Grass or Gravel?
Influences on the Visual Categorization of Naturalistic Structures in Infancy and Early Childhood

by

Karola Schlegelmilch

Doctoral Dissertation

Submitted to the

Faculty of Human Sciences

at the **University of Potsdam**

Department of Psychology

January 2021

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-52637
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-526370

I. Kumulative Dissertation, bestehend aus den empirischen Studien:

(1) Schlegelmilch, K., & Wertz, A. E. (2020). Grass and Gravel: Investigating visual properties preschool children and adults use when distinguishing naturalistic images. *PsyArXiv*. https://doi.org/10.31234/osf.io/tgmd3
(Submitted after revision to *Journal of Vision*)

(2) Schlegelmilch, K., & Wertz, A. E. (2019). The effects of calibration target, screen location, and movement type on infant eye-tracking data quality. *Infancy*, *24*(4), 636–662.

(3) Schlegelmilch, K., & Wertz, A. E. (2021). *Visual segmentation of complex naturalistic structures in an infant eye-tracking search task [manuscript in preparation]*.
(Submitted to *PLOS ONE*)

Acknowledgements

First of all, I want to thank my supervisor Dr. Annie E. Wertz for her council and advice, and the members of the Max Planck Research Group Naturalistic Social Cognition for great support and help.

Special thank also to Moritz Daum and the members of the Department of Psychology Developmental Psychology: Infancy and Childhood for the great time I could spend with them at the University of Zurich.

During the work on my dissertation, I was able to benefit greatly from my fellowship at the International Max Planck Research School on the Life Course (LIFE; www.imprs-life.mpg.de; participating institutions: Max Planck Institute for Human Development, Freie Universität Berlin, Humboldt-Universität zu Berlin, University of Michigan, University of Virginia, University of Zurich).

Moreover, I also want to thank my friend and partner for many inspiring and scrutinizing discussions.

Contents

List of Figures

Figures in the Supporting Information

List of Tables

Tables in the Supporting Information

General Abstract

Detecting and categorizing particular entities in the environment are important visual tasks that humans have had to solve at various points in our evolutionary time. The question arises whether characteristics of entities that were of ecological significance for humans play a particular role during the development of visual categorization.

The current project addressed this question by investigating the effects of developing visual abilities, visual properties and ecological significance on categorization early in life. Our stimuli were monochromatic photographs of structure-like assemblies and surfaces taken from three categories: vegetation, non-living natural elements, and artifacts. A set of computational and rated visual properties were assessed for these stimuli. Three empirical studies applied coherent research concepts and methods in young children and adults, comprising (a) two card-sorting tasks with preschool children (age: 4.1-6.1 years) and adults (age: 18-50 years) which assessed classification and similarity judgments, (b) a gaze contingent eye-tracking search task which investigated the impact of visual properties and category membership on 8-month-olds' ability to segregate visual structure. Because eye-tracking with infants still provides challenges, a methodological study (c) assessed the effect of infant eye-tracking procedures on data quality with 8- to 12-month-old infants and adults.

In the categorization tasks we found that category membership and visual properties impacted the performance of all participant groups. Sensitivity to the respective categories varied between tasks and over the age groups. For example, artifact images hindered infants' visual search but were classified best by adults, whereas sensitivity to vegetation was highest during similarity judgments. Overall, preschool children relied less on visual properties than adults, but some properties (e.g., rated depth, shading) were drawn upon similarly strong. In children and infants, depth predicted task performance stronger than shape-related properties. Moreover, children and infants were sensitive to variations in the complexity of low-level visual statistics. These results suggest that classification of visual structures, and attention to particular visual properties is affected by the functional or ecological significance these categories and properties may have for each of the respective age groups.

Based on this, the project highlights the importance of further developmental research on visual categorization with naturalistic, structure-like stimuli. As intended with the current work, this would allow important links between developmental and adult research.

Zusammenfassung

Spezifische Objekte der Umwelt zu entdecken und zu erkennen, sind wichtige visuelle Aufgaben, die Menschen im Laufe der Evolution zu lösen hatten. Es stellt sich die Frage, ob Charakteristika von ökologisch wertvollen Entitäten in der Entwicklung der visuellen Kategorisierungsfähigkeit bei kleinen Kindern eine besondere Rolle spielen.

Die vorliegende Dissertation untersucht, wie sich visuelle Fähigkeiten, visuelle Eigenschaften der Entitäten und Wertigkeit dieser Entitäten auf die frühe Kategorisierungsfähigkeit auswirken. Als Stimuli dienten monochromatische Fotografien, die strukturähnliche Ausschnitte der drei Kategorien Vegetation, natürliche Elemente und Artefakte zeigen. Ein Set visueller Eigenschaften wurden für diese Stimuli computational und durch menschliche Einschätzung erstellt. In drei empirischen Studien wurden übergreifende Forschungskonzepte und -methoden bei Babys, Kindern und Erwachsenen angewandt: Diese beinhalteten (a) zwei Kartensortieraufgaben mit 4-6-jährigen Vorschulkindern und Erwachsenen, worin Klassifizierung und Ähnlichkeitsbeurteilungen erhoben wurden, (b) eine Eyetracking-Suchaufgabe, die den Einfluss visueller Eigenschaften und Kategoriezugehörigkeit auf die visuelle Segmentierung natürlicher Strukturen bei 8-Monate-alten Babys untersuchte. Da Eyetracking mit Babys methodisch anspruchsvoll ist, wurde (c) in einer Studie mit 8-12-Monate-alten Babys und mit Erwachsenen der Einfluss verschiedener Vorgehensweisen auf die Datenqualität untersucht.

Die Kategorisierungsaufgaben zeigten, dass die Performanz aller Altersgruppen von Kategoriezugehörigkeit und visuellen Eigenschaften beeinflusst wurden. Sensitivität für die jeweiligen Kategorien variierte zwischen den Aufgaben und über die Altersgruppen hinweg. So erschwerten Bilder von Artefakten die visuelle Suche bei Babys, wurden aber von Erwachsenen am besten klassifiziert, während Vegetation Ähnlichkeitsurteile am stärksten beeinflusste. Insgesamt bezogen sich Vorschulkinder weniger als Erwachsene auf visuelle Eigenschaften, wobei aber einige davon (z. B. Tiefenwirkung, Schattierung) auch ähnlich stark beachtet wurden. Bei Kindern und Babys beeinflusste Tiefenwirkung die Performanz stärker als formbezogene Eigenschaften, zudem waren sie sensibel für Komplexitätsunterschiede bei statistischen Eigenschaften. Diese Ergebnisse deuten darauf hin, dass die Klassifizierung visueller Strukturen und die Aufmerksamkeit auf visuelle Eigenschaften von deren Wertigkeit für die jeweilige Altersgruppe beeinflusst sind.

Damit unterstreicht das Projekt den Bedarf, visuelle Kategorisierung naturalistischer, strukturähnlicher Stimuli in weiteren entwicklungsbezogenen Studien vertiefend zu erforschen. Das würde wichtige Verbindungen zwischen Entwicklungs- und Erwachsenenforschung ermöglichen, wie es auch von der vorliegenden Arbeit angestrebt war.

Chapter 1

General Introduction

The sense of vision guides many inferences of infants and young children when they interact curiously and attentively with the environment. Starting with birth, they engage with increasing competency in new tasks in accordance with their growing motoric abilities and autonomy (Bushnell & Boudreau, 1993). However, early visual abilities are described in incongruent, seemingly opposing ways: On the one hand, immaturities of visual functions and their protracted development are emphasized, part of which lasts up into adolescence (e.g., Ellemberg et al., 1999; Kovács et al., 1999; Siu & Murphy, 2018). Then again, literature highlights early onset and fast progress in the development of visual abilities within the first year of life (e.g., Atkinson & Braddick, 2013; for a discussion of this incongruency see: Kellman & Arterberry, 2007). Visual competencies beyond basic visual functions, such as visual categorization, also have their onset within the first year of life as indicated by behavioral (e.g., Mandler & McDonough, 1998a; Quinn, 2011; Rakison & Yermolayeva, 2010) and neuroscientific research (for review see: Hoehl, 2016). Some findings on early sensitivities to category distinctions or superior detection of particular stimuli are difficult to explain with the maturational state of the child's visual abilities. These sensitivities become evident within the first year of life and refer to faces (Fantz & Nevis, 1967; Mondloch et al., 1999), the animate-inanimate distinction (Opfer & Gelman, 2011; Rakison & Poulin-Dubois, 2001), and signals imposing ancestrally recurrent threats (LoBue & Adolph, 2019; Rakison & Derringer, 2008; Włodarczyk et al., 2018) among others.

Studies interested in early categorization abilities frequently adapt their stimuli to the child's visual abilities by selecting artificial or graphically simplified stimuli, or by extracting object stimuli from their background. These materials provide important insight into young children's visual processing of higher level visual information, including sensitivity to perceptual regularities (Bhatt & Quinn, 2011; Goldstone, 1998) and adaptation to deviations from familiar regularities (Fiser & Aslin, 2002; Kayhan et al., 2019). However, in every-day situations, young children are confronted with cluttered scenes consisting of diverse textures, colors and lighting gradients, making the detection and distinction of particular surfaces or entities difficult. Research with artificial or bounded stimuli might lack important aspects of visual processing stages necessary for the visual organization of scenes. Evidence that the human visual system is adapted to visual tasks provided by the natural environment (for reviews see e.g.: Geisler, 2008; Simoncelli & Olshausen, 2001) underscores the necessity to assess infant visual abilities with more naturalistic stimuli so that possible processing advantages can unfold. However, such research is still rare. Studies conducted in this field targeted aspects like the effect of low-level saliency on the detection of faces (Amso et al., 2014; Frank et al., 2012; Kelly et al., 2019), and perceptual adaptation to visual properties in naturally occurring textures (Balas et al., 2018; Balas & Woods, 2014), materials (Balas, 2017; Balas et al., 2020), or natural scenes (Ellemberg et al., 2012). These studies varied in the tasks conducted to assess processing ability, including allocation of gaze (e.g., Balas & Woods, 2014; Kelly et al., 2019), similarity judgment (Ellemberg et al., 2012), and classification (Balas, 2017; Balas et al., 2020), and they varied in the visual aspects which were investigated, covering global vs. local features (Balas et al., 2020), summary statistics (i.e., an algorithm by: Portilla & Simoncelli, 2000; included in: Balas et al., 2018; Balas & Woods, 2014), attention to social signals (Amso et al., 2014; Frank et al., 2012; Kelly et al., 2019), and the distribution of spatial scales (Ellemberg et al., 2012). The children's age ranged from as young as three months up into adolescence, with main foci laid on 6- to 10-month-old infants and children aged 5 to 12 years, complemented with comparison groups of adults. Due to this variability in research motivations, it is difficult to draw more general implications from these studies concerning young children's integrative abilities of complex naturalistic structures into visual tasks, provided their developing visual abilities. Still, this research points to three key aspects which affected processing of real-world images in young children. These were (i) children's sensitivity to naturally appearing visual regularities (i.e., statistically assessed vs. manipulated visual properties), (ii) their sensitivity to entities with ecological significance (e.g., faces, natural scenes), and (iii) the developmental state of the child's low- and higher-level visual abilities (e.g., spatial acuity, contrast sensitivity). The coherent consideration of these three aspects in study designs would provide a beneficial contribution to research on visual processing and categorization of naturalistic scenes. The current dissertation project was motivated by the wish to pursue this goal. It includes an eye-tracking search task with 8-month-old infants and two card-sorting tasks with preschool children and an adult comparison group. Images of real-world structures taken from different superordinate categories with diverse ecological significance were employed as stimuli. Furthermore, a study on infant eye-tracking procedures was conducted in preparation of the infant search task, motivated by the aim to implement methods which improve data quality and which support infants' engagement with the task. This was important because of (a) the challenges that unconstrained infant movement and attention generally pose on eye-tracking data quality (e.g., Hessels, Andersson, et al., 2015; Wass et al., 2014), and (b) the relative novelty of using an EyeLink eye-tracker with infant participants. The aim of the project was to answer the following questions:

- How do developing visual abilities relate to the discrimination of real-world structure?
- Does the ecological significance of the structure's category affect this relation?
- Do young children rely on similar visual properties during their discrimination of real-world scenes as adults do?

The following sections will introduce the research concept in more detail and give a selective overview of background literature. The introductions of the studies in Sections 2.2, 3.2, and 4.2 include additional detailed literature reviews. First, aspects of visual development, visual properties, and categorization which are of importance for the categorization of naturalistic scenes and structures will be defined (Section 1.1), followed by examples of how these aspects relate to each other (Section 1.2). Then, general methodological decisions which underlie the conceptual formulation and scope of the project are discussed (Section 1.3), and an overview of the particular studies is given in Sections 1.4.

1.1 Core Aspects of Categorization of Naturalistic Structures

Developmental literature on the categorization of bounded objects frequently investigates sensitivity to perceptual characteristics of these objects during categorization by manipulating the objects' appearance (e.g., controlling the presence or absence of features, creating morphs or adjusting visual similarity between two categories; for overviews see: Bornstein & Arterberry, 2010; Mareschal & Quinn, 2001; Sloutsky & Fisher, 2004). These manipulations relate to characteristics of the objects which are expected to be—in general—perceivable by the child. In contrast, the visual categorization of real-world scenes or structures is fused with the child's ability to visually process the properties which define naturalistic visual structures (e.g., Balas et al., 2018; Ellemberg et al., 2012). This difference to object categorization makes it necessary to specify the aspects comprised in the categorization of naturalistic structures. The following three sections introduce the current perspectives on these key aspects which refer to visual development, visual properties in real-world scenes, and concepts of categorization.

1.1.1 The Development of Visual Abilities

Literature on visual development in infants and young children assesses visual functions ranging from lower-level abilities, such as pattern vision or sensitivity to orientation and contrast, over color vision and sensitivity to depth, to higher-level abilities such as perceptual organization (for overviews see e.g., Braddick & Atkinson, 2011; Kellman & Arterberry, 2007; Zihl & Dutton, 2015). Aside from recruiting participants with normal vision—and also with corrected-to-normal vision if no eye-tracking was applied—we did not assess visual abilities of our participants as part of the conducted studies. Instead, an estimation of these were drawn from reports on the developmental state of visual abilities in similarly-aged participants (see Table 1.1). We mainly focused on (a) basic visual functions which are included in pattern vision, and (b) integrative processing abilities relating to perceptual organization. In particular, we considered spatial acuity and contrast sensitivity (i.e., the ability to discriminate light and dark lines of gradually decreasing size and contrast), and the refinement of spatial acuity, termed hyperacuity or vernier acuity (i.e., the ability to perceive small-scale shifts in lines or pattern; Almoqbel et al., 2017; Dekker et al., 2019; Ellemberg et al., 1999; Skoczenski & Norcia, 2002). Perceptual organization involves a number of aspects: the segmentation of distinct pattern or texture patches, the integration of distributed contour elements, and sensitivity to pictorial depth (for reviews see e.g., Kavšek et al., 2012; Landy & Graham, 2004; Taylor et al., 2014). Of these, we mainly considered texture segregation ability and sensitivity to pictorial depth. These abilities were expected to particularly affect the perception of the image materials we had chosen for the current study, which were monochromatic or greyscale naturalistic structures. Table 1.1 lists the age ranges over which the respective visual abilities develop until they are considered to be equal to those of a mature adult.

Table 1.1 Age-ranges in which visual abilities develop

Visual ability	Age at onset	Adult-like at age	References
Spatial acuity	After birth	8 years	1, 6
Hyperacuity	After birth	Teenage years	4, 7
Contrast sensitivity	After birth	8 years	1, 6
Texture segregation:			4, 5
Phase, intensity [a]	8 weeks	School age	
Orientation [a]	9 months	School age	
Contour processing	3–6 months	Teenage years	6, 8
Pictorial depth	6 months	ca. 7 years	2, 3

Note. References: 1 Almoqbel et al., 2017; 2 Freud & Behrmann, 2017; 3 Kavšek et al., 2012; 4 Kellman & Arterberry, 2007; 5 Sireteanu & Rieth, 1992; 6 Siu & Murphy, 2018; 7 Skoczenski & Norcia, 2002; 8 Taylor et al., 2014)
[a] Textures defined by phase, intensity, or orientation.

One key question concerning young children's visual processing of naturalistic stimuli is whether basic visual abilities, which are assessed with artificial stimuli, transfer to naturalistic scenes. In adults, visual responses to naturalistic stimuli exceed the performance that can be expected from responses to artificial stimuli (Kayser et al., 2004). Yet, comparisons of visual abilities in children assessed with these different stimuli types are rare. One example comes from Ellemberg et al. (2009), who found that children's contrast sensitivity was lower for naturalistic than for artificial stimuli, and also differed stronger from that of adults in naturalistic compared to artificial stimuli. This suggests that children's visual processing abilities related to naturalistic visual input may also differ more generally from those reported in Table 1.1. The current studies therefore investigated what visual information may have been available to children when exploring naturalistic stimuli through their performance in the respective tasks.

1.1.2 Visual Properties in Real-world Scenes

In the current project, visual properties are referred to as visual regularities, which can be assessed from real-world entities computationally or by human ratings. Examples of visual regularities may be that scene elements or objects have similar shading or shape characteristics, or that surfaces consist of coherent patterns. Visual properties contribute to grouping, segregation or categorization of the environment (e.g., Baumgartner et al., 2013; Geisler, 2008; Goldstone, 1998; Torralba & Oliva, 2003). The visual properties commonly investigated to understand (or predict) such visual processes mainly derive from literature on adults (e.g., Cichy et al., 2014; Olshausen & Field, 1996; Portilla & Simoncelli, 2000) or from computer vision (e.g., Clausi, 2002; Gonzalez & Woods, 2018; Tenenbaum, 1995). Although computational models do not necessarily mimic human visual processing, the fields inform one another and share an overlapping toolkit (e.g., Fleming, 2017; Hyvärinen et al., 2009; Wallis et al., 2017).

Relevant aspects of this literature which might help to approximate which visual regularities are extracted from real-world scenes by young children are (a) the relation between characteristics of the human environment and the neural architecture of the visual system leading to facilitated visual processing of the environment (Field, 1987; Geisler & Diehl, 2002; Isherwood et al., 2017) , and (b) the extraction of definable features and properties which underlie recognition or detection of scene elements at different levels of the processing hierarchy (e.g., Oliva & Torralba, 2006; Rao & Lohse, 1996; Schmidt et al., 2017). The following provides a brief summary of both subjects.

1.1.2.1 Properties Leading to Facilitated Processing

The human visual system has adapted to the visual tasks and physical properties inherent in the environment over evolutionary times (for reviews see e.g., Geisler, 2008; Simoncelli & Olshausen, 2001). The changing environments which humans inhabited were variously composed of vegetation, waters, stones, and also desert or snow (e.g., Potts, 2012). Such landscapes constituted visual backgrounds to every-day tasks, including the segmentation and rough classification of such background structures (Oliva & Torralba, 2006; Walther & Shen, 2014). These landscapes also enclosed areas or objects which were of particular interest because of their relevance to human survival and reproduction (e.g., ponds, game, or other human beings), making fast visual processing, segmentation, and the detection of significant visual targets necessary (Gegenfurtner & Rieger, 2000; Thorpe et al., 1996). Research in adults provides much evidence for processing advantages of characteristics that can be encountered in the natural environments. Examples are: particular aspects of the distribution of spatial frequencies (e.g., the slope of the distribution, its scaling invariance), or correlations between visual information (e.g., luminance and contrast) occurring at locations of varying distance in textures or environmental scenes, which both align well with neural computations of the visual system (e.g., Frazor & Geisler, 2006; Isherwood et al., 2017; Olshausen & Field, 1996). With regard to these characteristics, facilitating effects within the visual processing of naturalistic scenes were found for visual memory, detection of targets within the scene, and scene discrimination (H. E. Gerhard et al., 2013; Hansen & Hess, 2006; Hollingworth & Henderson, 2002; White et al., 2008)

1.1.2.2 Properties Involved in Recognition and Classification

The recognition and classification of entities is another task which is of particular importance in human every-day life. Contact with objects and substances involves many senses, and the visual properties adults rely upon during classification range from statistical properties equivalent to those involved in scene processing, to higher-order properties, including aspects of three-dimensional surfaces, or visually inferred material properties such as softness, fluidity or naturalness (Baumgartner & Gegenfurtner, 2016; Contini et al., 2017; Fleming, 2017; Schmidt et al., 2017). Because higher-order properties build upon the processing of basic visual information (Nassi & Callaway, 2009) or rely on prior experience (Goldstone, 1998), they are less reliably assessed with computational algorithms, but instead by human raters (e.g., Nosofsky et al., 2017; Rao & Lohse, 1996). Specific properties will be discussed below Section 1.3.3.

1.1.3 Categorization of Ecologically Significant Entities

"A category exists whenever two or more distinguishable objects or events are treated equivalently. This equivalent treatment may take a number of forms, such as labeling distinct objects or events with the same name, or performing the same actions on different objects." This definition introduces the review on categorization of natural objects by Mervin and Rosch (1981). The merit of this definition lies in its applicability to different types of categorization, and to categorization within all age groups. It does not restrict the indicators of categorization to any modality: they can be behavioral, physiological, or verbal.

1.1.3.1 The Current Concepts of Categorization

Developmental research suggests different models in which categorization and its underlying processes change (or do not change) over development, with particular focus on which resources and sensitivities categorical inferences are based upon (Gelman, 2004; Madole & Oakes, 1999; Mandler & McDonough, 1993; Rakison & Yermolayeva, 2010; Sloutsky, 2016; Westermann & Mareschal, 2012). If different age groups are compared, it is therefore important to assess categorization performance with measures that are sensitive to a variety of mechanisms.

Three distinct notions of categorization are considered here to uncover determinants of visual categorization with real-world images. They refer to (i) the ability to segregate visual structures as an indicator of the perceived distinctness of those structures, (ii) explicit judgments of the visual similarity of structures which does not refer to their identity or class, and (iii) classification, assessed as explicit assignment of a visual structure to a superordinate category.

The first kind of categorization—scene or texture segregation—can be seen as a very basic level of categorization. It occurs non-verbally and can be assessed for all age groups (e.g., by using eye-tracking). Segregation ability can be affected by the salience of a structure, as well as by distinctness in visual pattern and by image content (Amso et al., 2014; Kayser et al., 2006; Sireteanu et al., 2005). Examples of the assessment of this kind of categorization ability within different age groups are eye-tracking search tasks for faces in natural scenes, which take advantage of infants' particular interest in faces (Amso et al., 2014; Frank et al., 2012; Kelly et al., 2019).

The second and third kinds of categorization can be assessed from verbal participants of approximately 4 years of age and older (e.g., Markman, 1989). Judgments of visual similarity can be based on low- and higher-order visual characteristics, including past experiences obtained in contact with the depicted structures. Judgments of perceived similarity (e.g. via grouping or rating) is a useful categorization technique for visual structures for which labels are missing. Similarity judgments are therefore useful in studies investigating texture perception or unfamiliar objects (e.g., Heaps & Handel, 1999; Rao & Lohse, 1996; Schmidt et al., 2017) and can tolerate differences in the familiarity of entities between children and adults. In contrast, familiarity of entities does affect their classification. Still, since classification also relies on inferences based on appearance, similarity judgments as well as classification tasks may allow us to assess the degree to which a person is sensitive to visual properties and image content.

1.1.3.2 Definition of Ecological Significance

Human environments were sources of food and materials, and still are—although dramatic changes have occurred in how environments are structured. Next to the above-described processing advantages related to statistical properties of natural scenes, recurrent involvement with particular entities or objects provided by the environment similarly lead to enhanced performance in visual tasks or sensitivity related to these entities. For example, adults are very fast at classifying animals compared to non-living objects (e.g., Crouzet et al., 2012; Thorpe et al., 1996), and the animate-inanimate distinction is suggested to be evident in the visual processing hierarchy (e.g., Carlson et al., 2014; Sha et al., 2015). Similarly fast processing is evident for the recognition of materials (Fleming, 2017; Sharan et al., 2014). Moreover, information which is of significance to humans such as social signals (e.g., human faces: Amso et al., 2014; Crouzet, 2010), or signals indicating threat (LoBue & DeLoache, 2008) are detected fast and in a privileged way. Vegetation plays a distinctive role for humans in that while it provides both food and raw materials, it can also be hazardous. For this reason, the categorization of plants for subsistence strategies was an integral part of ancestral human life (Hardy, 2018; Şerban et al., 2008; Wertz, 2019).

Sensitivity to some of these entities and signals arises early in life. For example, infants showed distinct behavioral responses (i.e., avoidance, social referencing) and enhanced learning when they were confronted with plants compared to other entities (C. Elsner & Wertz, 2019; Wertz & Wynn, 2019; Włodarczyk et al., 2018). Categorical distinctions which refer to the animate-inanimate distinction become evident during infancy (B. Elsner et al., 2013; Opfer & Gelman, 2011; Rakison & Poulin-Dubois, 2001), and responses to threat signals from spiders or snakes are evident even from infancy (LoBue & Adolph, 2019; LoBue & DeLoache, 2010; Rakison & Derringer, 2008). These examples of early sensitivities support the claim that entities or events which were of ancestral relevance can lead to behavioral adaptations in infants (Pauen & Hoehl, 2015; Wertz, 2019).

