), 2535–2545. <http://doi.org/10.1016/j.visres.2006.02.003>
- Khan, M. (2013). *Thinking in Words : Implicit Verbal Activation in Children and Adults*.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11), 1762–1776. <http://doi.org/10.1016/j.visres.2005.10.002>
- Klaver, P., Lichtensteiger, J., Bucher, K., Dietrich, T., Loenneker, T., & Martin, E. (2008). Dorsal stream development in motion and structure-from-motion perception. *NeuroImage*, 39(4), 1815–1823. <http://doi.org/10.1016/j.neuroimage.2007.11.009>
- 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://doi.org/10.1017/S0048577201981399>

- Klenberg, L., Korkman, M., & Lahti-Nuutila, P. (2001). Differential Development of Attention and Executive Functions in 3- to 12-Year-old Finnish Children. *Developmental Neuropsychology*, 20(1), 407–428. <http://doi.org/10.1207/S15326942DN2001>
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.
- Koldewyn, K., Whitney, D., & Rivera, S. M. (2011). Neural correlates of coherent and biological motion perception in autism. *Developmental Science*, 14(5), 1075–1088. <http://doi.org/10.1111/j.1467-7687.2011.01058.x>
- Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision*, 13(2), 201–214. <http://doi.org/10.1163/156856800741216>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62(August), 621. <http://doi.org/10.1146/annurev.psych.093008.131123>
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25–26), 3559–3565. [http://doi.org/10.1016/S0042-6989\(01\)00102-X](http://doi.org/10.1016/S0042-6989(01)00102-X)
- Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311–1328. <http://doi.org/10.1068/p2935>
- 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://doi.org/10.1016/j.visres.2007.06.015>
- Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5), 802–817. <http://doi.org/10.1109/TPAMI.2006.86>
- Le Meur, O., & Liu, Z. (2015). Saccadic model of eye movements for free-viewing condition. *Vision Research*, 116, 152–164. <http://doi.org/10.1016/j.visres.2014.12.026>
- Lee, S. P., Badler, J. B., & Badler, N. I. (2002). Eyes alive. *ACM Transactions on Graphics*, 21(3), 637–644. <http://doi.org/10.1145/566654.566629>
- Loftus, G. R., & Mackworth, N. H. (1978). LoftusMackworth1978.pdf. *Journal of Experimental Psychology. Human Perception and Performance*, 4(4), 565–572.
- Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2013). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, 75(5), 1357–1372. <http://doi.org/10.1111/j.1467-8624.2004.00745.x>
- Luna, B., & Seeney, J. A. (2004). The Emergency of Collaborative Brain Function. fMRI Studies of the Development of Response Inhibition. *Annals New York Academy of*



*Sciences*, 1021, 296–309.

- Luna, B., & Sweeney, J. A. (2004). The emergence of collaborative brain function: fMRI studies of the development of response inhibition. *Annals of the New York Academy of Sciences*, 1021, 296–309. <http://doi.org/10.1196/annals.1308.035>
- Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, 68(3), 293–308. <http://doi.org/10.1016/j.bandc.2008.08.019>
- 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://doi.org/10.1177/0956797610373371>
- Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, 121(2), 196–206. <http://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), 363–86. <http://doi.org/10.1163/156856895X00052>
- Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3), 229–240. <http://doi.org/10.1038/nrn1348>
- Matin, E. (1974). Saccadic suppression: a review and an analysis. *Psychological Bulletin*, 81(12), 899–917.
- 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://doi.org/10.1016/S0163-6383\(97\)90035-7](http://doi.org/10.1016/S0163-6383(97)90035-7)
- Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs. Language affects low-level visual perception. *Psychological Science*, 18(11), 1007–1013.
- Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, 14(4), 710–716. <http://doi.org/10.3758/BF03196826>
- Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development*, 20(1), 19–31. <http://doi.org/10.1016/j.cogdev.2004.07.001>
- 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://doi.org/10.1167/11.8.17.Introduction>
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3), 774–85. <http://doi.org/10.1016/j.neuropsychologia.2007.10.005>