Yet, infants and young children are also very sensitive to perceptual regularities within their surrounding (Fiser & Aslin, 2002; Janacsek et al., 2012; Kirkham et al., 2002; Tummeltshammer et al., 2014). Attention to learning opportunities is of essential relevance for young children in order to get acquainted with and act in their every-day environment (e.g., Köster et al., 2020; Oudeyer & Smith, 2016). Such sensitivity also suggests that for young children, ecological significance may imply much more than the particular entities or category domains defined above—visual aspects or physical qualities (e.g., spatial information, contrast distributions) that are part of a child's developmental tasks in supporting his or her interaction with the environment may also be included.

1.2 Interrelations between Visual Abilities, Visual Properties and Ecological Significance in Young Children

1.2.1 Young Children's Gathering of Visual Information

In particular during the first years of life, infants attend to novel or increasingly complex visual pattern and events (Courage et al., 2006; Reynolds et al., 2013; Ruff & Rothbart, 2001). The gathering of such visual experience underlies the development of several visual functions (e.g., spatial acuity, contrast sensitivity, configural processing, and also the specialization of the visual cortex; for review see: Maurer & Lewis, 2013). Ongoing visual experience also leads to perceptual learning of progressively complex and differentiated environmental regularities (Goldstone, 1998). Young children rely on environmental regularities during categorization by 3 months of age (Madole & Oakes, 1999; Quinn, 2011; Rakison & Yermolayeva, 2010), and it does not seem feasible to demarcate the gathering of visual information supporting basic-level visual abilities from the gathering of environmental regularities. Since visual abilities are part of the organism's adaptation to relevant and reoccurring aspects of the environment (e.g., Geisler & Diehl, 2002), exposure to environmental regularities may support the development of such visual abilities to better capture these regularities. All these growing competencies enable children to visually organize their surrounding with decreasing effort (Madole & Oakes, 1999; Quinn, 2011; Rakison & Yermolayeva, 2010).

A key question in the study of young children's ability to visual process properties of real-world scenes is when and how visual adaptation to these properties occurs. This was addressed by investigating perceptual constancy and perceptual narrowing in infants. Visual adaptation generates sparse percepts which rely on the processing of relevant visual information, and on decreased sensitivity to irrelevant information. For example, gloss is perceived as a constant reflectance property in adults—in spite of the variability in appearance of glossy surfaces (Fleming, 2015; Yang et al., 2011, 2015). Surprisingly, already by seven months of age, infants were able to ignore this variability and perceive glossy objects (presented as 3-D graphics) as constant, while 3- to 4-month-olds' were still very sensitive to variations in the surface appearance if illumination changed (Yang et al., 2015). This early onset of perceptual constancy offers an intriguing example of early visual adaptation to characteristics of the environment.

Balas and collaborators (2018; 2014) investigated 3- to 10-month-old infants' sensitivity to the natural appearance of visual structure. They presented greyscale textures of natural substances and surfaces in a preferential looking paradigm, and colorful photographs of plastic objects in a study measuring event-related potentials with EEG. The images were contrasted with manipulated versions, which were obtained by either applying an algorithm that disrupted global image layout (texture synthesis; Portilla & Simoncelli, 2000), or by inverting image contrast to disrupt the typical appearance of local contours. Differential sensitivity to these manipulations was found in infants by 9 months of age in both studies, suggesting sensitivity to aspects of naturalistic visual input within the first year of life. However, a particular contribution of either of the respective manipulations was not conclusively revealed by the results, making it difficult to infer the involvement of underlying visual abilities (Balas et al., 2018).

1.2.2 Visual Properties and Their Significance for Young Children

Some properties of the environment can be understood as being of particular importance to humans because they fundamentally affect their actions. Spatial and physical characteristics, such as the three-dimensionality of an object and its plasticity, are examples for such significant properties. Because visual experience is acquired in interaction with the organism's changing physiology and motoric possibilities (Adolph & Tamis-LeMonda, 2014; Bushnell & Boudreau, 1993; Campos et al., 2000; Siu & Murphy, 2018), the gathering of visual regularities related to characteristics of the environment can similarly be affected by the child's changing organismic state (Bhatt & Quinn, 2011; Colombo & Cheatham, 2006; Kovács, 2000).

Alterations in depth within or between scene elements is one example of a visual quality whose role changes over the course of development (Adolph, 2000), but which is of constant ecological significance. Depth not only indicates spatial characteristics of the surroundings which are relevant for navigation when crawling or starting to walk (e.g., Gibson & Walk, 1960), but also specifies characteristics of surfaces or the shape of an object, which offer essential cues for explorative opportunities (Gibson, 2000) and categorization (e.g. in object-examination tasks; Mandler & McDonough, 1993; Pauen, 2002). Sensitivity to pictorial depth cues such as texture gradients, contour junctions and shading—which can be perceived with one eye and without movement—arises by 6 months of age (Kavšek et al., 2012; Kellman & Shipley, 1991). Since stereoscopic depth perception (i.e., integration of disparities between the retinal images) is still immature for central regions of the visual field until school-age (Giaschi et al., 2013), an early availability of pictorial depth may significantly support a child's perception and categorization of attended areas within their surroundings. The ongoing perceptual refinement of depth perception suggests that depth cues may trigger different behavior in infants compared to young children: rather explorative reactions to spatial cues might be elicited in infants (e.g., Atkinson & Braddick, 2013), whereas children with more experience might already be instantaneously able to rely on two-dimensional depth cues to segment scenes—enabling them to attend to more additional characteristics of a scene.

Another integral characteristic of the environment is its complex structure, consisting of heterogeneous collections of elements and physical qualities—confronting us with wide-stretched surfaces, collections of differently-sized elements, or combinations of them all, which consist of granular, plastic, fluid, or solid substances. Such structural and physical characteristics are not separable from the environment's entities, and they fundamentally affect ongoing tasks. These physical characteristics are also integral properties of categories referring to environmental domains. For example, non-living natural entities include sand, stone, or water, which are prototypical for granular, solid, or fluid substances, respectively. For vegetation, collections of similar looking elements such as branches, blossoms, or leaves are typical. In contrast, manmade things—ranging from consumables to architectural elements—can appear in rather diverse ways depending on their function or prior processing. In adults, the appearance of surfaces and general form essentially contributes to the visual classification of materials as well as to category distinctions such as between plants, minerals, or animals (Baumgartner et al., 2013; Schmidt et al., 2017). However, young children might still be collecting experience with such complex and variable properties, and it is difficult to estimate how much children already rely on these physical aspects during categorization decisions.

A general sensitivity to physical characteristics has been studied in infants by performing similar actions with solid, granular, or fluid substances, or with collections of objects. Indeed, infants' looking behavior when watching these actions was found to be affected by the physical characteristics of the objects and substances by the second half of the first year of life (Chiang & Wynn, 2000; Hespos et al., 2016; Hespos & VanMarle, 2012; Huntley-Fenner et al., 2002), suggesting that higher-level visual properties of substances already start to serve as a basis for cognitive expectations in infants (Hespos et al., 2009; VanMarle & Scholl, 2003). However, infants' reactions to actions performed with substances in these studies may refer to cues other than those available in a visual scene (e.g., the causal effect of the performed action). Thus, infants' reactions in these studies do not necessarily indicate if visual sensitivity to these substances can be drawn upon during other perceptual tasks, such as segmentation.

Visual processing of materials—which also differ in their physical characteristics—was investigated in children and adults in a categorization task by Balas (2017). Here, 5- to 7-year-olds still showed lower performance in their assignment and matching of material categories compared to older participant groups, suggesting a protracted acquisition of the visual properties defining materials. Yet, all participants of this study, and particularly the 5- to 7-year-olds, had greater difficulties with categorizing images of processed materials (i.e., metal) compared to images of water (Balas, 2017), which raises the question of whether other qualities of the materials beyond their visual properties—such as their naturalness—affected visual processing and categorization ability.

1.2.3 Does Ecological Significance Affect Visual Abilities?

Research on the effect of ecological significance on early visual processing and categorization ability shows rather incongruent findings.

For example, in spite of recurrent exposure to natural surroundings, processing performance of natural scenes is still affected by immature visual abilities up into late childhood. Ellemberg and colleagues (2012) compared children aged 6, 8, and 10 years to adults regarding their sensitivity to changes in the spatial characteristics of natural scenes. Only the 10-year-olds showed processing advantages of image statistics which are typical for natural scenes similar to those of adults. Processing difficulty of the younger age groups was explained by their immature contrast sensitivity for lower compared to higher spatial frequencies (Ellemberg et al., 2012).

In contrast, significant entities such as faces are spontaneously detected by infants with increasing sensitivity over the first year of age, even if the faces are included in complex visual scenes (Amso et al., 2014; Frank et al., 2012; Kelly et al., 2019). Yet, unlike infants' ongoing contact to faces, which might support face detection, infants' spontaneous reactions to potentially threatening entities such as spiders, snakes, or threatening faces are evident before they could have had similar exposure to those threat-inducing objects (LoBue & DeLoache, 2010; Rakison & Derringer, 2008). Infants' different responses to spiders and very similar looking cues such as blossoms, or their sensitivity to snakes or threatening faces may rely on evolved detection mechanisms sensitive to low-level visual information and movement pattern (DeLoache & LoBue, 2009; LoBue, 2014; Rakison & Derringer, 2008). However, how do young children process visual information which defines a significant entity, if it might exceed their visual abilities?

Sensitivity to plants—which is apparent in infants (e.g., C. Elsner & Wertz, 2019; Wertz & Wynn, 2014a, 2019)—is difficult to explain by an easy-to-detect perceptual template (Rakison & Derringer, 2008). The manifold involvement of humans with plants, together with the heterogeneous appearance of vegetation, make it difficult to suggest a general diagnostic property. Instead, when considering the reviewed literature, properties affecting the visual processing of vegetation may depend on the visual tasks as much as on the affordances determined by the person's age. Visual tasks related to vegetation include fast processing of natural background, as well as the categorization of plants, plant parts (e.g., branches, fruit), and plant properties for particular purposes (e.g., Gegenfurtner & Rieger, 2000; Hardy, 2018; Knill et al., 1990). Affordances which might be determined by the age of the child refer for example to the avoidance of plants due to potential hazards. Infants, who are not yet able to differentiate subgroups of plants nor to treat plants with the necessary care, show behavioral strategies when confronted with plants (e.g., avoidance, social referencing; C. Elsner & Wertz, 2019; Wertz & Wynn, 2014a; Włodarczyk et al., 2018). Infants do not need to categorize plants so that these strategies are triggered—any general visual aspect, and even visual uncertainty may be sufficient (e.g., Pauen & Hoehl, 2015). However, young children who start to move and act more independently in their environment may profit from increasing plant knowledge (Inagaki & Hatano, 1996; Nguyen & Gelman, 2002), including their categorization, to learn how to avoid hazards and benefit from specific plants.

Since in addition to changing affordances, visual abilities mature and the gathering of perceptual regularities builds up rapidly with increasing age, young children can be expected to draw upon quite different visual information when responding to plants—depending on their age.

1.2.4 Bringing Together Visual Abilities, Visual Properties, Categorization, and Ecological Significance

The current project will take a first step in investigating visual properties used by infants and preschool children in their visual processing and classification of vegetation in comparison to other superordinate categories. By considering the reviewed literature, a complex and dynamic interplay of abilities and environmental properties can be brought into focus, which develops in conjunction with the growing competencies and behavioral abilities of a child. It also becomes obvious that it is mainly the relation between visual abilities and categorization—including the possible effect of ecological significance on both abilities—which is least understood

1.3 General Methodological Decisions

The following will give details about general decisions on methods and materials underlying the current research objective. These methods predominantly relate to the categorization studies, which are : "Grass and Gravel: Investigating visual properties preschool children and adults use when distinguishing naturalistic images" (Study 1), and "Visual segmentation of complex naturalistic structures in an infant eye-tracking search task" (Study 3). The rationale of the study conducted to adapt and improve infant eye-tracking procedures and data quality (Study 2) is presented as well.

1.3.1 Choice of Participant Groups

We considered two age groups of young children as participants in the categorization studies, namely, 4- to 5-year-old preschool children together with a comparison group of adults were chosen for Study 1 which was investigating classification and similarity judgments of naturalistic structures by conducting card-sorting tasks. Additionally, 8-month-old infants were chosen for Study 3—an eye-tracking search task investigating the impact of category- and structure-related image characteristics on infants' detection performance.

By preschool age children have already have already gained substantial experiences of environmental regularities and developed naive concepts of environmental entities and events (Goldstone, 1998; Opfer & Gelman, 2011). Although the application of labels starts to affect categorization already during infancy, children's categories are increasingly affected by language with the beginning of school age (Markman, 1989; Nazzi & Gopnik, 2001; Westermann & Mareschal, 2012). Then, the impact of culture and language also effects visual scene processing (e.g., Köster et al., 2017a). In line with the current research interest contrast, we therefore chose preschoolers, who may provide more naive classifications than older children. Categorical inferences in preschoolers can be expected to still be more strongly based on experiences and possibly on their sensitivity to ecologically-driven significance.

Eight-month-old infants were chosen for the eye-tracking search task in which a discrepant structure patch had to be detected on a background structure. Eight-month-olds are expected to have already gathered some visual regularities of their environment including spatial and textural cues (Balas & Woods, 2014; Bertenthal, 1996), possess the ability to orient to locations of interest, and show sustained attention, such as during explorative tasks (Colombo & Cheatham, 2006; Ruff & Rothbart, 2001). Moreover, 8-month-olds had shown differential behavior to plants compared to manmade or other naturally occurring objects in previous studies (e.g., C. Elsner & Wertz, 2019).

In addition to these qualifications, the selected age groups of the children were the youngest who were able to perform the tasks included in the two main studies (Markman, 1989; Wang et al., 2012).

1.3.2 Choice of Images

The stimuli used in the categorization experiments (i.e., the card sorting study Chapter 2 and the eye-tracking search study Chapter 4) were all based on photographs taken of structure-like extracts of real-world entities. These entities belonged to the superordinate categories vegetation, non-living natural elements, and manmade artifacts.

Advantages of Using Photographs

Photographic images are a compromise between real-world settings and fully controlled artificial stimuli. Although the background of the current study relates to young children's perception of their real-world environment, we chose photographs for several reasons. In their reduction to two dimensions they preserve significant properties of real-world structure and complexity which exceeds that of graphics, line drawings, or Gabor patches. Properties of texture, shape, and pictorial depth are well represented in photographs, allowing the assessment of children's ability to perceive this kind of information. Further advantages are that algorithms to extract statistical properties are directly applied to digitized photographs. This allowed us to test the participant's reactions to the stimuli and compare them with the stimuli's statistical properties. When implemented in an eye-tracking experiment, photographs are presented on the eye-tracking monitor and the registration of participants' gaze can provide substantial information about their attention to regions of the scenes. Additionally, gaze-contingent feedback can be given during eye-tracking, which is a valuable means of experimental guidance, especially for non-verbal participants (Wang et al., 2012).

Selection of Contrasting Categories

The particular role that vegetation played for humans, as well as its heterogeneous visual characteristics inspired the project's research interest (see Section 1.2.3), and vegetation is one of the stimuli's superordinate categories. As contrasting categories, artifacts and non-living natural elements were chosen.

Artifacts in the form of tools, articles of daily use, or architecture nowadays dominate most living spaces of industrialized cultures. In contrast to vegetation, which was an important part of the human environment during ancestral times, early-history artifacts (e.g., leaves, stones or bones) visually resembled natural objects and only differed in their intentional use or nomination (Carrara & Mingardo, 2013), becoming increasingly manufactured and artificial only in more recent history. In comparison, non-living natural elements from the third category have been part of the human environment as long as vegetation, providing visual background as much as vegetation did, although the appearance of non-natural elements has not changed dramatically in the same way. Nowadays experiences with natural elements can be described as less detailed—compared to vegetation, to which differentiated contact occurs when plants are grown, watered, harvested, and used as decoration or food even in industrialized societies. The occurrence of natural elements in the environment in form of lakes, earth, rock or snow is mainly somewhat passive, for example, by providing prototypical substances on or in which locomotion occurs. Although during locomotion, action-related vision may well discriminate physical characteristics of such substances (e.g., Pelz & Rothkopf, 2007), the explicit differentiation and classification of exemplars or instances of natural elements is not as common in western cultures as it is for vegetation.

Important materials and substances of natural elements, which are made accessible mainly by physical or chemical transformation (e.g., metal, concrete), or manufactured goods which are produced from plant materials (e.g., noodles, fabric), are not considered here as natural elements nor vegetation, since they lost their original characteristics—instead, some were included as manmade artifacts (e.g., architectural elements. For image examples see Figure 2.**2**).

In spite of their diverse roles, all three categories provide visual surroundings in which humans act and move. They were depicted as extracts of structures instead of bounded objects in the images, because structures best represent the way categories appear in visual scenes (see also Section on ecological significance 1.2.3).

1.3.3 Relevant Features of the Selected Visual Properties

We chose common statistical and rated properties with the intention that they would provide variance within the images of a category, and that there would be the right amount of difference between the selected categories. However, they needed to fulfill several additional requirements. In particular, they needed to (i) be suitable for both descriptive and inferential statistical analysis, in that they are interpretable as a single, particular characteristic, (ii) cover basic statistical properties as well as higher-order characteristics, (iii) relate to aspects of visual abilities which change during development (including contrast, shape characteristics, spatial frequencies, surface structure and surface appearance), and (iv) be sufficiently distinct from each other. In order to keep statistical comparisons within a manageable range, computational properties were reduced to one property per aspect of interest (e.g., only the one property deviation is representing scaling invariance and fractality, which can be assessed in many different ways; e.g., Costa et al., 2012; Isherwood et al., 2017; Redies et al., 2007). Moreover, we adapted the choice and number of properties included in the infant search task to the small number of participants in this study by including mainly those visual properties which refer to more general aspects of image structure.

The visual properties included in both studies are listed in Table 1.2. Further details about the respective properties are provided in Tables 3, 6, and Figure 4.15.

Table 1.2 Properties included in the categorization studies

Property	Definition	Included in study:
Rated [a]		
Curvature	*Angular vs. curved.*	1, 3
Depth	*Plane vs. three-dimensional.*	1, 3
Gloss	*Dull vs. reflecting.*	1
Regularity	*Regular vs. chaotic.*	1, 3
Size	*Small vs. large pattern.*	1
Symmetry	*Asymmetrical vs. symmetrical.*	1, 3
Computational [b]		
Alpha	Steepness of the distribution of energy across spatial frequencies.	1, 3
Deviation	Deviation of a spatial frequency distribution from the fitted line defined by Alpha.	1, 3
CooCor	Co-occurrence Correlation.	1
Entropy [c]	Shannon entropy of pixel luminance values (Shannon, 1948).	3
Luminance [c]	Mean pixel luminance.	3
Skew	Skew of the pixel luminance histogram.	1, 3

[a] Rated properties were formulated as opposites and judged on a continuous scale by adult participants.
[b] Computational properties were assessed with functions implemented in Matlab (version R2017b) or provided by literature on image processing (Gonzalez & Woods, 2018)
[c] In the eye-tracking search task, computational properties were transformed to target-background difference variables (Section 4.9.1.1).

1.3.4 Methods of Analysis

Inferences drawn from the diverse studies conducted on the processing and categorization of naturalistically occurring visual structures (e.g., Balas, 2017; Balas & Woods, 2014; Ellemberg et al., 2012; Kelly et al., 2019) are manifold and frequently non-conclusive. Therefore, a project intending to obtain a basic and rather general approach to this subject cannot rely on previous findings to conduct scrutinizing experiments. Moreover, to best deal with the uncertainty of which visual processes and sensitivities are involved in categorization of naturalistic structures in young children, prior assumptions needed to be kept as unconstrained as possible. Therefore, at least when conducting the first investigation of this project, an application of exploratory methods was the most promising option. For example, multidimensional scaling and hierarchical cluster analysis allow to condense and generalize the structure of the participants' judgments on the images (e.g., Hair et al., 1998). These methods of analysis were employed in the card-sorting tasks (for further details see Section 2.4). With the help of inferences drawn from the exploratory analysis, a reference-frame could be established which then guided the more directed experimental setting of the subsequent eye-tracking search task.

As outlined above, the concepts of categorization considered here were well operationalized by the selected tasks (see Section 1.1.3.1). Moreover, the analysis of target detection performance and sorting decisions is similarly sensitive for visual and conceptual distinctions (Aslin, 2007; Markman et al., 1981; Van Gompel, 2007), which can be specified by the choice of stimulus materials. To best relate inferences drawn from the studies to each other, the same photographs of real-world structures were chosen as stimuli in the two categorization studies.

1.3.5 Challenges of Infant Eye-tracking

Eye-tracking with infants offers a spatially and temporally detailed analysis of infant attention and a flexible implementation of experimental paradigms (for overviews see: Aslin, 2007; Gredebäck et al., 2009; Oakes, 2012). Eye-tracking was therefore the appropriate method to be chosen for studying how visual and conceptual aspects affect scene segmentation in infants. However, eye-tracking with infants can be problematic, because eye-tracking instruments are commonly developed for participants who can constrain their movements. In contrast, infants are likely to spontaneously perform body movements and to turn their gaze away from the eye-tracking monitor, which both can markedly effect data quality (Haith, 2004; Hessels, Andersson, et al., 2015; Morgante et al., 2012; Wass et al., 2014). Furthermore, due to young children’s (relatively short) attentional span, particular experimental procedures might be necessary to support their engagement with the task. At the time at which the experiment was conducted, there was no data available on the success of infant calibration targets, nor on measurement quality that can be expected from unconstrained participants using the Eyelink eye-tracking technology (SR Research Ltd. 2015) which was chosen for the current project. In particular, infant calibration videos need to comprise several eligibilities. For example, the complexity of their characteristics should repeatedly attract infants' attention, but not lead to distress, which can be caused by overwhelming complexity (e.g., Aslin & Smith, 1988; Ruff & Rothbart, 2001). We therefore investigated which measures successfully support infants' attention to common experimental procedures, such as repetitive calibrations and trial sequences, and which actions can be taken to improve data quality (Schlegelmilch & Wertz, 2019).

In particular, with both 8- to 12-month-old infants and adults, we investigated (i) which types of calibration targets differing in shape and movement are preferred to others; (ii) which calibration targets lead to the most central fixations; (iii) how long a target attracts more central fixations before gaze shifts away–indicating the time in which fixations during a calibration procedure should be accepted; (iv) how body and head movements affect measurement accuracy and precision at different screen locations. Moreover, the study applied eye-tacking procedures—such as randomly alternating trials providing scenes with different layout and with changing sounds and background colors by keeping the luminance level stable—which were intended to increase the infant participants' interest in the experiment without causing distress.

Procedures and materials that worked well in this methodological study were then also implemented in the infant eye-tracking search task (Study 3). For example, calibration procedures of the eye-tracking search task used a different precise target and a different color scheme with every re-calibration. Additional precise calibration stimuli which worked simultaneously as attention grabbers and validation targets (in that they only disappeared if infants fixated a central region of the stimulus) were included to control measurement accuracy. These attention grabbers were placed at central and peripheral screen locations, because accuracy and precision were differentially affected at peripheral locations. In Study 3, the search stimuli also randomly alternated across three monochromatic colors, which had been found to increase infants' engagement. Yet, no underlying music was added to the search stimuli, because in the methodological study, some trials with especially happy music had affected some infants to joyfully perform rhythmic movements. A further insight adopted from Study 2 was the utility of analyzing movement during eye-tracking, which can account for a large part of variance in the statistical models on task performance.

1.4 The Studies

The following chapters present three studies conducted to investigate infants' and young children's categorization of naturalistic structures (Chapter 2 and 4) and to ensure methodological precision (Chapter 3).

In Chapter 2 we introduce the study "Grass and Gravel: Investigating visual properties preschool children and adults use when distinguishing naturalistic images" (Study 1). This study is an explorative investigation of 4- to 5-year-old children's and adults' card-sorting decisions in two tasks: (a) sorting cards depicting naturalistic structures into groups according to perceptual similarity, and (b) sorting a different set of these cards into boxes according to their superordinate category membership (i.e., vegetation, natural elements, artifacts). The sorting decisions of the participants were related to visual properties of the images by using explorative and inferential methods. Results revealed visual properties on which children and adults equally relied during categorization and similarity sorting, and other properties which differed between the participant groups in these tasks. In further analysis, differences in categorization performance were related to the three superordinate categories, and participants' sensitivity to the superordinate categories during their judgments of perceptual similarity was compared. This allowed us to make inferences about the significance of the particular categories for the age groups. Study 1 was preceded by a pilot-study in which 26 adults were performing variants of the sorting tasks on 141 images. With the pilot study, methodological decisions such as the selection of subgroups of images, the number of participants, the size of the image sets, and the amount of sorting options during each sorting session were determined in order to be able to successfully apply the intended statistical methods (see also Section 2.3.2).

Chapter 3 presents Study 2: "The effects of calibration target, screen location, and movement type on infant eye-tracking data quality", which investigates the virtue of materials and experimental procedures during infant eye-tracking. It was conducted in preparation for the infant eye-tracking search task (Chapter 4). Eight- to 12-month-olds and a comparison group of adults were watching variants of animated calibration targets, and the adult participants were additionally performing head and body movements—similar to those infants spontaneously perform—while their gaze was being tracked. The results yielded well-working calibration videos and procedures to be implemented in the main eye-tracking search task. Additionally, the adult movement tasks revealed that accuracy and precision were affected in diverse ways by the particular movement directions, and by the screen location which was fixated during movement. These findings provided measures useful for obtaining a more reliable assessment of infants' gaze.

The third study presented in Chapter 4, "Visual segmentation of complex naturalistic structures in an infant eye-tracking search task", investigated infants' ability to distinguish visual structures according to their superordinate categories, or to visual properties present in the structures. Search stimuli consisted of a background structure image in which a patch of another structure image was inserted. It was shown that detection success was affected by combinations of categorical and property-related characteristics of the target-background image combinations. Furthermore, the analysis incorporated additional variables which were extracted from the sorting tasks in Study 1—namely, category assignments and similarity judgments of preschool children and adults related to images which were likewise used in both studies. This further analysis showed that the preschoolers' sorting decisions predicted infants' detection performance more strongly than the adults' decisions, indicating a relationship between preschoolers and infants visual processing of the visual structures. These respective results will be discussed in the general discussion Section 5.4.

Chapter 2

Grass and Gravel:
Investigating Visual Properties Preschool Children and Adults Use When Distinguishing Naturalistic Images[1]

(Study 1)

Karola Schlegelmilch and Annie E. Wertz

Max Planck Institute for Human Development, Max Planck Research Group Naturalistic Social Cognition, Berlin, Germany.

Web link to data:

https://osf.io/8xy5n/?view_only=6ddced286c31456fae7d20dd86e072e6

---

[1] Submitted after revision to *Journal of Vision.*

Available as preprint: Schlegelmilch, K., & Wertz, A. E. (2020). Grass and Gravel: Investigating visual properties preschool children and adults use when distinguishing naturalistic images. *PsyArXiv*. https://doi.org/10.31234/osf.io/tgmd3

2.1 Abstract

Visual processing of a natural environment occurs quickly and effortlessly. Yet, little is known about how young children are able to visually categorize naturalistic structures, since their perceptual abilities are still developing. We addressed this question by asking 76 children (age: 4.1-6.1 years) and 72 adults (age: 18-50 years) to first sort cards with greyscale images depicting vegetation, manmade artifacts, and non-living natural elements (e.g., stones) into groups according to visual similarity. Then, they were asked to choose the images' superordinate categories. We analyzed the relevance of different visual properties to the decisions of the participant groups. Children were very well able to interpret complex visual structures. However, children relied on fewer visual properties and, in general, were less likely to include properties which afforded the analysis of detailed visual information in their categorization decisions than adults, suggesting that immaturities of the still-developing visual system affected categorization. Moreover, when sorting according to visual similarity, both groups attended to the images' assumed superordinate categories—in particular to vegetation—in addition to visual properties. Children had a higher relative sensitivity for vegetation than adults did in the classification task when controlling for overall performance differences. Taken together, these findings add to the sparse literature on the role of developing perceptual abilities in processing naturalistic visual input.

2.2 Introduction

Our daily environment consists of assemblies of complex structures, for example rock surrounded by shrubbery, or the differently patterned clothes piled up in our wardrobe. Still, adults are adept at distinguishing these entities visually, even if the visual information is reduced to two dimensions and color cues are lacking, like on a photograph (Renninger & Malik, 2004). Regularities of visual properties serve as cues for visual categorization (Geisler, 2008; Torralba & Oliva, 2003). However, for young children it might be more laborious to categorize the entities depicted in a photographed scene. One can think of several reasons for this additional effort. The reduced visual information of the photograph might make it difficult for the child to separate the image into regions which belong to separate entities, especially if they occlude each other. Alternatively, a child's perceptual abilities might not be sufficiently developed to perceive some fine-grained or complex visual information necessary for categorization. On either interpretation, a child's categorization success depends on the ability to perceive and differentiate diagnostic visual information, such as textures or contour elements that define the identity of the objects and their distribution in real-world space.

2.2.1 The Development of Visual Categorization

Categorization occurs when different discriminable entities are treated as equivalent, for example by labeling them with the same name, or by performing the same action on them. Individuals can generalize category membership to novel instances on the basis of some internalized representation of the category (Mervis & Rosch, 1981; Quinn, 2011). With increasing age and experience during development, categorical differentiations become more fine-grained (e.g., Fisher et al., 2015; Smith, 1979), and more distinctions as well as more similarities among entities are noticed, motivating an organization of categories into hierarchies in which more specific classes are included in more general ones (Markman, 1989).

Perceptual similarities between stimuli are found to elicit category formation and induction in infancy and early childhood (Badger & Shapiro, 2012; Madole & Oakes, 1999; Sloutsky & Fisher, 2004). If infants and young children are exposed to stimuli which have correlating properties, they incidentally form categories (e.g., Rakison & Yermolayeva, 2010; Younger & Gotlieb, 1988). These similarities can be based on properties like shape, color, or texture (e.g., Quinn & Eimas, 2000), on features like symmetry (Bornstein & Stiles-Davis, 1984), on the presence of salient parts like the legs of a chair (Tversky & Hemenway, 1984), or on types of motion patterns (Opfer, 2002). The inclusion of properties and relations between properties in children's categorical representations becomes more complex with age (Smith, 1979). Until preschool age, children are more likely to base category membership on within-category similarity rather than on between-category dissimilarities (Markman, 1989). Accordingly, they have a stronger tendency to attend to more distributed properties of unique stimuli, instead of primarily attending to diagnostic features which define category boundaries (Deng & Sloutsky, 2016), and young children have a stronger tendency to incidentally learn item-specific perceptual information compared to older children and adults when viewing a sequence of images (Köster et al., 2017b; Ofen & Shing, 2013).

Concurrently to perceptual information, non-verbal social signals and language direct a child's attention to relevant objects and features, facilitating the formation of categories and taxonomies (e.g., Bloom, 2002; Nazzi & Gopnik, 2001; Pauen & Hoehl, 2015). When adults cue a young child's attention to objects, the label they provide typically refers to basic level categories, like "flower" instead of the subordinate level term "tulip", or "chair" instead of the superordinate level term "furniture" (Markman, 1989). Basic level categories possess high within-category similarity and low between-category dissimilarity and are found to be processed faster than higher category levels in children and adults (Contini et al., 2017; Mervis & Rosch, 1981).

Other research emphasizes the relevance that more general superordinate categories such as plants, non-living natural entities, human-made objects or animals have to humans (Carrara & Mingardo, 2013; Gelman, 1988; Opfer & Gelman, 2011). Foundationally different statuses are given to these categories, for example, some categories are viewed as more natural whereas others are seen as being grounded on conventions (Rhodes & Gelman, 2009). In contrast to perceptually more similar basic level categories, the information included in the formation of perceptually heterogeneous superordinate categories is more likely to be drawn from multiple sources. For instance, categorization of artifacts strongly relies on their function or the way they are used, and preschool children were found to rapidly learn to group objects according to this kind of information (e.g., Casler & Kelemen, 2005; Matan & Carey, 2001; Truxaw et al., 2006). Similarly, living things can be categorized by drawing upon their goal orienting behavior (Opfer, 2002). The distinction of living and non-living things emerges during infancy (Rakison & Poulin-Dubois, 2001), but is found to develop still during school age to become based on less obvious attributes and overcome earlier naive beliefs. Plants are not explicitly acknowledged as living things until later in childhood (Carey, 1988; Opfer & Gelman, 2011), yet there is evidence that young children possess rich representations of this category. For example, preschool children understand that plants can grow, need water and sunshine to do so, and can die (Backscheider et al., 1993; Hickling & Gelman, 1995; Inagaki & Hatano, 1996; Nguyen & Gelman, 2002). Perceptual features which are included in superordinate category representations provide crucial cues for nonobvious properties (Gelman, 2004).

Researchers have also proposed that humans are equipped with specific sensitivities to certain categories due to their relevance over evolutionary time, and are prepared to process information about those categories in ways that allow for rapid responses and efficient learning (Barrett, 2014; Pauen & Hoehl, 2015). Visual sensitivity for dynamic and static features can enhance learning of functional and causal aspects that support the categorization of instances within the environment (Rakison & Poulin-Dubois, 2001). Moreover, visual sensitivity can enhance learning and facilitate adaptive responses. For example, research has shown that infants rapidly orient towards threatening stimuli presented as images of ancestrally relevant threats like spiders and snakes (LoBue et al., 2010). On the basis of these findings, researchers have proposed that the perceptual system is equipped with initial template representations that can be matched with real world visual input and trigger attention and learning (Rakison & Derringer, 2008). Plants are another category that has posed significant benefits and costs for humans over evolutionary time (Wertz, 2019). However, unlike snakes and spiders, plants are inconsistently shaped such that a representational template cannot easily be defined. Visual characteristics of plants are very complex and include symmetry, occlusion, and the repetition of parts varying in their orientation and size, for which young children were found to possess limited perceptual abilities (e.g., Siu & Murphy, 2018). Nevertheless, infants respond differently to plants than manmade artifacts, animals, and other natural kinds (e.g., rocks, shells; Elsner & Wertz, 2019; Mandler & McDonough, 1998b; Wertz & Wynn, 2014; Włodarczyk et al., 2018) and selectively learn about plant properties such as edibility (Wertz & Wynn, 2014a, 2019). To our knowledge, the perceptual features used to distinguish heterogeneous superordinate categories such as plants from other entities still need to be investigated.

2.2.2 Visual Features Included in Adult Visual Categorization

Research asking which visual information adults rely upon when identifying or differentiating entities in their environment covers category levels ranging from global distinctions between animate and inanimate things (e.g., Schmidt, Hegele, & Fleming, 2017) to subordinate category levels like subtypes of rock (Nosofsky et al., 2017). We will begin with an overview of visual properties in real-world scenes or extracts of scenes that have been investigated in the adult categorization literature.

Textures provide important visual information relevant for any level of categorization. Natural textures are defined as spatially homogeneous, consisting of repeated elements (Julesz, 1981; Portilla & Simoncelli, 2000; Zhu et al., 2005). Characteristics of the arrangement of these small elements such as their structural regularity, directionality, depth or roughness (Heaps & Handel, 1999; Rao & Lohse, 1993; Rao & Lohse, 1996) can be assessed via rating scales or card sorting with images of natural texture samples (e.g., Brodatz, 1966). Early stages of texture processing are fundamental for the segmentation of visual scenes (Marr, 1976). Because these early processing levels do not necessarily reach awareness, their characteristics are computationally assessed and based on pixel luminance levels of digitized images. Such computational approaches include characteristics of the luminance histogram and spatial frequency information, and can lead to complex parametric texture models (e.g., Portilla & Simoncelli, 2000; Wallis et al., 2017; Zujovic et al., 2013), but less complex image statistics which assess contrast, co-occurrence, or predictability of pixel luminance values can also provide diagnostic visual information (Clausi, 2002; Geisler, 2008). Within more heterogeneous scenes (e.g., urban or rural environments), statistical analysis is often generally applied to the entire image to predict the superordinate category of a scene. Natural scenes usually have a fractal-like quality in that their spatial frequency distribution stays approximately the same even if one zooms into the image (termed scale-invariance; Burton & Moorhead, 1987; Knill et al., 1990; Ruderman, 1997). The slope of a line fitted to this distribution, represented by the value of alpha, is another statistic applied to naturalistic scenes (see Figure 2.**1** for examples). Images within certain ranges of alpha are found to be more visually discriminable than images with alpha values outside these ranges (Hansen & Hess, 2006; Isherwood et al., 2017). Alpha has also been found to differ between categories of image content (Redies et al., 2007). Furthermore, studies that focus on the superordinate categories of the visual scenes (e.g., urban, forest, beach etc.) also investigate descriptors of general gist (e.g., Oliva & Torralba, 2006) or descriptors of the included contours (e.g., Walther & Shen, 2014).

An important focus of this categorization literature is unique objects. These objects are commonly extracted from their context, which increases attention to aspects of their shape. In particular the discrimination of basic level categories could well rely on shape features shared within each group (Goldstone, 1994; Rosch et al., 1976). However, visual categorization of perceptually heterogeneous higher-order categories is as much a subject in research with adults as it is with young children (e.g., Jozwik et al., 2016; Warrington & McCarthy, 1987; Zachariou et al., 2018). In a behavioral study, Schmidt, Hegele, & Fleming (2017) compared the classification of unfamiliar objects belonging to three superordinate categories—animals, plants and minerals—to judgments of the objects' properties. Specifically, mid-level shape opposites like non-symmetrical/symmetrical/, chaotic /regular, angular/curved, or if the object is branched or not were assessed. On the basis of this catalog, Schmidt et al. (2017) were able to classify distinctive shape features such as symmetry or roughness which increased the probability by which an object was assigned to a certain superordinate category. The materials that comprise an object have distinct surface qualities that can also contribute to the classification of objects. Material qualities sometimes refer to experiences induced by movement or touch (e.g., gloss or softness; Fleming, 2017; Hiramatsu et al., 2011), so that studies including these properties frequently rely on human judgments (Baumgartner et al., 2013; Fleming et al., 2013; Nosofsky et al., 2017).

Several behavioral studies found it fruitful to explore categorization of naturalistic images by interrelating data received through a combination of methods including rating of image features, visual similarity judgments, and classification of image content (Gegenfurtner & Rieger, 2000; Heaps & Handel, 1999; Nosofsky et al., 2018; Rao & Lohse, 1993). This approach has the advantage of being beneficially applicable to both children and adults (e.g., Sloutsky, 2003). Yet, in order to successfully interpret findings of developmental studies including the categorization of visually complex naturalistic images, it is important to consider the impact of immature vision on image perception (Aslin & Smith, 1988; Ellemberg et al., 1999).

2.2.3 Basic-level Visual Abilities Affect Perceptual Organization in Children

Visual abilities contribute to the child's developing understanding of his or her world and should not be isolated from the development of cognition, attention, and action (Atkinson & Braddick, 2013). During their first year of life, infants are already sensitive to visual signals which enable them to perceive some substantial properties of their environment (Aslin & Smith, 1988; Braddick & Atkinson, 2011; Daw, 2014; Kellman & Arterberry, 2007). However, lower order visual abilities such as spatial acuity and contrast sensitivity do not reach adult levels before late childhood (Almoqbel et al., 2017; Ellemberg et al., 1999; Leat et al., 2009). Vernier acuity—the ability to perceive detailed spatial relationships—continues to develop beyond spatial acuity into the teenage years (Bondarko & Semenov, 2012; Dekker et al., 2019). Such basic visual abilities promote the perception of fine-grained patterns and disrupted contours (Skoczenski & Norcia, 2002) and provide necessary detail for the development of perceptual organization.

In particular, if images depict scenes with overlapping and heterogeneous elements—such as rock surrounded by shrubbery—segregation and grouping of texture patches is essential to perceptually organize the scenes into meaningful areas. Texture segregation (the effortless discrimination of texture features; Kastner et al., 2000; Landy & Graham, 2004) or contour grouping (e.g., Elder & Goldberg, 2002; Geisler et al., 2001) therefore are prerequisites to successful identification (Panis et al., 2008; Perrinet & Bednar, 2015). Both abilities were found to mature around the age of 13 years (for a review see: Taylor et al., 2014).

Junctions within the contours of depicted objects can signal occlusion and depth (e.g., Kellman & Shipley, 1991). Sensitivity to monocular depth cues such as contour junctions, shading, or texture gradients is already present in infants around 6 months of age (Kavšek et al., 2012). Yet, more complex or detailed pictorial depth (e.g., Freud & Behrmann, 2017) is difficult to solve for preschool children, and likely relies on sub-processes which mature later in childhood.

2.2.4 The Impact of Visual Experience of Real-world Structures on Categorization Ability

Adults perceive cluttered natural scenes with little effort, suggesting that the adult visual system is shaped phylogenetically and ontogenetically by the tasks and physical properties inherent in human environments (Field, 1987; Geisler, 2008; Kayser et al., 2004; Nassi & Callaway, 2009; Shepard, 1992). The visual system is most susceptible to experience-driven adaptations and refinement up until early adolescence (Fantz & Nevis, 1967; Maurer & Lewis, 2013; Siu & Murphy, 2018). From their first months of life on, children show learning mechanisms which build up a basis for perceptual organization and visual categorization (Bhatt & Quinn, 2011; Goldstone, 1998). For example, 9-month-old infants as well as children beyond four years of age have the ability to detect visual statistical regularities such as spatial or temporal co-occurrence of feature combinations (e.g., Fiser & Aslin, 2002; Janacsek et al., 2012), which enables them to individuate objects and to perceive entities as unique (Scherf et al., 2009; Wilcox & Chapa, 2004). During development, there is an increase in the ability to differentiate variations and combinations of visual properties, and more versatile distinctions between object categories or single percepts can be made (Goldstone, 2003; Smith, 1979). In parallel, frequently occurring combinations of stimuli are learned to be processed as one unit (Goldstone, 1998), leading to more efficient and faster processing of complex stimuli.

Studies investigating visual development rely to a large extent on abstract graphic stimuli, while perceptual learning occurs in a natural environment. So far, there is inconclusive evidence that, similar to adults, children show processing advantages for image properties which they experience in their everyday environment. Findings with infants younger than one year of age suggest early sensitivity for differences between distorted compared to naturally appearing textures (Balas & Woods, 2014) or images of colorful plastic objects (Balas et al., 2018). In contrast, 10-year-olds—but not younger children—showed similar processing advantages to adults for spatial characteristics of natural scenes (Ellemberg et al., 2012). This finding was explained by younger children's more immature processing of lower compared to higher spatial frequencies.

The literature on visual development clearly suggests that the kind of information drawn upon by young children during visual categorization must be affected by their maturing visual abilities. The literature also suggests that analysis of typical real-world scene characteristics such as fine details, variations in the shape of repeated elements, or complex arrangements of scene components most likely requires significant processing effort. Still, children show a great ability to perceive and react to the entities of their environment. Consequently, we assume that children with immature visual abilities categorize their environment by drawing upon visual cues which they can perceive more efficiently, and which are sufficient to determine similarities of relevant entities and structures. These cues do not necessarily overlap with those perceived by adults. The current study aims to identify visual properties children use in their categorization of naturalistic images, and how children's use of visual properties during categorization differs from that of adults who receive identical visual information.

2.2.5 Selected Visual Properties

The current investigation is based on candidate visual properties which have been found to differentiate natural scenes, textures, or superordinate categories. They were selected from the literature on visual categorization (e.g., Baumgartner et al., 2013; Geisler, 2008; Heaps & Handel, 1999; Isherwood et al., 2017; Schmidt et al., 2017) and computer vision (Clausi, 2002; Costa et al., 2012). We specifically considered properties which had been included in developmental studies (Balas, 2017; Ellemberg, Hansen, & Johnson, 2012). Selection criteria were that the properties allowed us to quantitatively capture and distinguish the kind of image structures used in the present study. Moreover, the properties must describe a delimited visual characteristic or be associated with a certain visual phenomenon. The final selection of image properties is defined in Table 2.3.

Table 2.3 Definitions of the Visual Properties

Rated [a]		
Curvature	*Angular vs. curved.*	Shape characteristic, supports classification between animate and inanimate objects (Long et al., 2017; Schmidt et al., 2017).
Depth	*Plane vs. three-dimensional.*	Indicates spatial arrangement of scene elements as it is relevant for perceptual organization.
Gloss	*Dull vs. reflecting.*	Surface property, supports classification of materials (Fleming, 2017)
Regularity	*Regular vs. chaotic.*	Characteristic for texture discrimination (Heaps & Handel, 1999; Rao & Lohse, 1996).
Size	*Small vs. large pattern.*	The magnitude of elements as depicted on the cards.
Symmetry	*Asymmetrical vs. symmetrical.*	Shape characteristic, related to living things and plant parts, attracts attention in natural scenes (e.g., Açık et al., 2009).
Computational [b]		
Alpha	Steepness of the distribution of energy across spatial frequencies.	Typical alpha levels of natural scenes are efficiently processed in adults (for further details see Figure 2.**1**).
Deviation	Deviation of a spatial frequency distribution from the fitted line defined by Alpha.	Measure of scaling-invariance (fractality; e.g., Burton & Moorhead, 1987). Distinguishes between artifacts, plants, and natural scenes (Redies et al., 2007; see Figure 2.1)
CooCor	Co-occurrence Correlation.	Repeating probabilities of neighboring pixels over the image dimensions (Haralick et al., 1973). Sensitive for low-level irregularities of naturalistic textures.
Skew	Skew of the pixel luminance histogram.	Relates to the impression of lightness, reflectance, and original colors.

[a] Rated properties were formulated as opposites and judged on a continuous scale by an additional group of adult participants.

[b] Computational properties were assessed with functions implemented in Matlab (version R2017b, http://www.mathworks.com) or provided by literature on image processing (Gonzalez & Woods, 2018).

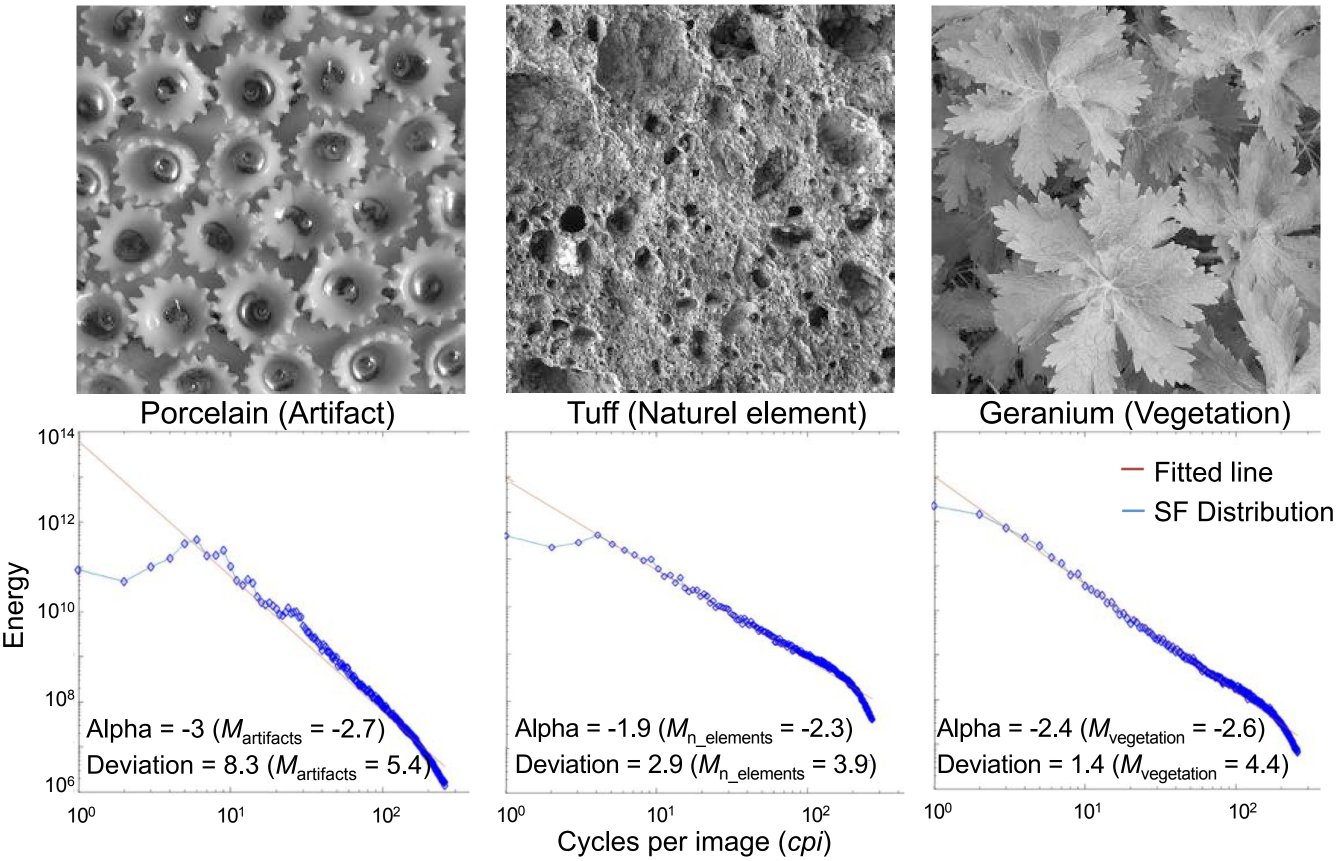


Figure 2.1: Examples of image spatial frequency characteristics.
Three images taken from the study (top row) and their distribution of spatial frequencies (*f*) plotted on logarithmic scales (bottom row). *F* is assessed by Fourier spectral analysis in cycles per image (cpi; alternations between light and dark) plotted on the x-axis, with a maximum of 256 cpi for our images analyzed in a resolution of 512 by 512 pixel. Energy refers to the magnitude of an *f* for all orientations. The fitted line falls with a slope of $1/f^{alpha}$. Porcelain (left) exhibits low Alpha, describing a steep fall of the slope, while high Deviation indicates a large variance of *f* around the fitted line. The more shallow slope of tuff (center) indicates more smaller compared lo large sized *f*, and moderate Deviation. Geranium (right) has moderate Alpha, and the distribution of *f* almost overlaps with the fitted line, resulting in low Deviation.