- Moeller, G. U., Kayser, C., Knecht, F., & König, P. (2004). Interactions between eye movement systems in cats and humans. *Experimental Brain Research*, *157*(2), 215–24. <http://doi.org/10.1007/s00221-004-1835-z>
- Moore, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *6*(2), 182–189. <http://doi.org/10.1038/nn996>
- Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*(2), 507–517. <http://doi.org/10.1016/j.neuropsychologia.2009.10.011>
- 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://doi.org/10.1007/s002210050473>
- Nakagawa, A., & Sukigara, M. (2013). Variable coordination of eye and head movements during the early development of attention: A longitudinal study of infants aged 12-36 months. *Infant Behavior and Development*, *36*(4), 517–525. <http://doi.org/10.1016/j.infbeh.2013.04.002>
- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science*, *6*(2), 136–142. <http://doi.org/10.1111/1467-7687.00263>
- Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*(1), 73–144. <http://doi.org/10.1017/S0140525X0200002X>
- Noström, 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. In *Neurobiology of Attention* (pp. 251–256). <http://doi.org/10.1016/B978-012375731-9/50045-8>
- Palmer, S. (1975). Visual perception and world knowledge: notes on a model of sensory-cognitive interaction. In D. Norman & D. Rumelhart (Eds.), *Explorations in Cognition* (LNRRes. Gr, pp. 279–307). San Francisco: Freeman.
- Pannasch, S., Helmert, J. R., Roth, K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, *2*(2), 1–19.
- Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception & Psychophysics*, *73*(4), 1120–1132. <http://doi.org/10.3758/s13414-011-0090-1>
- Pannasch, S., & Velichkovsky, B. M. (2009). Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. *Visual Cognition*, *17*(6–7), 1109–1131. <http://doi.org/10.1080/13506280902764422>

- Papageorgiou, K., Smith, T., Wu, R., Kirkham, N., Johnson, M., & Ronald, A. (2013). An investigation of the relationship between individual differences in infant fixation durations and later temperament and behaviour in childhood. *17th EUROPEAN CONFERENCE ON EYE MOVEMENTS Book of Abstracts*.
- 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.
- Parrish, E. E., Giaschi, D. E., Boden, C., & Dougherty, R. (2005). The maturation of form and motion perception in school age children. *Vision Research*, *45*(7), 827–837. <http://doi.org/10.1016/j.visres.2004.10.005>
- Paus, T. (1989). The development of sustained attention in children might be related to the maturation of frontal cortical functions. *Acta Neurobiologiae Experimentalis*, *49*(1), 51–55.
- Paus, T., Babenko, V., & Radil, T. (1990). Development of an ability to maintain verbally instructed central gaze fixation studied in 8- to 10-year-old children. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *10*(1), 53–61. [http://doi.org/10.1016/0167-8760\(90\)90045-F](http://doi.org/10.1016/0167-8760(90)90045-F)
- Pearson, D. a, & Lane, D. M. (1991). Auditory attention switching: a developmental study. *Journal of Experimental Child Psychology*, *51*(2), 320–334. [http://doi.org/10.1016/0022-0965\(91\)90039-U](http://doi.org/10.1016/0022-0965(91)90039-U)
- Phillips, J. O., Finocchio, D. V, Ong, L., & Fuchs, A. F. (1997). Smooth Pursuit in 1-to 4-month-old Human Infants. *Vision Research*, *37*(21), 3009–3020.
- Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory*, *9*(4–6), 423–432. <http://doi.org/10.1080/09658210143000182>
- Poirel, N., Mellet, E., Houdé, O., & Pineau, A. (2008). First came the trees, then the forest: developmental changes during childhood in the processing of visual local-global patterns according to the meaningfulness of the stimuli. *Developmental Psychology*, *44*(1), 245–253. <http://doi.org/10.1037/0012-1649.44.1.245>
- Poirel, N., Simon, G., Cassotti, M., Leroux, G., Perchey, G., Lanoë, C., ... Houdé, O. (2011). The shift from local to global visual processing in 6-year-old children is associated with grey matter loss. *Plos One*, *6*(6), e20879. <http://doi.org/10.1371/journal.pone.0020879>
- 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.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. <http://doi.org/10.1080/17470210902816461>
- Regal, D. M., Ashmead, D. H., & Salapatek, P. (1983). The coordination of Eye and head movements during early infancy: A selective review. *Behavioural Brain Research*, *10*,

125–132.

- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computational Neural System*, 10(4), 341–350. <http://doi.org/10.1088/0954-898X/10/4/304>
- Riby, D. M., & Hancock, P. J. B. (2008). Viewing it differently: social scene perception in Williams syndrome and autism. *Neuropsychologia*, 46(11), 2855–2860. <http://doi.org/10.1016/j.neuropsychologia.2008.05.003>
- Riby, D. M., & Hancock, P. J. B. (2009). Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. *Journal of Intellectual Disability Research : JIDR*, 53(2), 169–181. <http://doi.org/10.1111/j.1365-2788.2008.01142.x>
- Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: Evidence from eye tracking. *Developmental Science*, 12(4), 549–556. <http://doi.org/10.1111/j.1467-7687.2009.00795.x>
- Ridder, W. H., Borsting, E., & Banton, T. (2001). All Developmental Dyslexic Subtypes Display an Elevated Motion Coherence Threshold. *Optometry and Vision Science*, 78(7), 510–517. <http://doi.org/10.1097/00006324-200107000-00014>
- Rosander, K., & Von Hofsten, C. (2002). Development of gaze tracking of small and large objects. *Experimental Brain Research*, 146(2), 257–264. <http://doi.org/10.1007/s00221-002-1161-2>
- Roucoux, a., Culee, C., & Roucoux, M. (1983). Development of fixation and pursuit eye movements in human infants. *Behavioural Brain Research*, 10(1), 133–139. [http://doi.org/10.1016/0166-4328\(83\)90159-6](http://doi.org/10.1016/0166-4328(83)90159-6)
- Rütsche, A., Baumann, A., Jiang, X., & Mojon, D. S. (2006). Development of visual pursuit in the first 6 years of life. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 244(11), 1406–1411. <http://doi.org/10.1007/s00417-005-0248-4>
- Salapatek, P., Aslin, R. N., Simonson, J., & Pulos, E. (1980). Infant saccadic eye movements to visible and previously visible targets. *Child Development*, 51(4), 1090–1094.
- 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://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://doi.org/10.1016/j.neuropsychologia.2005.10.007>
- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception : A selective review, 11, 1–30. <http://doi.org/10.1167/11.5.9>
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5(4), 195–201.

- Scinto, L., & Pillalamarri, R. (1986). Cognitive strategies for visual search. *Acta Psychologica*, 62, 263–292.
- Shea, S. L., & Aslin, R. N. (1990). Oculomotor responses to step-ramp targets by young human infants. *Vision Research*, 30(7), 1077–1092. [http://doi.org/10.1016/0042-6989\(90\)90116-3](http://doi.org/10.1016/0042-6989(90)90116-3)
- Sigman, M., Cohen, S. E., Beckwith, L., Asarnow, R., & Parmelee, A. H. (1991). Continuity in cognitive abilities from infancy to 12 years of age. *Cognitive Development*, 6(1), 47–57. [http://doi.org/10.1016/0885-2014\(91\)90005-X](http://doi.org/10.1016/0885-2014(91)90005-X)
- Slater, A. M., & Findlay, J. M. (1975). Binocular Fixation in the newborn baby. *Journal of Experimental Child Psychology*, 20, 248–273.
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45(4), 447–481. [http://doi.org/10.1016/S0010-0285\(02\)00503-0](http://doi.org/10.1016/S0010-0285(02)00503-0)
- Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, 12(4), 282–286. <http://doi.org/10.7897/2277-4343.04135>
- Spotorno, S., Tatler, B. W., & Faure, S. (2013). Semantic consistency versus perceptual salience in visual scenes: Findings from change detection. *Acta Psychologica*, 142(2), 168–176. <http://doi.org/10.1016/j.actpsy.2012.12.009>
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–4. <http://doi.org/DOI 10.1126/science.7777863>
- Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye Movements and Spoken Language Comprehension. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Handbook of Psycholinguistics* (second ed., pp. 863–900). Oxford: Elsevier. <http://doi.org/10.1016/B978-012369374-7/50023-7>
- 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://doi.org/10.1167/7.14.4>
- Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition*, 17(6–7), 777–789. <http://doi.org/10.1080/13506280902869213>
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45(5), 643–659. <http://doi.org/10.1016/j.visres.2004.09.017>
- Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, 46(12), 1857–1862. <http://doi.org/10.1016/j.visres.2005.12.005>

- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision*, *11*(5), 5. <http://doi.org/10.1167/11.5.5>
- Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18.
- Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6/7), 1029–1054. <http://doi.org/10.1080/13506280902764539>
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection: Reply to commentaries. *Acta Psychologica*, *135*(2), 133–139. <http://doi.org/10.1016/j.actpsy.2010.07.006>
- Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K.-P. (2002). Neural mechanisms of saccadic suppression. *Science*, *295*(5564), 2460–2462. <http://doi.org/10.1126/science.1068788>
- Torkildsen, J. V. K., Sannerud, T., Syversen, G., Thormodsén, R., Simonsen, H. G., Moen, I., ... Lindgren, M. (2006). Semantic organization of basic-level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics*, *19*(6), 431–454. <http://doi.org/10.1016/j.jneuroling.2006.01.002>
- Torkildsen, J. V. K., Svangstu, J. M., Hansen, H. F., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience*, *20*(7), 1266–1282. <http://doi.org/10.1162/jocn.2008.20087>
- Torkildsen, J. V. K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, *20*(4), 332–351. <http://doi.org/10.1016/j.jneuroling.2007.02.003>
- Torralba, A., Oliva, A., Castelano, 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://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://doi.org/10.1016/S0959-4388\(03\)00105-3](http://doi.org/10.1016/S0959-4388(03)00105-3)
- Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *31*(4), 299–337. <http://doi.org/10.1007/BF00422717>
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology (2006)*, *59*(11), 1931–1949. <http://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*

- 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://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://doi.org/10.1016/j.concog.2006.11.008>
- Unema, P. J. a., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473–494. <http://doi.org/10.1080/13506280444000409>
- Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two Visual Systems and their Eye Movements: Evidence from Static and Dynamic Scene Perception. *Proceedings of the XXVII Conference of the Cognitive Science Society*, 2283–2288.
- Velichkovsky, B. M., Rothert, A., Dornhoefer, S. M., & Joos, M. (2002). Towards an Express-Diagnostics for Level of Processing and Hazard Perception. *Transportation Research*, 5(2), 145–146.
- Velichkovsky, B. M., Rothert, A., Dornhoefer, S. M., Joos, M., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 145–156. [http://doi.org/10.1016/S1369-8478\(02\)00013-X](http://doi.org/10.1016/S1369-8478(02)00013-X)
- Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 145–156. [http://doi.org/10.1016/S1369-8478\(02\)00013-X](http://doi.org/10.1016/S1369-8478(02)00013-X)
- Vinter, A., Puspitawati, I., & Witt, A. (2010). Children's spatial analysis of hierarchical patterns: construction and perception. *Developmental Psychology*, 46(6), 1621–1631. <http://doi.org/10.1037/a0020615>
- 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://doi.org/10.1167/9.3.24>
- Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, 10(3), 14.1-13. <http://doi.org/10.1167/10.3.14>
- 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://doi.org/10.3758/s13414-011-0150-6>

- Võ, M. L.-H., & Wolfe, J. M. (2013). Differential ERP signatures elicited by semantic and syntactic processing in scenes. *Psychological Science*, 24(9), 1816–1823. <http://doi.org/10.1177/0956797613476955>
- Volkman, F. C. (1986). Human visual suppression. *Vision Research*, 26(9), 1401–1416. [http://doi.org/10.1016/0042-6989\(86\)90164-1](http://doi.org/10.1016/0042-6989(86)90164-1)
- Von Hofsten, C., & Rosander, K. (1997). Development of Smooth-Pursuit Tracking in Young Infants. *Vision Research*, 37(13), 1799–1810.
- Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology*, 6(4), 632–650. [http://doi.org/10.1016/0022-0965\(68\)90108-2](http://doi.org/10.1016/0022-0965(68)90108-2)
- Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology*, 6(4), 632–650. [http://doi.org/10.1016/0022-0965\(68\)90108-2](http://doi.org/10.1016/0022-0965(68)90108-2)
- 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://doi.org/10.1111/inf.12049>
- Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, 5(February), 1–13. <http://doi.org/10.3389/fpsyg.2014.00054>
- Yarbus, A. L. (1967). Eye movements and vision. *Neuropsychologia*, 6(4), 389–390. [http://doi.org/10.1016/0028-3932\(68\)90012-2](http://doi.org/10.1016/0028-3932(68)90012-2)
- Ygge, J., Aring, E., Han, Y., Bolzani, R., & Hellström, A. (2005). Fixation stability in normal children. *Annals of the New York Academy of Sciences*, 1039, 480–483. <http://doi.org/10.1196/annals.1325.049>
- Yu, H. H., Chaplin, T. A., & Rosa, M. G. P. (2015). Representation of central and peripheral vision in the primate cerebral cortex: Insights from studies of the marmoset brain. *Neuroscience Research*, 93, 47–61. <http://doi.org/10.1016/j.neures.2014.09.004>
- Zelinsky, G. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835. <http://doi.org/10.1037/a0013118.A>
- Zhao, Q., & Koch, C. (2011). Learning a saliency map using fixated locations in natural scenes. *Journal of Vision*, 11, 1–15. <http://doi.org/10.1167/11.3.9>