2.2.6 The Current Study

Here, we explored the influence of developing visual abilities on categorization by comparing preschool children's and adults' categorization of images depicting real-world structures. Participants performed two tasks in which (1) perception of similarity of one set of images, and (2) inferences about the membership of images from another set to one of three superordinate categories—artifacts, non-living natural elements, and vegetation—were assessed. These categories were chosen because they cover different aspects of human daily life and are reliably distinguished from one another early in development. The selection of visual properties described above (Table 2.3) was assessed of these images. We evaluated children's and adults' performance in each of the tasks and the effect of the assessed visual properties on their performance and categorization decisions. By doing so, we aimed to infer the impact of developing visual abilities on children's categorization by asking the following questions:

- Do preschool children and adults differ in which visual properties they attend to when sorting naturalistic structures according to visual similarity?
- Do preschool children and adults draw on different visual properties during the classification of naturalistic structures depicting artifacts, non-living natural elements, and vegetation?
- How do decisions about visual similarity during categorization relate to the superordinate categories of the visual structures?

We assumed that category formation is influenced to some degree by immature visual abilities. Thus, we expected systematic differences between children's and adults' categorization.

2.3 Methods

2.3.1 Participants

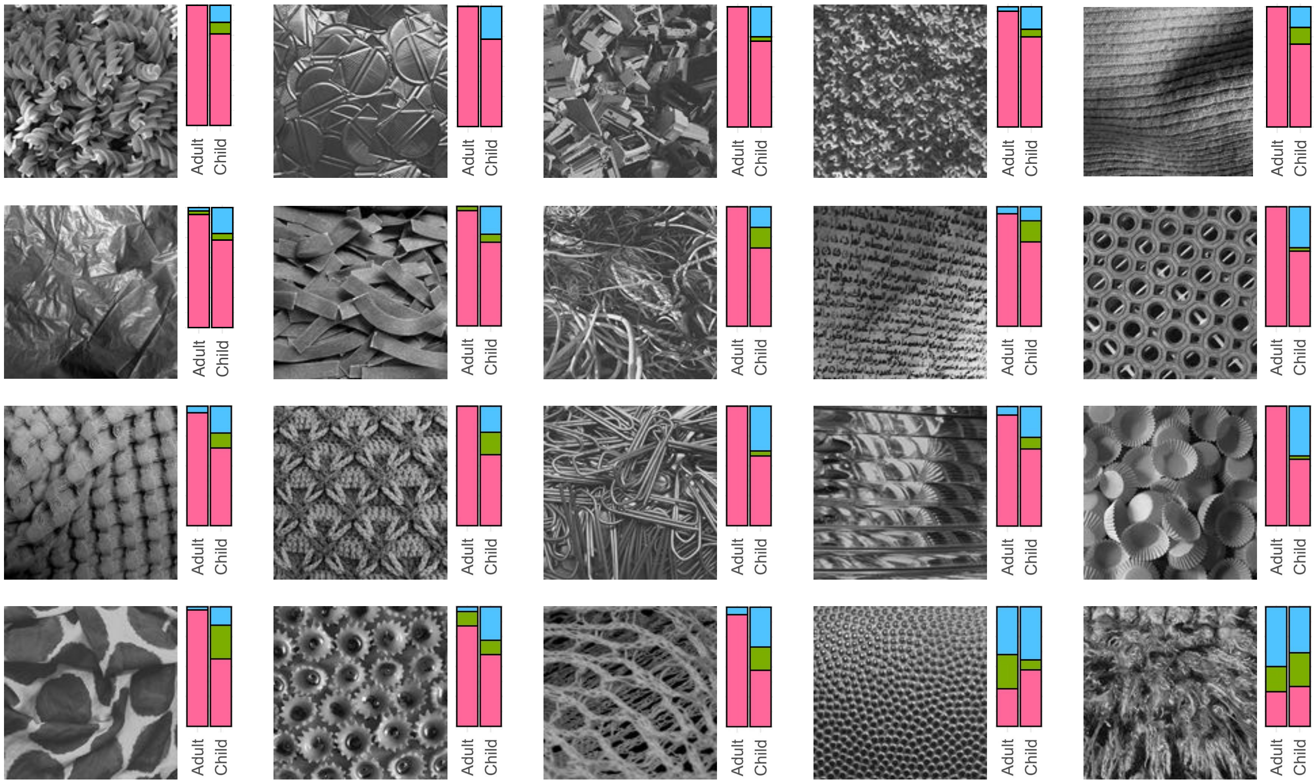
The final sample of child participants recruited from urban and suburban regions of a large European city were 76 preschool children (age: *M* = 4.8 years, *SD* = .7 years, range = 4.1 to 6.1 years; 40 female). We chose the youngest age range which could be expected to be able to perform our tasks. One additional child was invited but did not want to participate and was excluded from the analysis. Three children did not want to participate in the sorting task but provided data for the classification task, and remained in the analysis of this task. The adult sample consisted of 72 participants (age: *M* = 32.9, *SD* = 9.2, *range* = 18 to 55 years; 37 female). All participants had normal or corrected to normal vision. We chose a sample size based on a pilot study which was conducted using tasks similar to those used in the present study (see Stimulus section below for details). Participants were tested either in a laboratory or in a day care center. Participants who were invited to the laboratory received 10 Euros while the day-care center received a general donation of toys and books. All children additionally received a participation certificate.

All procedures involving human subjects in this study were approved by the Ethics Committee of the Max Planck Institute for Human Development. Written informed consent was obtained from each adult participant and from a parent or guardian for each child participant before any assessment or data collection.

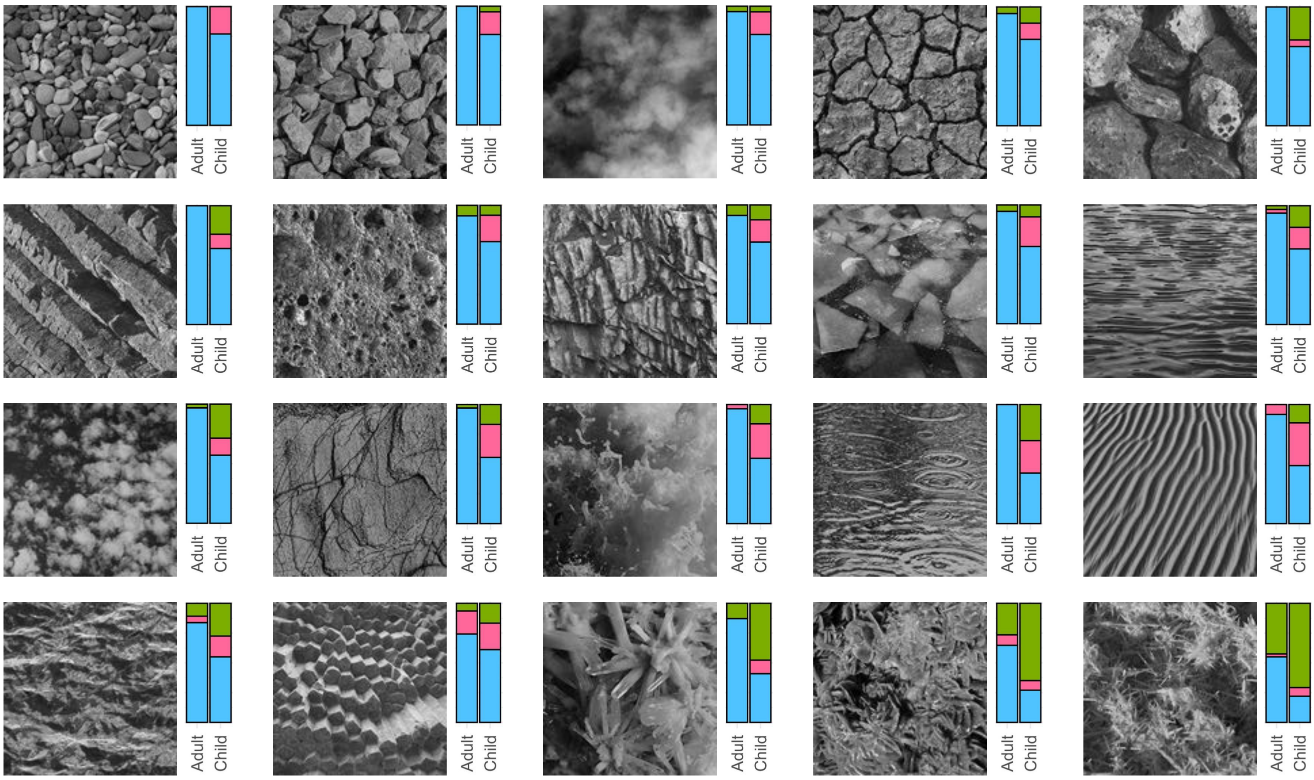
2.3.2 Stimuli

The photographs included in the study were taken by the first author using a digital camera (Canon Ixus 85 IS), or downloaded from license-free online databases. Selection criteria were that they (1) depicted extracts of real-world structures representing one of the three superordinate categories non-living natural elements (e.g., water surfaces or rocks; in the following abbreviated to *natural elements*), vegetation, or artifacts, and (2) were homogeneous in that each entity was covering the whole image space in a non-manipulated way. The images were transformed to greyscale using the software Adobe Photoshop (Version 2017.0.0). Although color contributes to the identification of natural entities (Gegenfurtner & Rieger, 2000), we decided against its inclusion because we expected color to dominate similarity perception in the sorting task and hide the impact of visual properties which are based on pattern. In our picture set, we adjusted the luminance distribution using the software Adobe Photoshop to reduce the impact of technical decisions during photographic exposure, such as overall dark or light images, by stretching each range of grey-levels to the full range of 1–256. In parallel, we took care to keep meaningful characteristics of the luminance distribution which relate to lightness and the original color of the depicted entities by keeping the averaged grey within the range of 71–183.

Artifacts



Natural Elements



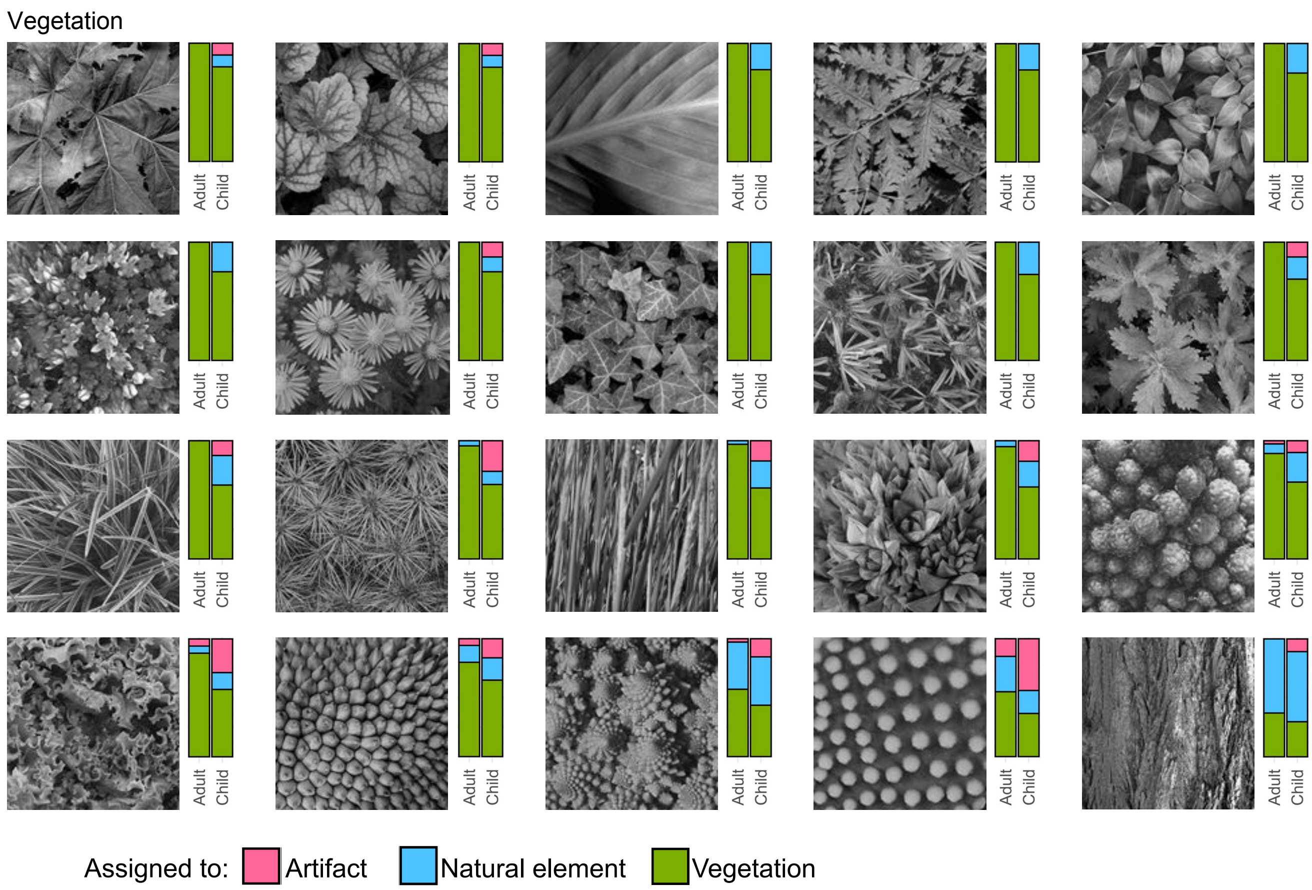


Figure 2.2: Images and their category assignments by children and adults.
The images of the current study are ordered from top to bottom by their true category, and within each category by the proportion of correct assignments of children and adults taken together. Next to each image, colored bars indicate the groups' proportional assignments to each of the categories. These proportional assignments were used as variables in the sorting task.

Classification difficulty was balanced between the categories according to the results of the pilot study in which 26 adults had categorized 141 of our images. To promote variance within a category's appearance we also matched some overarching properties between the categories by visual judgment (i.e., image complexity, irregularity, and repetitiveness of elements; Rao & Lohse, 1996), and we chose images providing approximately similar variability within the levels of the categories' visual properties. This reduced stereotypical representation of the superordinate categories and was intended to produce more perception-related variance (for a similar approach see: Schmidt et al., 2017).

The selection process resulted in a final set of 60 images with 20 images per category. All images are shown in Figure 2.**2**. The images were then downsized to 1024 × 1024 pixel, printed on card board photo paper and laminated to obtain sorting cards of size 9.7 × 9.7 cm. A square format was chosen to reduce biases of orientation and gravity. These cards were split into two halves so that different images could be used in the two categorization tasks, respectively (see Procedure section).

2.3.3 Procedure

The experiment started with a card sorting task in which participants were asked to group images together according to visual similarity. Directly after sorting, adult participants received a questionnaire asking about criteria for their similarity judgments. After a short break, a classification task followed in which participants were asked to place images into one of three boxes representing the categories vegetation, natural elements and artifacts, respectively. Stimuli and tasks were identical for children and adults except that children were told cover stories around the instructions. For the sorting task, the story included a bat puppet who liked image patterns but could not sort the cards itself. When the child finished sorting, the bat happily flew over the sorted image groups (see Supporting Information (SI) for the precise instructions of the sorting task, Table S 2.1 and Section S1 Results).

Sixty images were selected to obtain as many exemplars of each category as possible with respect to the attentional span of preschoolers. Before each session the full set was separated into two halves of 30 cards with equally balanced categories, so that each participant viewed different images during the two tasks. The separation of cards followed a routine which provided similar probabilities for each image to be sorted into a group with any other image during the sorting task. We also ensured that the cards given to the participants for sorting were always balanced over the three categories. In preparation of the classification task, the other half of the images were shuffled to a random order and piled to a stack.

2.3.3.1 Design of the Sorting Task

Participants were instructed to group images together that they perceived as visually similar. Visual similarity was asked to be judged subjectively without taking the identity of the objects depicted on the images into consideration. There were no restrictions on the number of groups or the number of images within each group. Nine cards were already placed on the table in a circle at the beginning of the task, and participants were asked to point to a card pair that they perceived as visually similar. The card pair was put aside and replaced by two new cards. This was repeated until all cards were assembled. Participants could request additional cards to be put on the table if they did not find a match. Furthermore, if we noticed that participants were only assembling pairs of two cards we reminded them that it was also possible to add cards to these pairs if wished (see SI, Table S 2.1).

2.3.3.2 Design of the Classification Task

Participants were seated in front of three boxes. They were told that cards with different images would be presented to them. In contrast to the sorting task, we pointed out that it was now important to attend to the things depicted on the cards, and guess into which of the three boxes the card belonged. Definitions of the categories represented by the boxes were then given orally. These definitions were adapted in their wordings to be suitable for the age groups (see Table 2.4). Participants were then shown one image after another and asked to decide for one of the boxes even if they did not fully classify the depicted things. Participants did not receive feedback regarding the correctness of their decision, but the children were praised periodically and were reminded to look at the images carefully.

Table 2.4 Category definition during the classification task.

Category	Definition for Adults	Definition for Children
Vegetation	"Cards that depict plants or parts of plants."	"Anything that you think is a plant or tree."
Artifacts	"Everyday objects and utensils – everything that you would say isn't naturally occurring, but has instead been produced by people."	"Things that are man-made. You would know them from seeing them in your kitchen shelves or in your bedroom, as parts of houses or on the street."
Natural Elements	"Things that you would know as natural materials – that is, things that you would see in a natural environment, and which are not plants, nor animals, nor manmade objects."	"Natural things – things you might see in the mountains, or by the sea. However, plants can't go in there, since they belong in the first box, and the same goes for man-made things, which have their place in the second box."

2.3.4 Data Preparation

We noted the images which were assembled within a group by each participant during the sorting task. All possible combinations of images within a group were classified as pairs of images perceived as similar by the participant. Occasionally, a participant did not find a match for an image. These individual images stayed in the analysis. We then calculated proportional similarity matrices by dividing the frequencies of joined image pairs by the number of participants who had received the respective images with the 60 images on the x- and y-axis, separately for children and adults.

In the classification task, we noted participants' answers for each image, no matter if it was correct or not, leading to the variables *Assigned Artifact, Assigned Natural Element,* or *Assigned Vegetation*.

2.4 Results

Data comparisons in both tasks were conducted as analysis of variance (ANOVA; R-package ez; Lawrence, 2016) or generalized linear model (GLM) with the function glm implemented in R (R Core Team, 2019). In the classification task, we conducted generalized linear mixed effect models (GLMM) with the function lmer (R-package lme4; Bates et al., 2015) if the comparisons included multiple levels of comparisons. Binomial error structures were defined for GLMs and GLMMs. Residual and specification diagnostics of the GLMMs were carried out with the R package DHARMa (Hartig, 2020). Influential cases within the units of our comparisons were diagnosed with regard to DFBetas (function influence; R-package lme4), and Cook's D (R-package influence.ME; Nieuwenhuis et al., 2012). In the GLMMs, we assessed the significance of predictors by comparing the current model with a model reduced by the respective predictor with the R-function Anova (package car; Fox & Weisberg, 2019), which provided p-values for fixed effects based on Chi-square likelihood-ratio test. In all comparisons, p-values $\leq .05$ were considered as significant.

The data underlying the statistical analysis of this study is accessible under the link https://osf.io/8xy5n/?view_only=6ddced286c31456fae7d20dd86e072e6.

2.4.1 Classification Task

We pursued several goals with the analysis of the classification data. First, we assessed performance measures as a function of participant group (children, adults) and image category (artifact, natural element, vegetation). The rationale for this was to determine the difficulty of the task and sensitivity for the three categories. Then, we assessed the impact of visual properties on participant groups' assignment of category membership to each of the images. This was done to infer which visual property affected classification. Next, we compared visual properties between the participant groups to understand which visual properties were specifically affecting children's decisions. Finally, we assessed six indices for each image, representing the proportions to which a category was assigned to an image by either participant group. These indices were included in the analysis of the sorting task. We therefore begin by reporting the classification results, even though the classification task was conducted after the sorting task in the experimental procedure. We first report the children's results, followed by those of the adults.

2.4.1.1 Children's Classification Task Results

Each of the 76 children who participated in the classification task sorted a complete set of 30 cards into artifact, natural element, and vegetation boxes. In sum, they correctly classified $N = 1586$ (69.6%) of a total of $N = 2280$ images. No difference in the proportion of correctly classified images was found between girls (71.3%, $SD = 45\%$) and boys (67.5%, $SD = 47\%$; $t = 1.3$, n.s.). The continuous variable Age, however, predicted children's proportion of correctly classified images ($F(1) = 10.8$, $p = .001$). We therefore included the covariate Age in the GLMMs conducted in the later analysis of children's assignment of categories.

Classification Performance Children. A confusion matrix (Table 2.5) shows the structure of responses to each of the presented images. Children most correctly assigned cards in the vegetation category, and least in the natural elements category. We assessed the sensitivity measure d-prime (d'; Wickens, 2002) for each of the true categories (Table 2.5). Higher values of d' indicate a better discriminability of one category from the others. Analysis of variance of the three categories on d' revealed that children's sensitivity differed between categories ($F(2, 150) = 20.6$, $p < .001$, $\eta^2 = .07$), in that sensitivity for natural elements was lower than for vegetation and artifacts (both adjusted $p < .001$; Post-hoc Tukey's HSD test), while sensitivity for vegetation and artifacts did not differ.

Table 2.5 Classification performance of children and adults.

Image category	Assigned category [a]						Decision measures	
	Artifact		Natural element		Vegetation		Sensitivity [b]	
	M	SEM	M	SEM	M	SEM	d'	SEM
	Children						Children	
True artifact	**6.88**	**(0.26)**	2.89	(0.23)	1.67	(0.16)	1.70	(0.10)
True natural element	2.22	(0.20)	**6.34**	**(0.24)**	2.39	(0.17)	1.33	(0.09)
True vegetation	1.45	(0.11)	2.04	(0.23)	**7.64**	**(0.21)**	1.78	(0.07)
	Adults						Adults	
True artifact	**9.08**	**(0.76)**	1.19	(0.62)	1.05	(0.21)	2.99	(0.05)
True natural element	1.13	(0.35)	**9.07**	**(0.89)**	1.43	(0.70)	2.63	(0.07)
True vegetation	1.20	(0.42)	1.49	(0.68)	**8.99**	**(1.00)**	2.76	(0.07)

Note. Participants' category assignments (left), and decision measure (right) as functions of true image categories, separate for children (N = 76) and adults (N = 72).
[a] Cells show responses given by one participant, averaged over all participants within the participant group (*M*) and their standard error (*SEM*). Rows correspond to the true image categories. Diagonal cells in bold case indicate correct responses (*hits*), while the remaining cells in a column indicate if an image was assigned to one category, but belonged to one of the remaining categories termed false alarms (*fa)*. The remaining cells of a row indicate correct rejections of the true category, respectively. Decision measures are averaged over all participants within the participant group.
[b] Sensitivity (*d'*) was calculated by subtracting the normalized proportion of the sum of false alarms from the normalized proportion of hits, *d'* = z(hits) - z(sum(fa)).

Visual Properties Predicting Children's Category Assignment. In order to extract visual properties that predicted a child's assignment of an image to one of the three categories, we conducted GLMMs with a binomial error structure. Participants' responses were binarily coded, resulting in three DVs which indicated if an image was assigned to one of the categories or not (1, 0). To account for intra-class correlation, we included participants and images as the units of random intercepts. Visual properties were included as fixed effects, and age was included as covariate in all models. We assessed the impact of visual properties on category assignment by including each visual property individually in a model, resulting in 10 tests. These tests equivalently explored the predictive value of each property. The reason for choosing separate analysis of visual properties were 1) missing prior expectations about their significances, and 2) the fact that interrelations of simultaneously included IVs in a full model might obscure the impact of some predictors (correlation matrix shown in S2 Results, Figure S 2.**2**). Alternative methods like the agglomeration of properties by PCA might hide their unique contributions, whereas variable selection with specialized methods like lasso regression (Groll & Tutz, 2014) could generate different results if additional properties than those currently selected were included. These alternative types of analysis would not be as appropriate for the exploratory approach taken here. We controlled the false discovery rate in multiple comparisons by adjusting *p*-values within the 10 tests conducted for each DV using the method of Benjamini and Hochberg (Benjamini & Hochberg, 1995).

Visual properties that contributed to children's category assignment are specified in **Figure 2.**3a. Overall, the visual properties depth, symmetry, skew, and deviation were drawn upon by children during the classification task. Children's assignment to artifacts was best predicted for images with high skew, high values of deviation, and low depth. Assignment to natural elements was mainly predicted by low symmetry, whereas the assignment to vegetation was predicted by high depth. In particular, children drew upon pictorial depth cues in their decisions about category assignment. Although alpha was an important predictor of the categories in our image set (see S2 Results), it was not drawn upon significantly by children. All coefficients are reported in S3 Results.

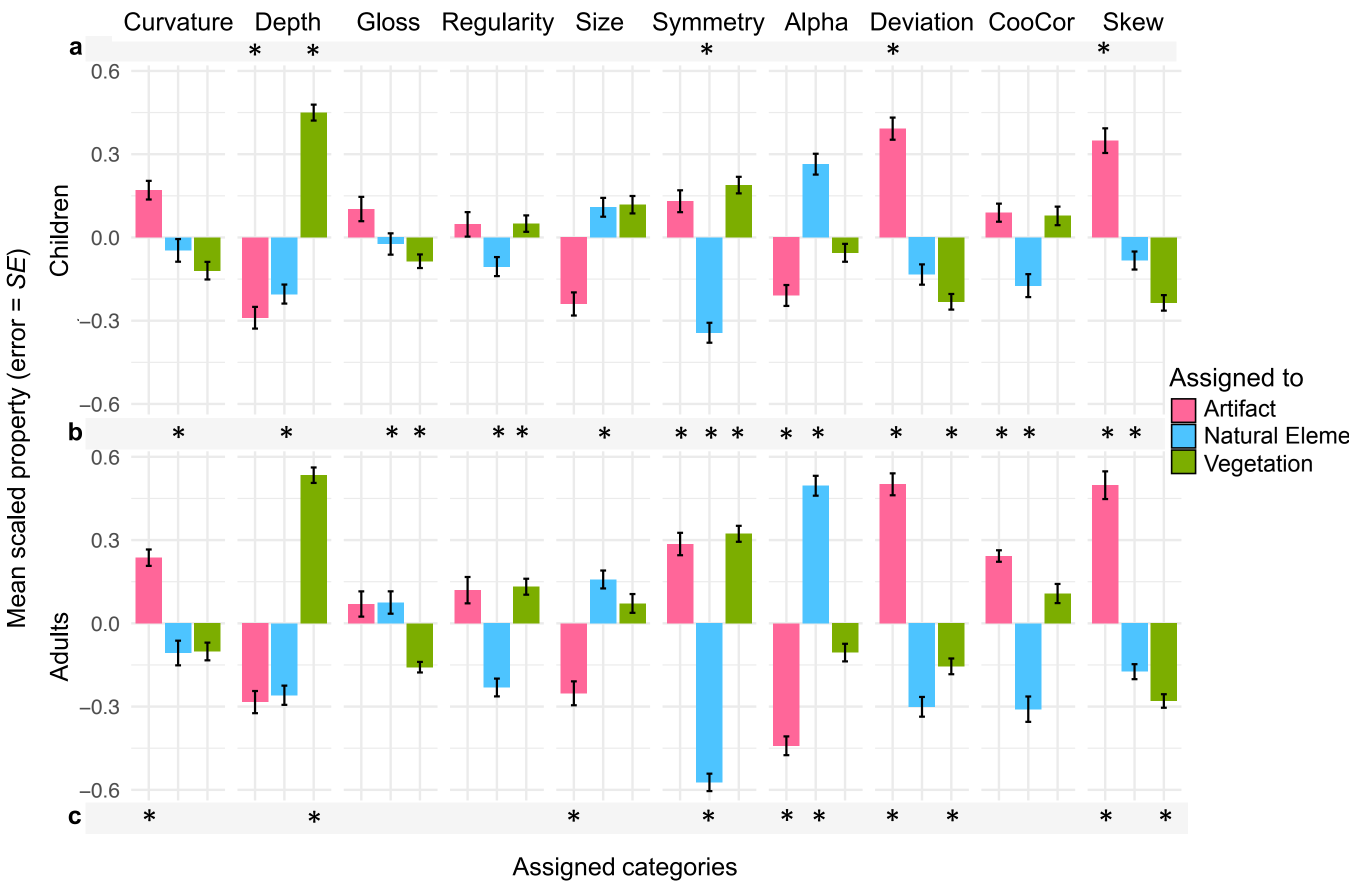


Figure 2.3: Visual properties as functions of assigned categories by children (top) and adults (bottom).

Properties are z-standardized and averaged within assigned categories. Asterisks in (a) and (c) indicate significant main effects of a visual property in the GLMMs conducted on the respective assigned category for children (a) and adults (c). Asterisks in (b) indicate significant interaction terms between participant group and visual property, of the GLMM conducted with the data of both participant groups. All adjusted *p* < .05 (method: Benjamini and Hochberg,1995). Coefficients and *CI* are provided in S4 Results, Table S 2.5Table S 2.7.

2.4.1.2 Adults' Classification Task Results

Each of the 72 adult participants sorted a full set of 30 cards into artifact, natural element, and vegetation boxes, which resulted in a total of $N = 2160$ images sorted. Adults correctly classified $N = 1586$ (90.5%) of the sorted images.

Classification Performance Adults. A three level within-subject ANOVA on *d′* showed that adults' sensitivity for the categories differed, $F(2, 142) = 21.4$, $p < .001$, $\eta^2 = .07$, in that sensitivity for artifacts was higher than for vegetation and natural elements. No other contrasts were significant. Inspection of the confusion matrix in Table 2.5 (bottom left) suggests that this effect can mainly be attributed to fewer incorrect assignments to the artifact category.

Visual Properties Predicting Adult's Category Assignment. Adults drew upon the visual properties deviation, skew and alpha, as well as upon curvature, size, depth, and symmetry when assigning categories to the images. Visual properties which significantly predicted adults' assigned categories are specified in **Figure 2.**3c. Although adults correctly classified 90% of the images, some differences remained when comparing the visual properties predicting adult category assignment with those predicting our true image categories as shown in SI S2 Results. In particular, adults included curvature and size in their assignments to the artifact category although these properties did not predict the true category of artifacts in our image set.

2.4.1.3 Comparison of Children's and Adults' Results in the Classification Task

The correlation of correctly classified images between children and adults was $r(59) = .74$, 95% $CI = [0.59, 0.83]$, $p < .001$. Children's sensitivity for the image categories was generally lower than adults' sensitivity. Of great interest for our study were differences between the participant groups within particular visual properties or image categories. These particular differences could specify the effect of immature visual abilities on visual categorization in children. We therefore conducted comparisons between children and adults which included the factor group (children, adults) and an interaction term between the IV of interest and the factor group.

A 2 × 3 ANOVA on the sensitivity measure *d'* with the factors category (i.e., true image category) and group confirmed the general difference in sensitivity between children and adults with a main effect of group ($F(1, 146) = 164$, $p < .001$, $\eta^2 = .45$). Moreover, there was a main effect of category ($F(2, 292) = 33.8$, $p < .001$, $\eta^2 = .055$) and an interaction between group and category ($F(2, 292) = 7.6$, $p = .001$, $\eta^2 = .013$). Post-hoc comparisons (Tukey's HSD) indicated that within both participant groups, classified natural elements were classified with lower sensitivity than artifacts and vegetation (both $p < .001$). Adults' sensitivity for vegetation was lower than children's sensitivity, if compared to the other categories within groups ($p = .02$; see Table 2.5). No other contrasts were significant.

We also compared the impact of the specific visual properties on the assignment of categories between children and adults. GLMMs included the IVs group, a particular individual visual property, their interaction term, and the random effects image and participant id. **Figure 2.**3b indicates visual properties that significantly differed between children and adults. Children differed from adults in their inclusion of all visual properties in at least one of the three assigned categories, in particular in their assignment of images to the natural elements category. Symmetry differed between the participant groups in all of the assigned categories, whereas depth, curvature and size only differed in natural elements.

All significant interaction coefficients, except for deviation on the assigned vegetation category, indicated a weaker inclusion of the property for children compared to adults (all adjusted $p < .05$; SI S3 Results).

2.4.2 Sorting Task

The sorting task measured participant's attention to visual information (e.g., pattern, shading, the arrangements of elements) apart from the identity of the depicted entities. We aimed to identify visual properties relevant for the discrimination of naturalistic structures. However, reference characteristics to which similarity is judged without further instruction can vary between individuals (A. Tversky, 1977). These variations might have been even more emphasized with our images because the perception of photographic images might be influenced by experiences with the objects depicted on the photographs. In the analysis, we therefore also explored whether the categories which were assigned to the images by the participant groups in the classification task had affected the similarity judgments (Figure 2.**2**).

Similarity judgments can be based on dimensions of image characteristics, in which similarity relies on similar levels within a property's continuum (e.g., Nosofsky et al., 2017; Rao & Lohse, 1996). Alternatively, similarity judgments can be based on the salience of a set of unique attributes which images share in a context (e.g., Heaps & Handel, 1999; Tversky, 1977). In order to include these different aspects, we applied the statistical methods hierarchical cluster analysis (HCA; e.g., Friesen et al., 2015; Hair et al., 1998), and non-metric Multidimensional Scaling (nMDS; e.g., Läge et al., 2011), which allow a dimensional as well as a categorical perspective on multivariate data. The similarity data is accessible under the link https://osf.io/8xy5n/?view_only=6ddced286c31456fae7d20dd86e072e6.

We first conducted HCAs using the R-function hclust (R Core Team, 2019) with the Ward2 agglomerative clustering method (Murtagh & Legendre, 2011; Ward, 1963) separately for children and adults. HCAs determine a progressive series of more inclusive clusters, and the Ward2 clustering method attempts to merge clusters which lead to a minimum total of within-cluster variance. The obtained HCA solutions were then related to image characteristics. Next, we conducted nMDS specified for ordinal data with the R-function mds (package smacof; De Leeuw & Mair, 2011) separately for children and adults. The fit of the nMDS was evaluated according to suggestions of Mair et al. (2016). We will again report the children's results first.

2.4.2.1 Children's Similarity Sorting

Seventy-three children participated in the sorting task. 70 children received the full set of 30 images, while 3 children stopped participating after receiving 20, 24 and 24 images, respectively. Children sorted between 15 and 30 images into groups (M = 29.5, SE = .24). The remainder of the images were not sorted because the child terminated sorting early, or they were not included into a group because no match was found. Each child sorted their images into M = 9.8 (SE = 0.3) groups which each included M = 3.2 (SE = 0.1) images. In sum, child participants assembled 2153 images to 2874 image pairs.

Children's HCA. The dendrogram of the children's HCA is shown in

Figure 2.4 (left). The scale at the y-axis indicates the distances between clusters which are merged at a certain height. Inter-cluster distances of the children's sorting data ranged from a minimum of .18 between the two most similar images to a maximum of 1.89. In order to assess the impact of visual properties on the participants sorting decisions, we added the visual property values of the images to the data indicating the images' cluster membership. The impact of a visual property for each step in the clustering hierarchy was assessed by calculating the proportion of variability between the individual clusters to the total variability of the property, specified by R2 (frequently termed "explained variance"; for a similar approach see: Friesen et al., 2015). Higher levels of the resulting R2 values indicate a stronger variation of the visual property between clusters, interpreted as a stronger impact of the property on participants' sorting decisions. The impact of assigned categories was assessed with the same procedure.

In Figure 2.**5** (top row) we plotted the visual properties' R^2 values as a function of the height of the dendrogram. The top left of Figure 2.**5** illustrates the development of the impact of visual properties on children's sorting decisions. At the origin of the children's x-axis, each of the 60 images belonged to an individual cluster, resulting in values of $R^2 = 1$ for each of the properties. Visual inspection indicates that the impact of particular visual properties started to vary late with increasing height at about height .4 (56 Clusters). Moreover, visual properties alternated in the strength of their impact depending on the number of clusters in which images were organized, and on the corresponding inter-cluster distance. To evaluate the overall impact of the visual properties on children's similarity judgments, we included the R^2 values of each step in the agglomeration process (60 to 2 Clusters) in one test. Analysis of variance showed a main effect of visual properties on R^2, $F(9, 522) = 77$, $p < .001$, $\eta^2 = .03$. Post-hoc analyses using Tukey’s HSD indicated that depth had the highest impact on children's similarity decisions, differing from all other properties, all $p < .005$. Skew with the second highest impact and regularity with the third highest impact differed from all other properties except each other, all $p < .005$. Gloss and alpha had the lowest impact on children's similarity sorting (see Figure 2.**7**, for all contrasts SI, S4 Results).

The impact of assigned categories on children's similarity decisions is depicted in Figure 2.**5** bottom left. Variability between categories increased late around height .75 (30 clusters). From here onward, the assigned vegetation category explained the largest proportion of variance compared to the other categories, with a value of $R^2 = .37$ at maximum height (2 clusters). An ANOVA with the factor assigned category on R^2 revealed a main effect for assigned categories ($F(2, 116) = 26$, $p < .001$, $\eta^2 = .03$), qualified by high levels of vegetation which differed from the two other categories ($p < .001$), while artifacts did not differ from natural elements within children (Figure 2.**7** top right). These results confirm our assumption that children's similarity judgments were affected by the categories they perceived in the images.

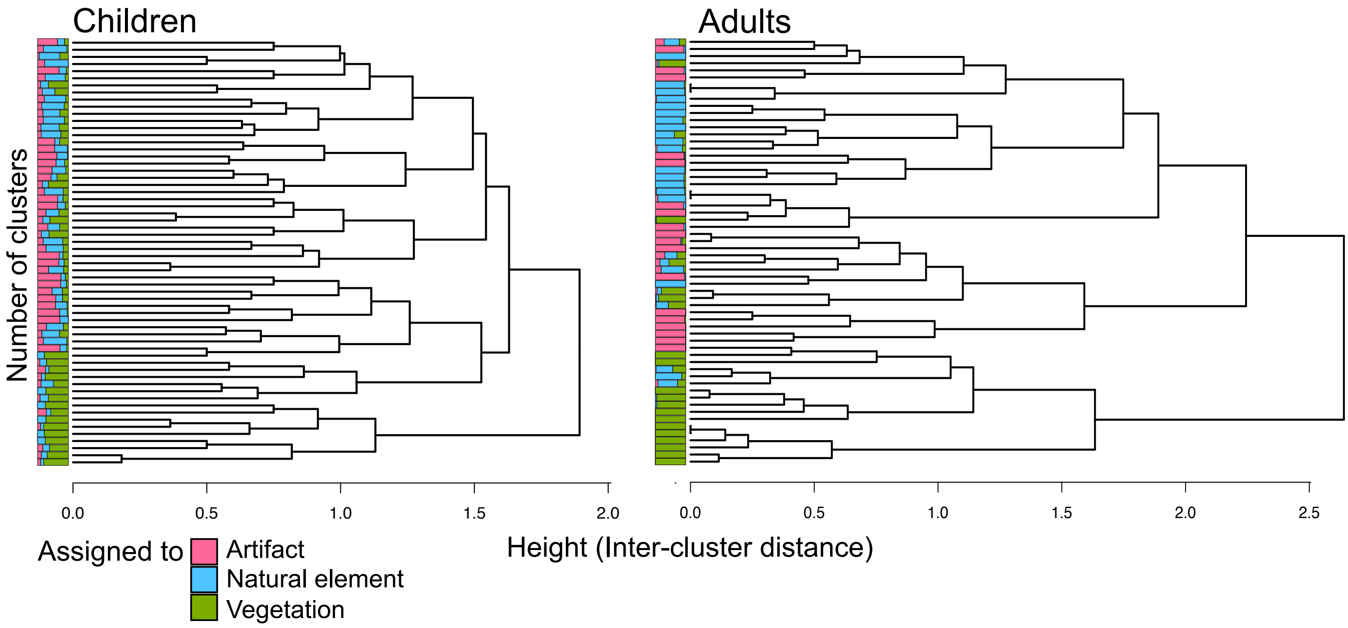


Figure 2.4: Hierarchical clustering results of the sorting task.
Dendrograms representing the structure of image similarities as hierarchical clusters received from the children's (left) and adults' (right) sorting task. For each dendrogram, zero height at the origin of the x-axis was the starting point from which individual images were agglomerated to decreasing numbers of clusters (method Ward2). Colored bars represent the proportion with which individual images were assigned to one of the categories (see Figure 2.**2**). The levels of height indicate the dissimilarity of the merged image clusters (inter-cluster distance).

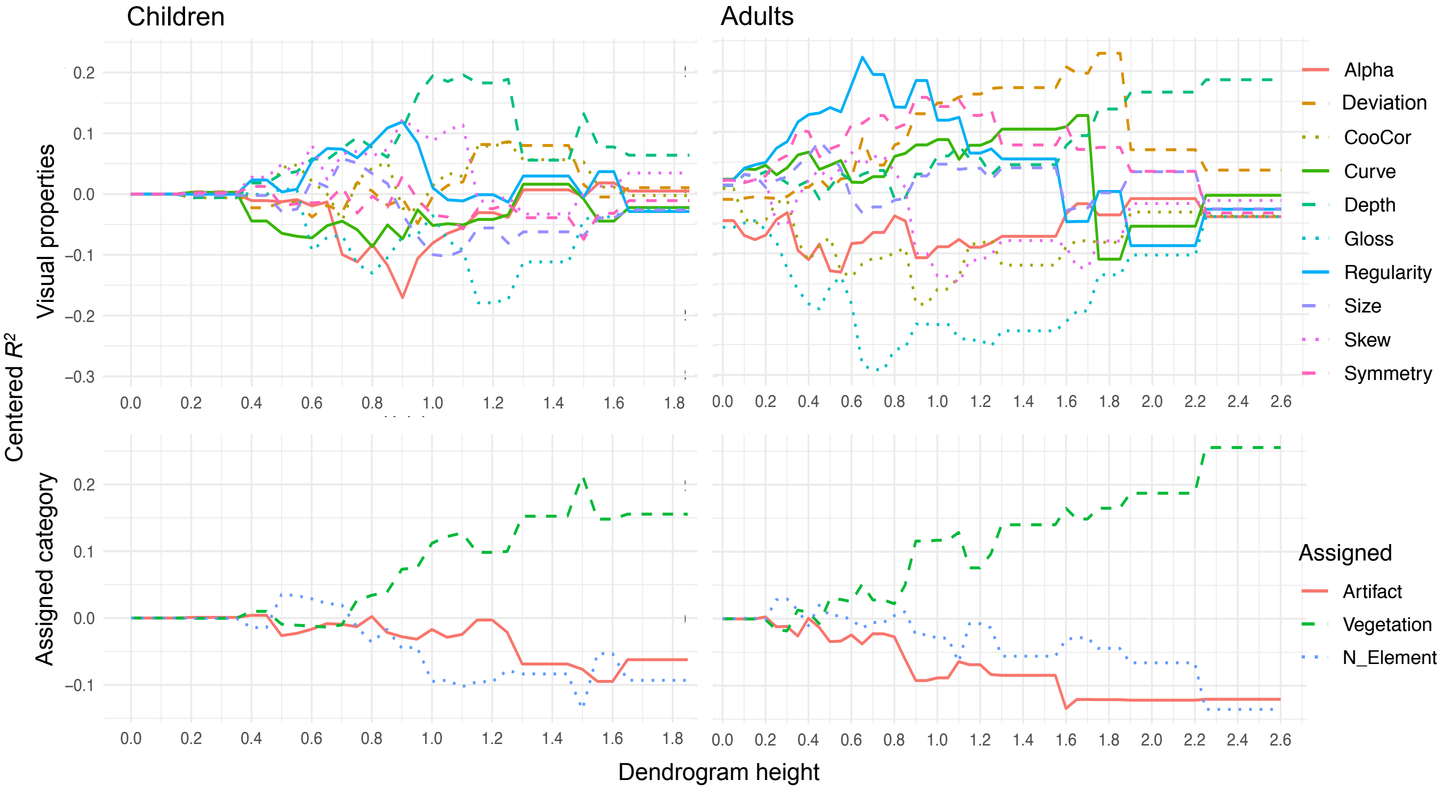


Figure 2.5: The impact of image characteristics on participants' sorting as function of the number of clusters.
Levels of R^2 for image characteristics as a function of the height of the dendrogram, for children (left) and adults (right). Height (x-axis) indicates the inter-cluster distance of images merged within successive hierarchical clusters. R2 (y-axis) represents the impact of each of the image characteristics on similarity judgments (i.e. explained variance), separately for the 10 visual properties (top row) and for the categories assigned to an image in the classification task (bottom row). R^2 was assessed in steps of height .05. For better comparability of the differences between the properties, R^2 values were centered for each indexed height. A detailed discussion on the impact of image characteristics is provided in SI, Section 1 in S4 Results.

2.4.2.2 Adult's Similarity Sorting

Seventy-two adults participated in the sorting task and received the full set of 30 images. Adults sorted between 26 and 30 images into groups (M = 29.3, SE = .12). The remainder of the images were not sorted into groups because no matches were found. Each adult sorted on average 7 (SE = .3) different groups which included on average 4.7 (SE = .2) images. Adults assembled a total of 2110 images to 4592 image pairs.

Adults' HCA. The right dendrogram of Figure 2.4 shows the HCA conducted on the adult sorting task data. Inter-cluster distances range between close to zero and the maximum of 2.64. The median of the decrease in the number of clusters lies at height .54. At minimum height (57 Clusters) visual properties already varied in their impact, indicating that 3 pairs of images were grouped together by adults frequently, and that these images included corresponding image properties (Figure 2.5 top right). Visual properties varied in their impact during the full cluster agglomeration process, with strong differences between groups of properties. Beyond height 1.9 (3 clusters), depth showed the highest impact until maximum height, indicating a very general impact on similarity decisions of adults. Analysis of variance of the factor visual property on R^2 values of the whole agglomeration process showed a main effect of visual property, $F(9, 522) = 121, p < .001, \eta^2 = .08$. Post-hoc analyses using Tukey's HSD indicated that regularity had the highest impact on adults' similarity judgments, differing from all other properties except symmetry, with all $p < .001$. Symmetry differed from the remaining properties except Deviation, all $p < .05$. Gloss had the lowest impact, differing from the other low impact properties alpha and CooCor, both $p < .001$ (Figure 2.**7** and SI, S4 Results).

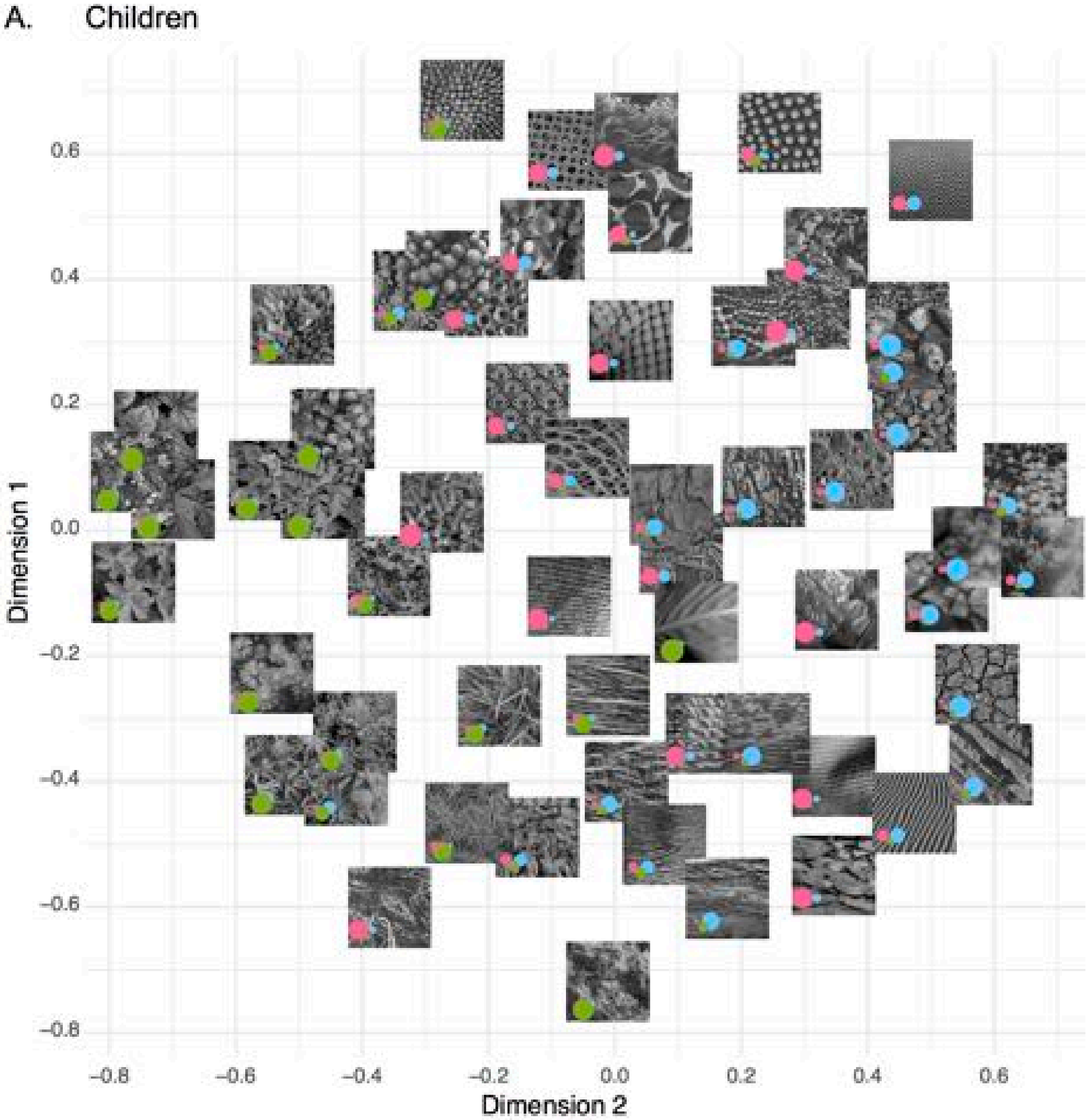
As shown in Figure 2.**5** (bottom right), assigned categories started to vary in their impact at height .2 (51 clusters) much later than visual properties. Vegetation showed a strong impact on similarity judgments during most of the clustering process until maximum height, where vegetation still explained more than 50% of the variance within the more general clusters. All R^2 values are reported in the online data repository. Analysis of variance on the whole agglomeration process revealed a main effect for assigned categories ($F(2, 116) = 33, p < .001, \eta^2 = .03$) in adults, qualified by a high impact of vegetation which differed from the two other categories ($p < .001$), and by a higher impact of natural elements compared to artifacts, $p = .006$ (Figure 2.**7**). These results show that adults relied on the category they perceived in the images when judging their visual similarity.

2.4.2.3 Nonmetric Multidimensional Scaling

In multidimensional scaling, objects are arranged in distances in a multidimensional space so that the configuration of the objects represents the distances of the underlying matrix. Including configuration plots in our analysis provides visualization of the similarity judgments related to concrete images. We decided on a two-dimensional nMDS model because it had better clarity and because scree plots did not indicate a more preferable number of dimensions (Mair et al., 2016).

Children's nMDS. Figure 2.6a shows the nMDS solution obtained for child**re**n. Clusters of images in the nMDS configurations relate to images which were assembled together frequently in the sorting task due to their perceived similarity.

Adult's nMDS. The two-dimensional nMDS configuration plot for the adult data is shown in Figure 2.6b. We will discuss the nMDS of both participant groups in more detail below.


A. Children
Dimension 1
Dimension 2
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6

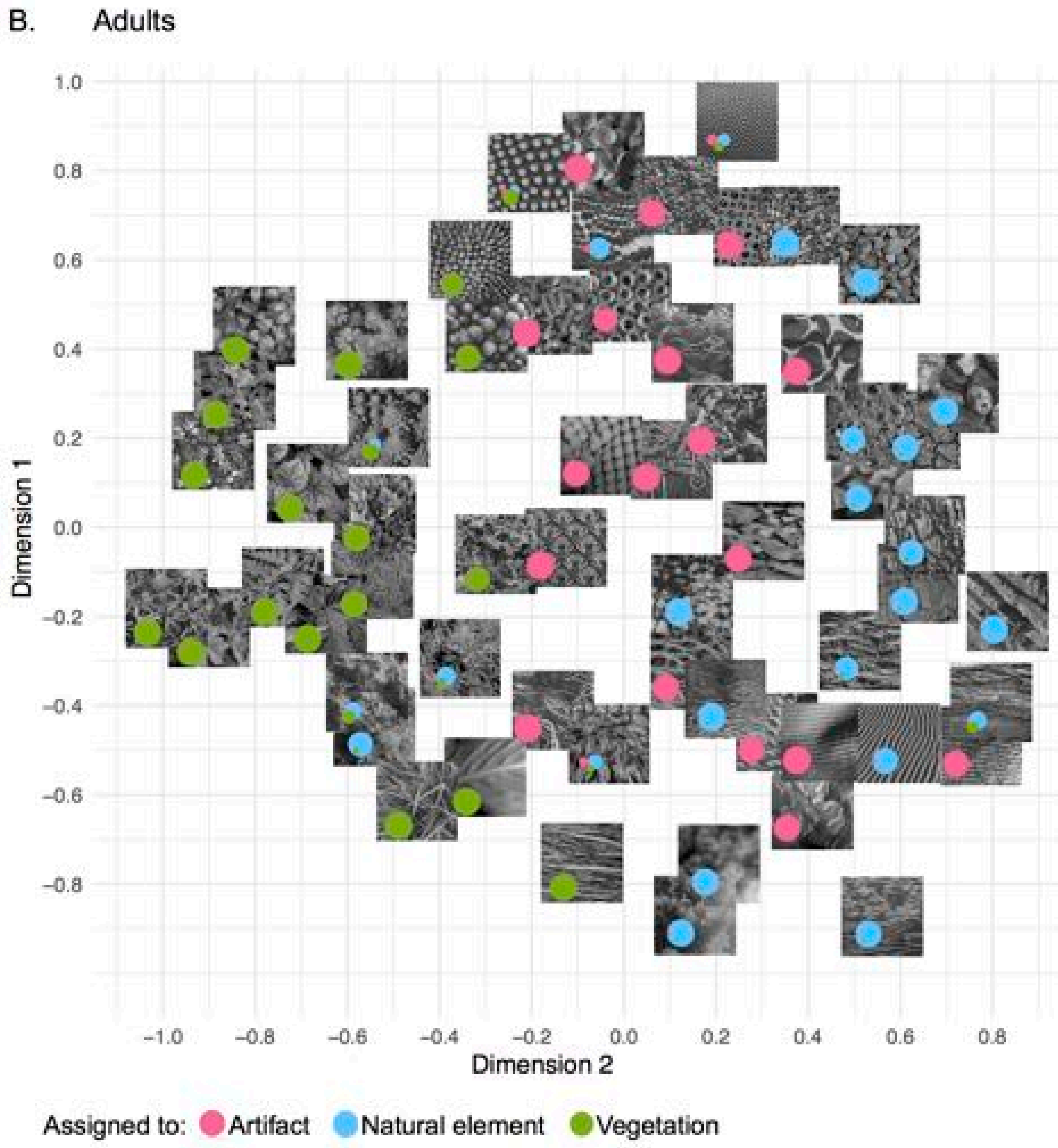


Figure 2.6: Two-dimensional nMDS solutions and the distribution of assigned categories.

Images used in the task are plotted in configurations obtained by nMDS models, for children (A) and adults (B). Stress fit index was .293 for children, and .252 for adults. We applied Procrustes transformation to the models to obtain comparability between the participant groups (congruence coefficient = .93). The size of the colored discs superimposed on each image represents the proportional assignments of categories to these images by the participant groups.

2.4.2.4 Comparison of Children's and Adults' Results: Sorting Task

To answer the question of whether children and adults attended to different visual properties in their similarity judgments, we statistically compared their HCA solutions with focus on the impact of visual properties (Figure 2.**7**). A 10 × 2 ANOVA on R^2 values of the entire cluster sequence with the factors participant group (children, adults), visual property (10 levels), and their interaction term showed a main effect for participant group ($F(1, 28) = 5.6$, $p = .025$, $\eta^2 = .14$) which was qualified by an overall greater impact of visual properties on adults' similarity judgments. We also found a main effect for visual properties ($F(9, 252) = 110$, $p < .001$, $\eta^2 = .38$), and a significant interaction between participant group and visual properties ($F(9, 252) = 51$, $p < .001$, $\eta^2 = .22$). Post hoc comparisons of the interactions revealed that differences between children and adults went in both directions. Children attended less than adults to Deviation, Curvature, Regularity, Size and Symmetry (all $p < .001$). In contrast, children included CooCor and Gloss to a greater extent than adults (both $p < .001$), while no differences were found for Depth ($p = .1$), Skew ($p = .921$), and Alpha ($p = .372$). Differences in means and *CIs* of the comparisons are provided in SI, Tables S2.12 and S2.13 in S4 Results).

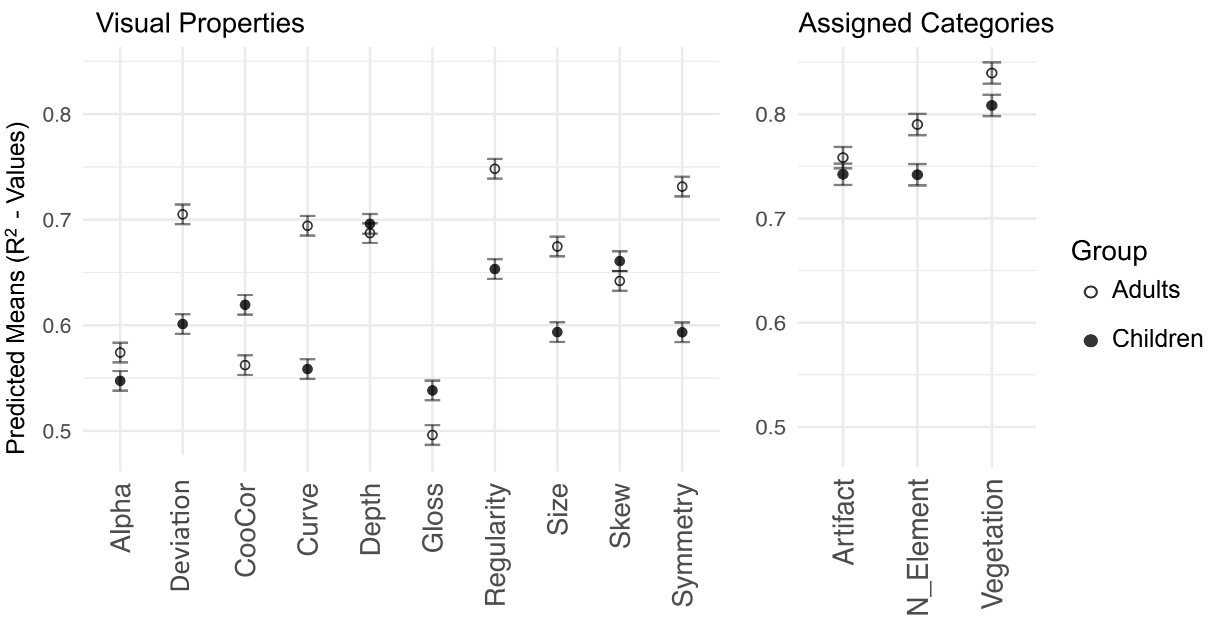


Figure 2.7**: The impact of visual properties (left) and assigned categories (right) separately for children and adults.**

Predicted means are obtained by analysis of variance. Comparisons are based on R^2 values obtained at each step in the clustering process from 60 to 2 clusters, error bars are Fisher's Least Significant Difference (FLSD). All means and the differences of post-hoc comparisons are provided in SI, S4 Results.

Inspection of the nMDS configurations allows inferences about the combinations of visual properties which might have been included in similarity decisions. This is of interest because children differ from adults in their ability to differentiate visual regularities (e.g., Smith, 1979). On first sight, nMDS model configurations of children and adults (Figure 2.6) show generally similar arrangements. For example, the bottom right corners include images with elongated, parallel elements, and the top of the configurations include images depicting small distributed particles. Next to this more general overlap, disparities in the characteristics of more direct image neighbors are noticeable. In the children's nMDS, image neighbors frequently overlapped in clearly visual characteristics. For example, the 4-image-group at the crossing of dimension 1 (D1) = .4, and D2 = -.3 to -.1 shows round, erected elements with jagged contours, or the 3-image group at D1 -.4, D2 = -.4 includes radially-configured elongated elements. In contrast, adults also appeared to use non-visual similarities to assemble images. One example is the 2-image-group at D1 = -.5 to -.4, D2 = -.2 to 0, in which two artifacts with irregular structured threads and cables are assembled in spite of their different patterns. Another example is D1 = .1 to .3, D2 = .5 to .7, where a 4-image-group of natural elements overlaps in some image characteristics (i.e., shape, material, or the subcategory stone surfaces), but none seems to be included in every image. To a lesser degree, these kinds of thematic combinations are also obvious in children, such as the two crochet works at D1 = .1 to .2, D2 = -.1. Still, these examples provide illustration of how similarity might be perceived in children and adults.

For children and adults, assigned vegetation was a major image characteristic related to similarity decisions. In the nMDS configurations, we found that both children's and adults' images were organized in accordance with their assigned categories (Figure 2.6). In particular, images on the left tended to predominantly depict vegetation. Artifacts and natural elements shared the remaining space, with less clearer boundaries. In the HCA, vegetation showed its major impact on both participant groups, especially within the organization of images to larger clusters. A 3 × 2 ANOVA on R2 values compared the impact of assigned categories on similarity judgments between the participant groups. A main effect for assigned category ($F(2, 232) = 56$, $p < .001$, $\eta2 = .25$) was qualified by a stronger impact of vegetation compared to artifacts and naturel elements, which did not differ. The participant groups did not differ ($F(1, 116) < 1$, $p = .38$), and the interaction term did not reach significance ($F(2, 232) = 2$, $p = .12$). These results indicate that similarity judgments were affected by all assigned categories, of which assigned vegetation had the strongest impact for both children and adults.

2.5 Discussion

In the present study, we investigated preschool children's and adults' categorization of images depicting real-world structures by focusing on the impact of developing perceptual abilities on categorization ability. Participants performed a sorting task in which they assembled a set of images to groups according to visual similarity, and a classification task in which another set of images were put into boxes representing the superordinate categories artifacts, natural elements, and vegetation. We then related categorization decisions of the participant groups to the visual properties of the images.

The results of both tasks show that—in spite of their still maturing visual system—preschool children readily perceive and interpret complex visual structures by 4 years of age (for 5-year-olds see also: Balas, 2017; Cohen et al., 2019; Köster et al., 2017b). Nevertheless, children's sensitivity for the superordinate categories was generally lower than adults'. Moreover, children’s lower reliance on many visual properties than adults' during classification might reflect spontaneous choices of a superordinate category. Such choices are not necessarily based on the properties children perceive in the unfamiliar images, but perhaps instead result from inattentiveness, playfulness, or a personal preference for one category type over the others. These kinds of factors are common in children's task performance and lead to more noise in the data. However, that some of the relevant properties did not differ between children and adults, or were even included to a greater extent by children, indicates that children also showed systematic sensitivities for image characteristics that seemed to depend on the image's assigned category.

2.5.1 Visual Properties Used by Children and Adults

Rated pictorial depth was an important predictor for children's assignment of images to artifacts and vegetation. In the sorting task, *depth* had the overall highest impact on children's similarity judgments, directly followed by *skew*, and the impact of both properties did not differ between children and adults. This was also the case for the inclusion of depth in the assignment to vegetation. The importance of depth cues for sensorimotor coordination and the early onset of sensitivity for pictorial depth (Kavšek et al., 2012) are both in line with the high impact of depth on children's sorting decisions. However, previous research has reported immature analyses of depth in 3D-figure-drawings for young children (Freud & Behrmann, 2017). It is possible that the depth cues in our naturalistic images were easier to perceive for children because, in contrast to Freud and Behrmann (2017), they are supported by shading information. Depth in the current images additionally refers to real-world distances between structure-elements rather than to the 3-dimensionality of the elements (e.g., leaves of a plant are flat, but assembled in space). The availability of shading information could have made the analysis of fine-grained and distributed depth information (e.g., contour junctions) unnecessary, which may be more demanding for young children.

Skew refers to an overall impression of shade within an image. Similar to depth cues, variations in skew were possibly instantaneously perceived without the necessity to attend to details. This might have allowed a spontaneous visual comparison of both depth and skew for children as much as adults during similarity perception.

In contrast, *symmetry* and *curvature* cues, which are both shape related properties, had high impact on adults' similarity judgments, but were included to a significantly lesser extent in children's similarity perception. Moreover, symmetry also differed strongly between children and adults in the classification task—in spite of its importance as a predictor of natural elements in our image set (see SI, Section 1 in S2 Results). There is evidence that the development of symmetry perception is still ongoing during preschool age (Bornstein & Stiles-Davis, 1984), whereas curvature cues were already found to be perceived during infancy (Kellman & Arterberry, 2007). The literature suggests that the most likely explanation for the lower impact of these shape properties could lie in the cognitive demand on young children raised by a necessary attention to distributed details such as contour junctions or texture elements. The integration of detailed visual cues into larger shapes may be difficult for children because of their immature vernier acuity and less efficient read-out of visual information (e.g., Dekker et al., 2019; Skoczenski & Norcia, 2002). Neuroimaging studies additionally support this explanation by indicating that adult-like efficiency of visual object processing is not reached before school age (Dekker et al., 2011, 2015; Gathers et al., 2004). In order to find additional support for this explanation, we conducted a post-hoc analysis of the children's classification data which included the continuous variables age, visual property, and the interaction term Age × Visual Property. With this, we aimed to investigate the inclusion of visual properties which rely on shape details in relation to children's age. And indeed, as shown in the SI, S5 Additional Analysis, the reliance on some detail-based visual properties changed significantly within the age-range of the child participants. The inclusion of symmetry and size became more differentiated in older children for the assigned artifact and natural element categories. Similarly, curvature, deviation and CooCor became more differentiated in older children for the assigned artifact category. In contrast, depth and skew did not show differences between younger and older children's classification, indicating that even 4-year-olds successfully relied upon these important predictors for their category assignments (see SI, S5 for the full analysis).

The impact of *regularity* on children's sorting was as high as the impact of skew. Regularity provides an overall impression of the structure of an image. Its inclusion during similarity sorting requires the ability to distinguish irregular from regular arrangements and shapes. It may be the case that this property is more accessible to children because a detailed analysis of shape is not necessary. Yet, that older children relied more on regularity than younger children in the classification task (SI, Figure S 2.3 in S5) could indicate that these arrangements of varying elements are still difficult to process in younger children. This is supported by a recent finding in an eye-tracking search task with 8-month-olds using the same images. There, the degree of regularity of a background image did not affect target detection (Schlegelmilch & Wertz, 2021)

These examples suggest that visual properties with less emphasis on detailed information and more global or holistic information had a stronger impact on children's categorization. In adults, global image features provide an instant perceptual and semantic "understanding" of a naturalistic scene (Oliva & Torralba, 2006). Developmental studies support this explanation, in that 6- and 8-year-olds were found to be faster in comparing global rather than detailed local visual information (Mondloch et al., 2003; but see e.g. Vinter et al., 2010 for more variable findings in younger children). Moreover, preschool children attend to the overall appearance of objects in categorization tasks, in contrast to adults who instead attend to category-relevant features (Deng & Sloutsky, 2016; but see: Mash, 2006). However, a stronger impact of properties which can be perceived globally in the current study does not rule out that young children attend to details as well. Their inclusion might rely on their relevance for a visual task and on the amount of details the child needs to solve (e.g., during material judgments: Balas et al., 2020). In the current study, children were confronted with many pictures present in the visual scene , so that it was perhaps more difficult to perceive details of particular images (Kimchi, 2015; Vinter et al., 2010).

2.5.2 Do Children Rely on Incomplete or Sub-threshold Visual Information During Categorization?

Our finding that children included several properties to a lower degree than adults in their categorization decisions raises the question: how were they still able to perform relatively well in the classification task?

One explanation could be that the classification of visual structures with ecological relevance relies on over-generalization of incomplete visual information. This suggestion is based on findings that young children are as able as adults to gather and adapt environmental information (e.g., Amso & Davidow, 2012). Moreover, the design of the visual system is argued to be based on task-related contact to statistical regularities of the natural environment during human evolution (Frazor & Geisler, 2006; Geisler & Diehl, 2002). In our study, the visual property *Deviation* may have been over-generalized by preschoolers. Deviation had a rather moderate impact on children's similarity perception, except at a point in the HCA clustering sequence in which vegetation increased its impact (height 1.2 to 1.5, relating to 10 to 5 clusters). Correspondingly, the inclusion of area differed between children and adults (adjusted p = .002) in that area was the only property which was drawn upon to a greater extent by children than adults in the assignment of vegetation. In contrast to the higher values of area which predicted children's assignment of images to the artifact category, low values of area must include high spatial frequencies to meet scaling invariance, which was found to be typical for natural scenes (Ruderman, 1997). Visual information that deviates from scaling invariance can therefore be seen as qualitatively different from scaling invariant visual information. The reliance on low area in children's assignments to vegetation exceeded even its actual predictive power for the vegetation category in our image set (SI, Figure S 2.**1** in S2 Results,). This might indicate that children over-generalized a category to unknown cases on the basis of visual information which was gathered from their environment. Balas (2017) came to a similar conclusion in a material categorization task, suggesting that available but incomplete visual information had led to quickly-adopted representations in 5- to 7-year-olds which then led to incorrect responses.

Yet, how could children include high spatial frequencies—which are difficult to perceive for them—into their classification decisions? One might conjecture that it is not necessary to perceive the precise spatial information provided by the full range of frequency bands included in low values of area. Moreover, sub-threshold stimuli properties might have affected classification preattentively, as they did affect the perception of ambiguous visual input in adults (e.g., Pearson & Brascamp, 2008; Sterzer et al., 2009). Behaviorally, sensitivity to sub-threshold properties might provide developmental advantages, because uncertain or overwhelmingly complex regions of the environment could then be avoided or treated with care (e.g., ambiguous visual input affected infants' search for social signals from adults; Pauen & Hoehl, 2015). Visual uncertainty can also trigger increased attention and provide a ground for adaptation and learning (e.g., specific sensitivities support the categorization of instances within the environment; Rakison & Poulin-Dubois, 2001). Adaptive reactions might in particular be beneficial for young children and infants when confronted with difficult or uncertain visual information as it is included in vegetation (Wertz, 2019).

2.5.3 The Effect of Developing Perceptual Organization

In the sorting task, participants viewed about nine images placed in front of them, while the growing number of already-assembled image groups was placed right beneath. This situation made the detection and grouping of similar images within the visual scene necessary. Possibly, one reason why children sorted smaller image groups than adults would be immaturities of perceptual integration. During childhood, neuroimaging studies found an increase of horizontal intra- and interhemispheric connectivity, as well as increasing feedback connectivity from extrastriate visual areas to V1—changes which are thought to be involved in the detection, grouping and spatial integration of distributed visual elements (Fornari et al., 2014; Knyazeva, 2013; Kovács, 2000; van den Boomen et al., 2014). Larger distances between the images would have increased the difficulty to find matches. Therefore, children preferred to compare similarity within the region around where the new images were placed, and produced more image groups which consisted of only 2 to 3 images. Non-metric MDS and HCA reflect the distinct sorting behavior of the participant groups. In the children's nMDS, there were predominantly groups consisting of 2 to 3 images, yet many separate images were distributed over the nMDS space. In contrast, images in the adult nMDS were assembled into larger constellations which were loosely connected to each other (Figure 2.6).

However, children's smaller image groups could alternatively be explained by more strict decision rules applied to similarity judgments, based on their better memory for the general appearance of images compared to adults (Deng & Sloutsky, 2016; Ofen & Shing, 2013). The difference between children's judgments and adults' more flexible property inclusions are reflected in the HCA. In the children's dendrogram, the inter-cluster distances between individual images were much higher than in the adults' dendrogram, which additionally showed a more distributed variability (Figure 2.4). These examples show how higher order visual abilities might affect the perception of elements in a naturalistic scene, which provides insight into the cognitive demand on children during categorization in every-day situations.

2.5.4 Both Age Groups Attended to Assigned Categories in their Similarity Judgments

The impact of assigned categories on similarity decisions did not differ between children and adults. Young children as well as adults seem to have processed images in a semantic fashion, while simultaneously the available combinations of images triggered category formation based on visual properties (for similar suggestions see: Hoehl, 2016; Mandler, 2000). This would indicate an equivalent inclusion of visual properties and semantic categories in similarity perception. Consistent with this proposal, answers provided by adults to the questionnaire which asked about their similarity criteria included only 55% of terms which related to visual properties, whereas the remaining terms related to the identity of or experiences with the depicted entities. In children, visual properties were mentioned in 62% of the comments referring to similarity, which were recorded during the sorting task (SI, Section 1 and 2 in S1 Results). The equivalence of semantic and property-related perception is additionally supported by a detailed comparison of the R^2 values referring to visual properties and categories, provided in the SI, Section 3 in S4 Results. In western cultures, photographic representations are understood to include at least two levels of information, which either relate to the image object itself (the sorting card), or to the entity which it represents. Children become acquainted with this cultural habit from an early age (DeLoache, 2011; Liben, 2003), and it might have been difficult for some participants to ignore the referent of the image but exclusively attend to its visual properties.

Of the three categories, vegetation was the strongest predictor for similarity in children and adults. Vegetation had a very high impact in the HCAs at organizations of the images to only few clusters, indicating a domain-like differentiation (Figure 2.5). Such a differentiation is also reflected in the nMDS solutions (Figure 2.6, and SI, S4 Results). Further, when controlling for overall performance differences in the classification task, children had higher sensitivity for vegetation than adults did. These findings suggest that preschool children possess a superordinate category representation for plants (see e.g., Gelman, 2004) and add to previous findings that preschoolers have a rich conceptual representation of plants and their properties (Backscheider et al., 1993; Hickling & Gelman, 1995; Inagaki & Hatano, 1996; Nguyen & Gelman, 2002). Research with infants provides evidence that such rich representations start to develop very early in life. Infants treat plants differently than artifacts and other natural kinds (Elsner & Wertz, 2019; Mandler & McDonough, 1998b; Wertz &

Wynn, 2014; Włodarczyk et al., 2018) and rapidly learn about plant properties such as edibility (Wertz & Wynn, 2014, 2019). These results with infants suggest that they were able to visually distinguish the plants from the other types of entities which were used as stimuli in the studies. In the current study, an age effect was lacking in children for any of the visual properties within the assigned vegetation category (see SI, S5). That the categorization of what is assumed to be vegetation in the current study is only marginally influenced by the age of the child may also be explained by the particular importance of vegetation (see e.g., Wertz, 2019).

Yet, substantial changes within the visual cortex make a direct link between infants and preschool children difficult (Siu & Murphy, 2018). Moreover, many of the visual features of plants are still difficult for infants to perceive (Ellemberg et al., 1999; Skoczenski & Norcia, 2002; Taylor et al., 2014). For example, depth cues, which were used by both children and adults in the current study to categorize vegetation, could also affect infant's visual categorization, albeit in a different way. Sensitivity to pictorial depth and coarse stereopsis is already present in infants around 6 months of age. However, fine stereopsis, which determines three-dimensional depth in the central visual area (where the plants were placed in some of the infant studies) continues to develop up into the school-age years (Giaschi et al., 2013; Kavšek et al., 2012). Therefore, future studies investigating whether infants rely on similar visual information when distinguishing naturalistic structures as those we have identified in the current study will be particularly informative.

2.5.5 Limitations

It might be argued that the images showing vegetation were perhaps more similar within their category than the images of the other categories. Consequently, this would have led to an increase in assembled vegetation images. Although we did not assess the impact of the true categories, but only the categories to which the images were assigned, this argument could still apply to the relatively large number of correctly classified images in children and adults. We therefore aimed to statistically evaluate whether the categories differed in the variance of properties they included by conducting an analysis of variance on the combined visual-property values in the image data (see SI, S2 Results). The ANOVA did not reveal a main effect for the factor category: $F(2, 8) = 1.1$, $p = .37$, *n.s*; Means (*SD*) of artifacts = .09 (1.1), natural elements = -.14 (1), vegetation = .05 (.8). This analysis shows that statistically, there were no differences in the variance of the visual properties between the categories.

It should also be kept in mind that the results of the classification task do not provide absolute descriptors of properties relevant for the classification of artifacts, natural elements, and vegetation. The current results were obtained by comparing visual property values included in the assignment of a particular category to the values of the two remaining categories. The diagnostic properties which we found therefore depend on differences between the images selected for the three categories. However, by choosing superordinate categories which differ in several conceptual and functional aspects we hope to have extracted meaningful properties.

Still, the selection of visual properties investigated here might have missed other properties that have a significant impact on categorization. For example, it could be argued that algorithms which mimic early-stage visual processing of textures (e.g., Portilla & Simoncelli, 2000) may have provided insight into substantial differences between children's and adults' perception of the images. However, we decided against the inclusion of this algorithm in the current study because our images were not well represented by it, and our restricted stimulus set was not suited for a necessary reduction of the variables obtained by the algorithm (for a similar argument see: Okazawa et al., 2015).

Importantly, it should be noted that classification decisions may have also been affected by higher-level characteristics. For example, Figure 2.**2** shows that of the stimuli from the natural elements category that children assigned to the vegetation category, crystals were the most-represented sub-category. Crystals are formed by crystal growth, which is visually reflected in their shape and resembles, to some degree, the growth of plants. Moreover, crystals might not be as familiar to children as other natural elements (e.g., pebbles or clouds). High-level characteristics were not assessed in the current study and might provide alternative, yet not mutually-exclusive, explanations for classification decisions.

Our selection of images was aimed at providing the most lifelike stimuli that could be used with our participant age groups and the type of tasks that were being performed. However, it cannot be ruled out that the inherent absence of color, movement or binocular disparity, which are important visual cues for visual categorization in a naturalistic environment, strongly influenced the results. Nevertheless, textures, as mentioned above, also provide important visual information that substantially supports the segmentation and categorization of the environment. We therefore believe that our results provide important insight into the impact of immature vision on categorization. Future studies could investigate the role of cues like color and motion in processing naturalistic images.

2.6 Conclusion

The current findings reveal visual properties that preschool children and adults use to distinguish complex naturalistic stimuli. Some visual properties were similarly included in categorization decisions by both age groups, but there were systematic differences which cannot be explained by noise. In general, children were less likely to include characteristics that relied on the analysis of detailed and distributed visual information. These findings suggest that immaturities of the developing visual system affect visual categorization through the preschool years.

Of the categories assigned to images by the participant groups, vegetation was the strongest predictor of similarity in both age groups. Moreover, children were most sensitive to vegetation in the classification task, differing from adults who were most sensitive to artifacts. Children's strong sensitivity for vegetation is consistent with recent work showing that infants respond differently to plants than artifacts or natural elements. Studies of visual properties used to process naturalistic stimuli in infancy may be a particularly fruitful line of future inquiry.

Finally, our findings suggest that during categorization, children flexibly draw upon visual signals when confronted with the ambiguity of complex naturalistic stimuli.

The data underlying the statistical analysis of this study is accessible under the link https://osf.io/8xy5n/?view_only=6ddced286c31456fae7d20dd86e072e6.

2.7 Acknowledgements

This research was funded by the Max Planck Society. During this research, Karola Schlegelmilch was a pre-doctoral fellow of the International Max Planck Research School on the Life Course (LIFE, www.imprs-life.mpg.de). We thank our participants and their parents, Britta Lewke, Jasmin Eichelsdörfer, Theresa Pokojski, Carmen Kaiser, Daniel Tatlow-Devally, Marie Wiedenhöft and the members of the Max Planck Research Group Naturalistic Social Cognition for their assistance. We thank the team of the Department of Developmental Psychology: Infancy and Childhood at the University of Zurich, Damian Läge, Karl Gegenfurtner and members of the Department of Experimental Psychology at Justus-Liebig-University Giessen, for their support.

2.8 Supporting Information

Table S 2.1 Translations of the Instructions for the Sorting Task

2.8.1 S1 Results Questionnaires

2.8.1.1 Adults' Questionnaire on their Criteria for Assembling Images.
2.8.1.2 Children's Comments during Similarity Sorting
2.8.1.3 Analysis of Questionnaires on Prior Exposure to the Categories

2.8.2 S2 Results Materials.

2.8.2.1 Prediction of the True Categories in our Image Set
2.8.2.2 Correlations between Visual Properties in the Image Set

2.8.3 S3 Results Classification Task

2.8.4 S4 Results Sorting Task

2.8.4.1 The Course of Visual Property Variance during the Cluster Agglomeration Process
2.8.4.2 R2 values of the Agglomeration Process HCA
2.8.4.4 Can Variance of Visual Properties be Separated from Variance of Assigned Categories in the Sorting Task?

2.8.5 S5 Additional Analysis: The Effect of Children's Age on the Inclusion of Visual Properties during Classification

Table S 2.1 Translations of the Instructions for the Sorting Task

German Original	English Translation
Children	
„Hast Du Lust, mit mir zu spielen?"... Experimentator*in holt die Puppe.	*"Would you like to play a game with me?" - the Experimenter picks up the puppet.*
„Erst einmal erzähle ich eine Geschichte - von Uri. Das ist Uri!" Die Puppe wird gezeigt. Sie bewegt sich kurz und winkt. „Uri kann nicht sprechen, aber er kann fliegen! Weißt Du, was Uri für ein Tier ist?" ... „Ja, eine Fledermaus. Und weißt Du, wann Fledermäuse fliegen?" ... „Genau, die fliegen, wenn es dunkel ist. Und dann sehen sie Sterne."	*"First, I'm going to tell you a story about Uri. Here he is!" The puppet is shown to the child. The puppet is moved around and waves. "Uri can't speak, but he can fly! Do you know what sort of animal Uri is?"…"Yes, he's a bat! And do you know when bats normally fly?"…"Exactly, they fly when it's dark. And in the dark, there are lots of stars in the sky."*
"Uri lebt eigentlich in einer Welt, wo es ganz besondere Sterne gibt: Mustersterne! Und die vermisst Uri sehr." Puppe wird hingesetzt. „Mustersterne sind etwas ganz Besonderes: In ihren Strahlen haben sie Bilder, die sich ganz ähnlich sind, weil sie nämlich das gleiche Muster haben. Jeder Strahl hat ein anderes Muster. Und weil Uri selbst keine Mustersterne basteln kann, kannst Du das vielleicht für ihn machen. Ich habe mir ein Spiel ausgedacht, damit es für Dich einfacher ist. Hättest Du Lust, Uri zu helfen?"	*"Now, Uri comes from a world where there are very special kinds of stars: patterned stars! And Uri misses them a lot." The puppet is put back down. "Patterned stars are something special: their rays have little pictures in them, and these pictures look quite similar to one another because they have the same pattern. Every ray has a different pattern. Because Uri can't make a patterned star himself, maybe you can make one for him. Would you like to help Uri?"*
„Siehst Du diese Karten? Da sind unterschiedliche Muster drauf. Du kannst mir jetzt zwei Karten zeigen, die sich ähnlich sehen, weil sie ein ähnliches Muster haben, dann legen wir die zwei zusammen, und das ist der Anfang vom ersten Strahl.	*"Do you see these cards? There are different patterns on them. For a start, give me two cards that look similar to one another. We'll put those cards down beside one another, and that will be the start of the first ray."*
Die Kinder fangen an. Die beiden kombinierten Karten werden als Paar beiseitegelegt, und eine neue Karte (oder bei Bedarf auch mehrere neue Karten) füllen die Lücken. Immer wieder sollten - in Sätze eingebaut - die Hinweise kommen: Strahlen mit Bildern, die sich ähnlich sind ...	*The children begin the game. The first set of combined cards are set down as a pair, and a new card (or if necessary multiple new cards) fill the gaps. Again and again the experimenter should—built into appropriate sentences—repeat the following sorts of hints: "rays with pictures*

Bilder, die gleiche Muster haben ... ein schöner Musterstern mit langen Strahlen usw. Evtl. der Hinweis: „Es ist gar nicht wichtig, was für Dinge oder Sachen auf den Karten sind, Uri geht es nur um die Muster." Falls nur 2er-Paare gefunden werden wird darauf hingewiesen, dass die Strahlen auch aus mehr Karten bestehen können. Das Kind spielt so lange weiter, bis es nicht mehr möchte oder die Karten verbraucht sind.	*that look similar"…"pictures with the same pattern"…"a nice star should have long rays" and so forth.* *Eventually the experimenter may say "it's not at all important what the things in the pictures are—Uri is only interested in the pattern."* *In the case of the child only setting down 2-card groups/rays they will be reminded that they can put more cards onto the rays that are already there. The child plays the game until they no longer want to, or until all the cards are used up.*
„Oh, das sind schöne Muster! Uri, gefallen sie Dir? Freust Du Dich über diesen Musterstern?" Uri fliegt über die Karten und bedankt sich dann durch nicken.	*"Oh, what lovely patterns! Uri, do you like it? Does this star make you happy?"* *Uri flies over the cards and thanks the child by nodding.*
Adults	
„Wir möchten Sie bitten, Karten, die ich Ihnen gleich geben werde, mit Karten zusammen zu legen, denen sie visuell ähnlich erscheinen. Die Gründe, weshalb Sie etwas visuell ähnlich empfinden entscheiden Sie selbst. Dabei ist es nicht wichtig, welche Gegenstände auf den Karten abgebildet sind - die Karten, welche am ähnlichsten aussehen, kommen zusammen in eine Gruppe. Die Größe einer Gruppe ebenso wie die Anzahl der entstandenen Gruppen hängt davon ab, wie viele Karten Sie jeweils als ähnlich empfinden. Da gibt es keine Vorgaben. Haben Sie Fragen dazu?"	*"We ask you to lay down cards—which I will give to you shortly—with other cards on the basis of how visually similar they are. What it is for them to be "visually similar" is something you decide for yourself. It's not at all important what objects are depicted on the cards—if cards strike you as similar, you put them together in a group. The size of these groups and the number of those groups both depend on how similar you find the cards. There are no other requirements.* *Have you any other questions?"*
„Sie können mit Paaren von 2 ähnlichen Karten beginnen, und später noch andere passende Karten dazu legen, so dass die Gruppen größer werden." (Es sollten keine Fragen beantwortet werden, welche sich auf die Motive beziehen)	*"You can start with pairs of two similar cards, and later add other matching ones to make the group bigger."* *(The experimenter cannot answer questions relating to the images' identities)*

2.8.1 S1 Results Questionnaires

2.8.1.1 Adults' Questionnaire on their Criteria for Assembling Images.

After the finishing the sorting task, adult participants received a questionnaire asking about criteria underlying their similarity judgments. Adults were asked a) to provide terms which described image similarity within their image groups in general (question 1), and b) to choose two to three of their assembled image groups and describe image similarity within each of these groups (question 2a-c).

We coded the answers by noting if they included terms which belonged to one of four qualities, defined by the variables:

- Appearance–descriptions of pattern, shape, or grey tone.
- Entity–labels of depicted objects.
- Haptic–adjectives which describe experiences with the depicted objects.
- Paraphrases, which include labels of entities (e.g. "leave-like", "rock-pattern").

We then calculated the proportion in which each quality contributed to similarity decisions. This was done by dividing the total of cases in which the particular quality was mentioned by the total of all mentioned qualities. We only included answers to question 2 because some of the general criteria were difficult to understand, and because answers to question 1 included terms which were repeated in the examples of question 2. The results are shown in Table S 2.2.

Table S 2.2 Qualities included in adults' answers

Quality	Frequency	Proportion
Appearance	152	0.55
Entity	76	0.27
Haptic	17	0.06
Paraphrase	33	0.12

2.8.1.2 Children's Comments during Similarity Sorting

In order to assess children's criteria on what they perceived as similar, we had video-recorded children during the sorting task. If caregivers did not agree to video-recording, we had taken notes of children's spontaneous comments on the images during the sorting task. We then coded all comments by first separating them according to the context in which they occurred, and the intention we assumed behind the comment (e.g. describing similarity between images versus naming an object or describing an impression independent of similarity to another image). In the analysis of the comments, we only included those that referred to the similarity between images. Because the quality of the terms children used were more difficult to categorize than those of adults, we only assessed the qualities: a) appearance, and b) entity because they were of great interest for our analysis. Their frequencies and proportions are shown in Table S 2.3. The assessment of criteria determining similarity indicates that children and adults attended to visual appearance as well as the depicted entities during the sorting task.

Table S 2.3 Qualities included in children's comments

Quality	Frequency	Proportion
Appearance	53	.62
Entity	33	.38

2.8.1.3 Analysis of Questionnaires on Prior Exposure to the Categories

After the experiments, a questionnaire was given to the adult participants or the caregivers of the children which asked about frequencies of prior exposure to each of the categories: artifacts, natural elements, and vegetation. The questionnaires were developed for the present study. They included questions about exposure to the categories due to activities which the participant him- or herself performed, or exposure to activities a participant passively experienced—for example, when the partner of an adult or a child's parent was involved in the activities. We also assessed general exposure to pictures or picture books, and exposure to more abstract, computer-related activities such as text processing. These more general questions were expected to indicate visual exposure to two dimensional visual information, which differs from that of the naturalistic environment. For each of the questions, five possible frequencies could be chosen. These were: a) more than 4 times a week; b) 1 - 4 times a week; c) 1 - 4 times a month; d) less than once a month; c) never.

We then averaged the frequencies over active and passive exposures and correlated the averages with performance data of the classification task (i.e., the sensitivity measure dprime) separately for each of the categories depicted in the images. No significant correlations were found between exposure frequencies and sensitivity for a particular category after adjusting *p*-values (Benjamini & Hochberg, 1995). However, when correlating participants overall sensitivity values with frequencies of general exposure to pictures or picture books, we found that more frequent exposure to activities including pictures and picture books led to lower sensitivity for the categories depicted in the study's images in children (Spearman’s $r(223) = -.19$, $p = .02$), but not in adults. Exposure to abstract, computer-related activities was not significantly related to sensitivity in children or adults. One possible explanation for this finding could be that children who spend much time with children's books learn graphical versions of entities, which do not include visual information as it is useful for the perception of photographs. Additionally, the more time a child spends with picture books and pictures, the less outdoor activities this child is exposed to. In contrast to what one might expect, frequent visual exposure to pictures did therefore not lead to an increase in the ability to perceive two-dimensional visual information. Future studies could compare the effect of frequent exposure to graphics designed for children with the effect of frequent outdoor activities on developing perceptual abilities. Moreover, this finding questions the validity of graphical representations of natural entities in categorization studies conducted with young children.

2.8.2 S2 Results Materials.

2.8.2.1 Prediction of the True Categories in our Image Set

In order to determine visual properties which were additionally included in participants' decisions during classification although they did *not* predict categories in our images, or which were *not* included by participants although they *did* predict category membership in our image data, we assessed which visual properties statistically predicted the category membership in our image set.

For each of the three categories, separate GLMs were conducted (R-function glm, R Core Team, 2019) on the visual properties of the 60 images used in our study. The binary dependent variables (DV) indicated if an image depicted the respective category or not (1, 0). We assessed the significance of visual properties by including each visual property individually in a model, resulting in 10 tests (see Section 2.4 Results, Statistical Analysis in the main text). and adjusted *p*-values with the method Benjamini and Hochberg (1995). Figure S 2.**1** shows the distribution of visual properties as a function of the categories depicted in the images. Significant main effects are indicated by asterisks. Coefficients for all visual properties are provided in Table S 2.4.

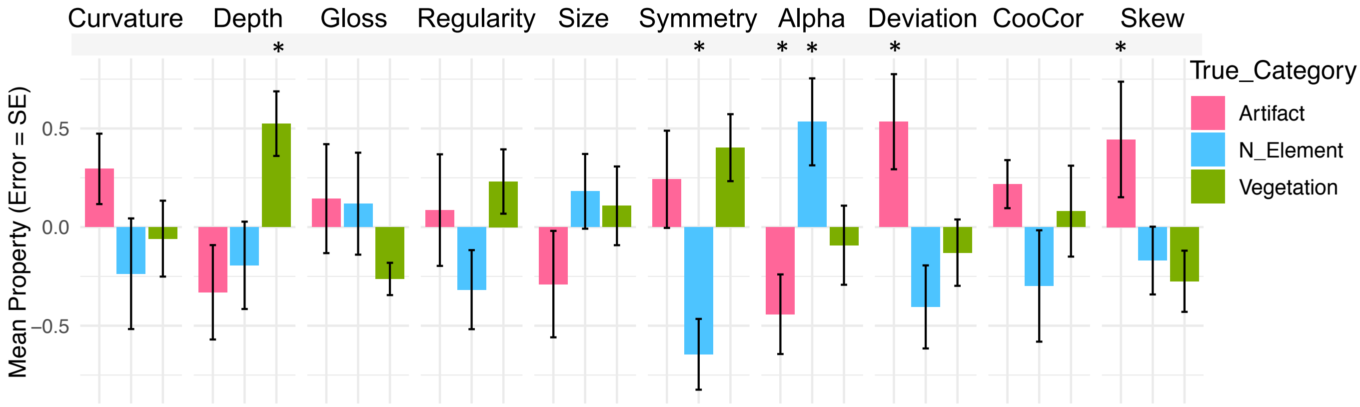


Figure S 2.1**: Visual properties as function of the images' categories.**
Visual properties are z-standardized and averaged over true categories.
* adjusted *p* < .05, assessed in the GLM, Table S 2.4.

Table S 2.4 Category membership of the images predicted by visual properties

	True categories								
	Artifacts			Natural Elements			Vegetation		
Property	Log-odds	*CI* low	High a	Log-odds	*CI* low	High a	Log-odds	*CI* low	High a
Curvature	0.46	-0.10	1.02	-0.37	-0.93	.19	-0.09	-0.63	0.45
Depth	-0.51	-1.07	0.05	-0.29	-0.84	.25	0.99*	0.27	1.70
Gloss	0.21	-0.32	0.74	0.18	-0.35	.71	-0.46	-1.09	0.18
Regularity	0.13	-0.41	0.67	-0.52	-1.11	.07	0.35	-0.20	0.90
Size	-0.44	-1.00	0.11	0.29	-0.28	.86	0.17	-0.39	0.72
Symmetry	0.38	-0.18	0.95	-1.13**	-1.81	-.45	0.67	0.06	1.27
Alpha	-0.78*	-1.44	-0.12	0.93*	0.26	1.59	-0.14	-0.69	0.41
Deviation	0.87*	0.25	1.49	-0.72	-1.39	-.06	-0.20	-0.76	0.36
Coocor	0.40	-0.27	1.06	-0.44	-1.00	.11	0.13	-0.44	0.70
Skew	0.70*	0.10	1.29	-0.28	-0.86	.31	-0.48	-1.11	0.16

* adjusted p < .05; ** adjusted p <.01 (method: Benjamini and Hochberg, 1995)
[a] Confidence intervals with low = 2.5%, high = 97.5%.

2.8.2.2 Correlations between Visual Properties in the Image Set

The visual properties we chose for the current study statistically relate to each other, indicated by correlations between some visual properties in our image set. We still decided not to agglomerate the correlating properties because a) they were chosen for theoretically distinct reasons, and b) even the members of highly correlating property pairs (i.e., alpha-CooCor, or regularity-symmetry) were found to be included very differently in the categorization decisions of the participant groups and led to distinct significance patterns between the categories. This important information would have been obscured by including for example the principle component of property pairs (see main Result section for further discussion on this decision). We present the correlation matrix of the visual properties in Figure S 2.**2**.

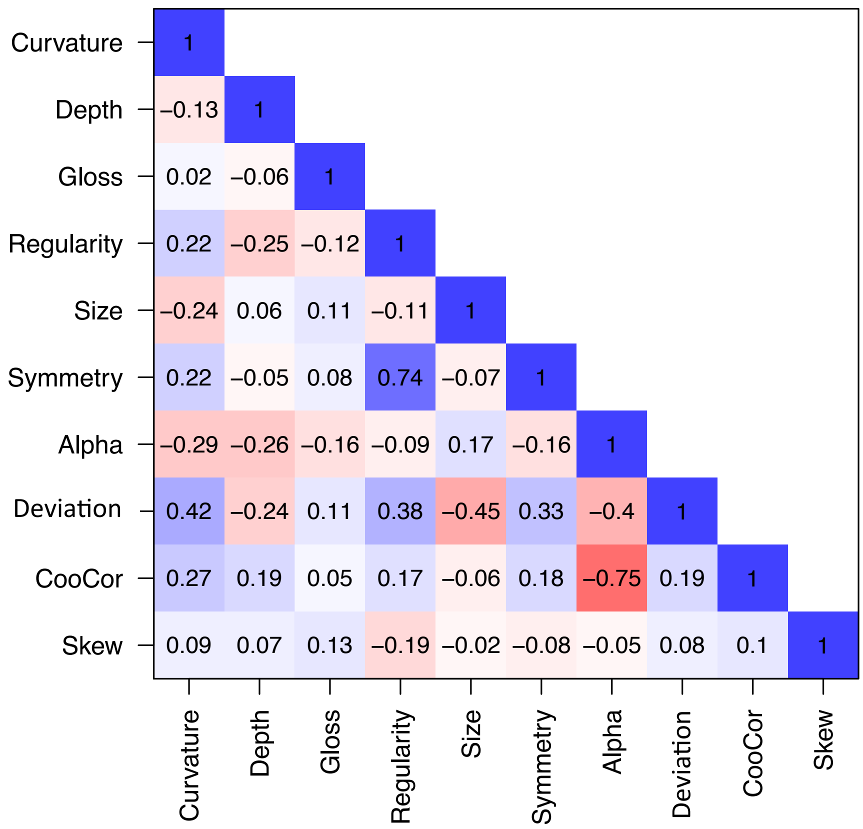


Figure S 2.2: **Correlation matrix of visual properties included in our image set.** Numbers are Pearson correlation coefficients including the data of 60 images.

2.8.3 S3 Results Classification Task

The following tables include all results of the GLMMs conducted individually on visual properties and assigned categories.

Table S 2.5 Visual Properties Predicting Assigned Categories in Children

	Assigned categories children								
	Artifacts			Natural elements			Vegetation		
Property	Log-Odds	*CI* low	high [a]	Log-Odds	*CI* low	high [a]	Log-Odds	*CI* low	high [a]
Curvature	.38	-.11	.87	-.08	-.48	.32	-.24	-.85	.37
Depth	**-.69***	-1.16	-.22	-.44	-.83	-.05	**1.15****	.60	1.69
Gloss	.27	-.22	.76	.01	-.39	.41	-.27	-.88	.35
Regularity	.05	-.45	.55	-.21	-.60	.19	.22	-.39	.83
Size	-.51	-1.0	-.03	.16	-.25	.56	.32	-.29	.93
Symmetry	.16	-.34	.65	**-.65****	-1.01	-.28	.44	-.16	1.05
Alpha	-.32	-.81	.17	.49	.11	.87	-.06	-.67	.54
Deviation	**.80****	.35	1.25	-.18	-.58	.22	-.64	-1.24	-.05
CooCor	.06	-.44	.56	-.32	-.71	.07	.13	-.47	.74
Skew	**.71***	.25	1.17	-.11	-.52	.29	-.65	-1.25	-.05

* adjusted p < .05. ** adjusted p <.01, (method: Benjamini and Hochberg, 1995).
[a] Confidence intervals with low = 2.5%, high = 97.5%.

Table S 2.6 Visual Properties Predicting Assigned Categories in Adults

	Assigned categories adults								
	Artifacts			Natural elements			Vegetation		
Property	Log-Odds	*CI* low	high [a]	Log-Odds	*CI* low	high [a]	Log-Odds	*CI* low	high [a]
Curvature	**3.0***	.85	5.14	-.33	-1.73	1.08	-1.64	-3.85	.58
Depth	-.68	-2.99	1.63	-1.37	-2.81	.07	**2.93****	1.17	4.70
Gloss	1.33	-2.32	4.99	.42	-.97	1.81	-.59	-2.56	1.38
Regularity	.76	-1.81	3.34	-.83	-2.28	.62	.26	-1.72	2.23
Size	**-3.05***	-5.26	-.85	.34	-1.06	1.74	1.04	1.04	1.04
Symmetry	2.28	-.64	5.19	**-2.81****	-4.16	-1.46	1-06	-.83	2.95
Alpha	**-3.04***	-5.12	-.97	**1.89***	.61	3.18	-.30	-2.20	1.60
Deviation	**3.72****	1.60	5.84	-.99	-2.37	.38	**-2.21***	-4.08	-.34
CooCor	2.08	-.34	4.49	-1.26	-2.59	.07	.31	-1.43	2.06
Skew	**4.54***	1.48	7.60	-.97	-2.47	.54	**-2.66****	-4.34	-.99

* adjusted p < .05. ** adjusted p <.01, (method: Benjamini and Hochberg, 1995)
[a] Confidence intervals with low = 2.5%, high = 97.5%.

Table S 2.7 Visual Properties, Participant Groups, and Their Interaction Predicting Assigned Categories

	Assigned categories group interaction					
	Artifacts		Natural elements		Vegetation	
Property	Log-Odds	*CI* low high [a]	Log-Odds	*CI* low high [a]	Log-Odds	*CI* low high [a]
Group [c] (Curvature)	-.01	-.26 .24	-.19	-.45 .06	.24	-.07 .56
Curvature	.57	-.10 1.24	-.23	-.79 .33	-.23	-.99 .52
Curvature × Group [c]	-.18	-.37 .02	**.17***	.01 .34	-.05	-.25 .14
Group [c] (Depth)	-.06	-.31 .19	-.19	-.44 .07	.29	-.03 .60
Depth	-.76	-1.42 -.11	**-.66***	-1.20 -.11	**1.46****	.78 2.14
Depth × Group [c]	-.08	-.28 .11	**.18***	.00 .36	-.23	-.45 -.01
Group [c] (Gloss)	-.06	-.30 .19	-.16	-.42 .10	.24	-.08 .55
Gloss	.25	-.42 .92	.22	-.34 .78	-.52	-1.28 .24
Gloss × Group [c]	.11	-.07 .29	**-.25***	-.42 -.09	**.30***	.09 .52
Group [c] (Regularity)	-.03	-.28 .22	-.17	-.43 .09	.25	-.07 .56
Regularity	.18	-.50 .86	-.54	-1.10 .01	.50	-.26 1.25
Regularity × Group [c]	-.18	-.35 -.00	**.38****	.21 .55	**-.32****	-.52 -.12
Group [c] (Size)	-.04	-.29 .21	-.21	-.46 .05	.27	-.05 .59
Size	-.66	-1.33 -.00	.26	-.30 .83	.21	-.54 .97
Size × Group [c]	.05	-.14 .24	**-.18***	-.36 -.00	.21	.00 .42
Group [c] (Symmetry)	-.02	-.27 .23	-.11	-.37 .15	.22	-.09 .54
Symmetry	.54	-.14 1.22	**-1.40****	-1.92 -.89	.85	.11 1.60
Symmetry × Group [c]	**-.50****	-.69 -.31	**.76****	.58 .95	**-.48****	-.69 -.27
Group [c] (Alpha)	-.00	-.25 .25	-.14	-.40 .12	.24	-.07 .56
Alpha	**-1.02***	-1.70 -.34	**1.19****	.65 1.73	-.18	-.93 .57
Alpha × Group [c]	**.81****	.60 1.02	**-.76****	-.95 -.57	.15	-.06 .35
Group [c] (Deviation)	.03	-.22 .28	-.20	-.45 .06	.23	-.09 .54
Deviation	**1.19****	.57 1.81	-.60	-1.15 -.04	-.46	-1.20 .28
Deviation × Group [c]	**-.30****	-.48 -.12	**.50****	.33 .67	**-.39****	-.60 -.18
Group [c] (CooCor)	-.01	-.26 .24	-.18	-.43 .08	.25	-.07 .56
CooCor	.52	-.17 1.21	**-.70***	-1.25 -.15	.19	-.56 .93
CooCor × Group [c]	**-.57****	-.80 -.34	**.41****	.23 .58	-.04	-.22 .15
Group [c] (Skew)	.01	-.24 .25	-.19	-.45 .07	.26	-.06 .58
Skew	**1.23****	.60 1.87	-.39	-.95 .17	-.95	-1.70 -.21
Skew × Group [c]	**-.45****	-.63 -.26	**.33****	.14 .52	.23	.00 .45

* adjusted p < .05. ** adjusted p <.01, (method: Benjamini and Hochberg, 1995).
[a] Confidence intervals with low = 2.5%, high = 97.5%.

2.8.4 S4 Results Sorting Task

2.8.4.1 The Course of Visual Property Variance during the Cluster Agglomeration Process

HCA Children. Figure 2.6 (top row) of the main result section shows the proportion of variance explained by the visual properties as a function of the height of the dendrogram. The top left of Figure 2.6 illustrates the development of explained variance for children. At the origin of the children's x-axis, each of the 60 images belonged to an individual cluster, resulting in values of $R^2 = 1$ (in the centered R^2 values used in the figure the information of overall explained variance is not included anymore, so R^2 values are zero). With increasing height, the agglomeration of the images to clusters reduced R^2 values. According to visual inspection, explained variance of the individual visual properties first developed homogenously, but became more distinct at height .75 when images were merged to 30 clusters. At height .9 (23 Clusters), which includes sizes of clusters approximately corresponding to the actual mean of sorted image group sizes, visual properties reached a point of high variability. Here, Regularity, Depth and Skew had the highest impact on children's similarity perception with values of $R^2 > .61$, while Alpha had the lowest impact with $R^2 = .33$. This changed at the second region of high variability between height 1.05 and 1.25 (14 to 9 clusters). While Depth had remained the property with the highest impact with $R^2 > .6$, Deviation and CooCor (both $R^2 = .36$) elevated above Regularity and Skew. The impact of Gloss decreased below all properties to $R^2 = .1$. Depth, Deviation and CooCor shared their high impact until height 1.6 (3 clusters), from where on variability of the properties diminished.

HCA Adults. The impact of visual properties on adults' sorting decisions is depicted in Figure 2.6 top right. At zero height, the visual property values already varied due to the above described instantaneous agglomeration, leading to R^2 values between 1 and .92. They then continued in a very similar order, arriving at a first region of high variability between heights .65 and .75 (21 to 19 clusters). Here, regularity had the highest impact with $R^2 > .72$, whereas the lowest value belonged to gloss ($R^2 < .26$). Within the next region of high variability (height .9 to 1.1, 16 to 12 clusters), the visual properties separated into a high and a low impact group. In the group with the impact, regularity, symmetry and area similarly explained the most variance ($R^2 > .55$), while in the group with lower impact gloss remained at the lowest level with $R^2 < .22$. Beyond this region, area elevated above the other properties, replaced by depth at height 1.9 (3 clusters), which remained at the elevated position until maximum height.

2.8.4.2 *R^2 values of the Agglomeration Process HCA*

This data is uploaded to:
https://osf.io/8xy5n/?view_only=6ddced286c31456fae7d20dd86e072e6
file: Explained_Similarity_HCA_sorttask_public.txt

2.8.4.3 Contrasts of the ANOVAs on the HCA R^2 values:

Table S 2.8 Tukey's HSD Contrasts of R^2 Visual Properties, Children's HCA

Contrast	Difference	*CI* low	*CI* high [a]	adjusted *p*
Deviation-Alpha	.054	.027	.08	<.001
CooCor-Alpha	.072	.046	.099	<.001
Curve-Alpha	.011	-.015	.038	.944
Depth-Alpha	.149	.122	.175	<.001
Gloss-Alpha	-.009	-.036	.017	.986
Regularity-Alpha	.106	.079	.132	<.001
Size-Alpha	.046	.02	.073	<.001
Skew-Alpha	.113	.087	.14	<.001
Symmetry-Alpha	.046	.019	.072	<.001
CooCor-Deviation	.018	-.008	.045	.469
Curve-Deviation	-.043	-.069	-.016	<.001
Depth-Deviation	.095	.068	.121	<.001
Gloss-Deviation	-.063	-.089	-.036	<.001
Regularity-Deviation	.052	.026	.079	<.001
Size-Deviation	-.008	-.034	.019	.996
Skew-Deviation	.06	.033	.086	<.001
Symmetry-Deviation	-.008	-.034	.019	.995
Curve-CooCor	-.061	-.087	-.034	<.001
Depth-CooCor	.077	.05	.103	<.001
Gloss-CooCor	-.081	-.108	-.055	<.001
Regularity-CooCor	.034	.007	.06	.002
Size-CooCor	-.026	-.052	.001	.063
Skew-CooCor	.041	.015	.068	<.001
Symmetry-CooCor	-.026	-.053	0	.059
Depth-Curve	.137	.111	.164	<.001
Gloss-Curve	-.02	-.047	.006	.311
Regularity-Curve	.095	.068	.121	<.001
Size-Curve	.035	.009	.061	.001
Skew-Curve	.102	.076	.129	<.001

(continues: Table S 2.8)

Contrast	**Difference**	***CI* low**	***CI* high** [a]	**adjusted *p***
Symmetry-Curve	.035	.008	.061	.001
Gloss-Depth	-.158	-.184	-.131	<.001
Regularity-Depth	-.043	-.069	-.016	<.001
Size-Depth	-.102	-.129	-.076	<.001
Skew-Depth	-.035	-.062	-.009	.001
Symmetry-Depth	-.103	-.129	-.076	<.001
Regularity-Gloss	.115	.088	.141	<.001
Size-Gloss	.055	.029	.082	<.001
Skew-Gloss	.122	.096	.149	<.001
Symmetry-Gloss	.055	.029	.082	<.001
Size-Regularity	-.06	-.086	-.033	<.001
Skew-Regularity	.007	-.019	.034	.996
Symmetry-Regularity	-.06	-.086	-.033	<.001
Skew-Size	.067	.041	.094	<.001
Symmetry-Size	0	-.027	.026	1
Symmetry-Skew	-.067	-.094	-.041	<.001

Table S 2.9 Tukey's HSD Contrasts of R^2 Assigned Categories, Children's HCA

Contrast	**Difference**	***CI* low**	***CI* high** [a]	**adjusted *p***
N_Elements-Artifacts	0	-.026	.025	.999
Vegetation-Artifacts	.066	.041	.091	<.001
Vegetation-N_Elements	.066	.041	.092	<.001

Table S 2.10 Tukey's HSD Contrasts of R^2 Visual Properties, Adults' HCA

Contrast	Difference	*CI* low	*CI* high [a]	adjusted *p*
Deviation-Alpha	.131	.097	.164	<.001
CooCor-Alpha	-.012	-.045	.022	.981
Curve-Alpha	.12	.087	.153	<.001
Depth-Alpha	.113	.08	.147	<.001
Gloss-Alpha	-.078	-.111	-.045	<.001
Regularity-Alpha	.174	.141	.207	<.001
Size-Alpha	.1	.067	.134	<.001
Skew-Alpha	.068	.034	.101	<.001
Symmetry-Alpha	.157	.124	.191	<.001
CooCor-Deviation	-.143	-.176	-.109	<.001
Curve-Deviation	-.011	-.044	.023	.99
Depth-Deviation	-.018	-.051	.016	.802
Gloss-Deviation	-.209	-.242	-.176	<.001
Regularity-Deviation	.043	.01	.077	.002
Size-Deviation	-.03	-.064	.003	.109
Skew-Deviation	-.063	-.097	-.03	<.001
Symmetry-Deviation	.026	-.007	.06	.27
Curve-CooCor	.132	.099	.165	<.001
Depth-CooCor	.125	.092	.158	<.001
Gloss-CooCor	-.066	-.1	-.033	<.001
Regularity-CooCor	.186	.153	.219	<.001
Size-CooCor	.112	.079	.146	<.001
Skew-CooCor	.08	.046	.113	<.001
Symmetry-CooCor	.169	.136	.202	<.001
Depth-Curve	-.007	-.04	.026	1
Gloss-Curve	-.198	-.232	-.165	<.001
Regularity-Curve	.054	.021	.087	<.001
Size-Curve	-.02	-.053	.014	.692
Skew-Curve	-.052	-.086	-.019	<.001
Symmetry-Curve	.037	.004	.071	.016
Gloss-Depth	-.191	-.225	-.158	<.001
Regularity-Depth	.061	.027	.094	<.001
Size-Depth	-.013	-.046	.021	.971

(*continuous: Table S 2.10*)

Contrast	**Difference**	***CI* low**	***CI* high [a]**	**adjusted *p***
Skew-Depth	-.045	-.079	-.012	.001
Symmetry-Depth	.044	.011	.077	.001
Regularity-Gloss	.252	.219	.286	<.001
Size-Gloss	.179	.145	.212	<.001
Skew-Gloss	.146	.112	.179	<.001
Symmetry-Gloss	.235	.202	.269	<.001
Size-Regularity	-.074	-.107	-.04	<.001
Skew-Regularity	-.106	-.14	-.073	<.001
Symmetry-Regularity	-.017	-.05	.017	.847
Skew-Size	-.033	-.066	.001	.062
Symmetry-Size	.057	.023	.09	<.001
Symmetry-Skew	.089	.056	.123	<.001

Table S 2.11 Tukey's HSD Contrasts of R^2 Assigned Categories, Adults' HCA

Contrast	**Difference**	***CI* low**	***CI* high [a]**	**adjusted *p***
N_Element-Artifact	.031	.008	.059	.006
Vegetation-Artifact	.081	.057	.105	<.001
Vegetation- N_Element	.049	.025	.073	<.001

Table S 2.12 Tukey's HSD Contrasts of R^2 Children and Adults, Visual Properties

Contrast	**Difference**	***CI* low**	***CI* high [a]**	**adjusted *p***
Children-Adults	-.045	-.051	-.039	<.001
Alpha:Children-Alpha:Adults	-.027	-.061	.007	.372
Deviation:Children-Deviation:Adults	-.104	-.138	-.07	<.001
CooCor:Children-CooCor:Adults	.057	.023	.091	<.001
Curve:Children-Curve:Adults	-.136	-.17	-.101	<.001
Depth:Children-Depth:Adults	.009	-.026	.043	1
Gloss:Children-Gloss:Adults	.042	.008	.076	.002
Regularity:Children-Regularity:Adults	-.095	-.129	-.061	<.001
Size:Children-Size:Adults	-.081	-.115	-.047	<.001
Skew:Children-Skew:Adults	.019	-.015	.053	.921
Symmetry:Children-Symmetry:Adults	-.138	-.172	-.104	0

Table S 2.13 Tukey's HSD Contrasts of R^2 Children and Adults, Assigned Categories

Contrast	Difference	*CI* low	*CI* high [a]	adjusted *p*
N_Element -Artifact	-.016	-.001	.032	.072
Vegetation-Artifact	-.074	-.057	-.090	<.001
Vegetation- N_Element	-.058	-.041	-.075	<.001

[a] Confidence intervals of table 7–13 with low = 2.5%, high = 97.5%

2.8.4.4 Can Variance of Visual Properties be Separated from Variance of Assigned Categories in the Sorting Task?

We assessed and presented R^2 values of visual properties and of assigned categories– both are indicating their impact on the participant's similarity decisions (Figure 2.**5**). One can argue that assigned categories and visual properties are not independent of each other, and that it is not clear whether participants attended to the category of an image, or the visual properties which are predicting the category. If categories were primarily attended to, then the R^2 values of the visual properties should develop in patterns which are congruent to those provided by the R^2 values of the categories. We therefore evaluated the relationship between assigned categories and visual properties by visual inspection and did not find a clear relationship. For example, children predominantly relied on Depth, Skew and Deviation in their assignment of artifacts and vegetation. In the sorting task, these properties had elevated impact on similarity perception in accordance with higher values of assigned vegetation and artifacts above natural elements (height 1.1 to 1.2). This gives the impression that visual properties illustrate the impact of assigned categories. However, around height 1.4 (6 clusters), when vegetation is elevated high above the other properties, predictors of assigned vegetation only play a secondary role, while Alpha–which predicted natural elements– increased its impact. Concerning the general sequence, we found that Symmetry which was predicting children's assignment to natural elements and vegetation only played a minor role in their similarity decisions. In contrast, CooCor which was *not* found to predict category assignment, explained a moderate to high proportion of variance during the children's clustering hierarchy, compared to the other properties. This inspection gives the impression that assigned categories and visual properties do not play an exclusive role for similarity judgments, but were attended in parallel.

Visual inspection of the relationship between assigned categories and visual properties in the adult sorting task did not reveal a clear overlap. Recall that depth, skew and area had been found to predict adults' assignment to vegetation. Between height 1.7 and 2.6 (5 to 2 clusters), depth and area were elevated in parallel to assigned vegetation. However, skew generally had a minor impact on similarity perception in adults. Moreover, regularity, which did not reach significance in the adults' classification task, was one of the properties with the highest impact in the sorting task (Figure 2.8). As with children, these examples show that the impact of particular visual properties cannot be fully attributed to participants' inclusion of assigned categories.

Nevertheless, we cannot exclude that relationships exist which cannot be observed in this way. Alternative explanations of the partial overlap are discussed in the discussion part of the main text.

2.8.5 S5 Additional Analysis: The Effect of Children's Age on the Inclusion of Visual Properties during Classification

During the analysis of the classification task, we compared which visual properties predicted the assignment of categories in children and adults. We had included the covariate age in the analysis of the children's classification task, because of its significant impact on a child's general performance in this task (i.e., Spearman correlation of correctly identified images; $r(226) = .21$, $p < .001$). The findings indicated that some visual properties which relied on detailed visual information were not included in an adult-like way by children. To receive better understanding of this finding, we decided to analyze the relationship between the inclusion of a visual property and the age of a child. This analysis could show, if younger children were including less visually-detailed information in their decisions than older children and support our interpretation. We ran additional GLMMs including the continuous variables age, visual property, and the interaction term Age × Visual Property. We ran separate models for each of the assigned categories and each particular visual property (further information about procedures and software are provided in the main result section).

After adjusting *p*-values (Benjamini & Hochberg, 1995), we found significant interactions between age and visual property for the visual properties regularity, size, and symmetry on both of the assigned categories artifacts and natural elements, and, moreover, for the visual properties curvature, area and CooCor on artifacts (all $p < .05$). No properties led to an interaction with age on assigned vegetation. The directions of the effects are shown in Figure S 2.3.

These results indicate that with regard to images of artifacts and natural elements, preschool children's inclusion of some visual properties drawn upon during categorization changed with age. Moreover, depth and skew, which were the strongest predictors of children's similarity judgments and did not differ between adults and children in the sorting task also do not show differences between younger and older children during classification.

Visual properties included in the assignment of images to vegetation did not differ between younger and older children. This may be explained by the particular importance of vegetation (see e.g., Wertz, 2019). An additional explanation could be that visual properties–except depth–were only included randomly in children's assignment to the vegetation category and therefore showed no interactions with age. In contrast, depth, which was the mayor predictor of vegetation, had the same impact on all children's classification decisions.

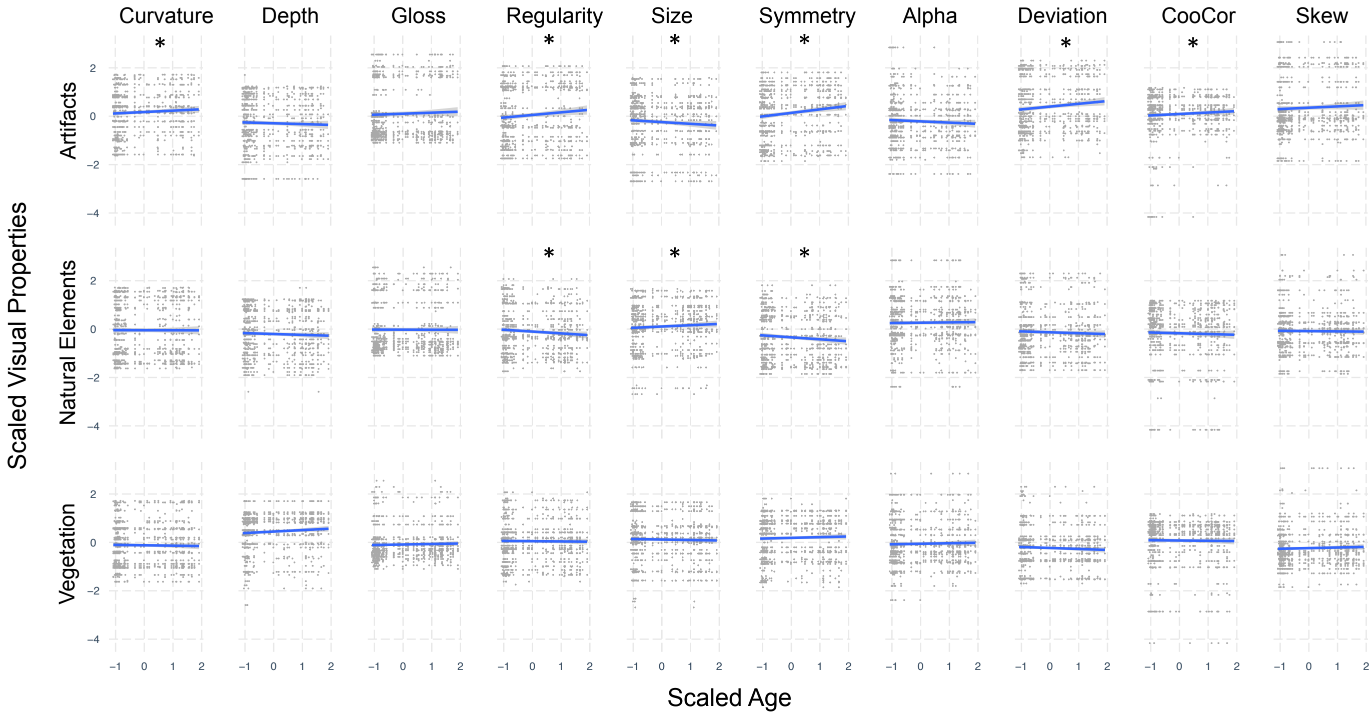


Figure S 2.3: Visual Properties as Function of Children's Age and Assigned Category.
Note. Error bars are *SE*.
***** adjusted $p < .05$ (method: Benjamini and Hochberg, 1995).

Chapter 3

The Effects of Calibration Target, Screen Location, and Movement Type on Infant Eye-tracking Data Quality [2]

(Study 2)

Karola Schlegelmilch and Annie E. Wertz

Max Planck Institute for Human Development,

Max Planck Research Group Naturalistic Social Cognition, Berlin, Germany.

---

[2] This is the accepted version of the following article:
Schlegelmilch, K., & Wertz, A. E. (2019). The effects of calibration target, screen location, and movement type on infant eye-tracking data quality. *Infancy*, *24*(4), 636–662, which has been published in final form at [https://onlinelibrary.wiley.com/doi/abs/10.1111/infa.12294].

3.1 Abstract

During infant eye-tracking, fussiness caused by the repetition of calibration stimuli and body movements during testing are frequent constraints on measurement quality. Here, we systematically investigated these constraints with infants and adults using EyeLink 1000 Plus. We compared looking time and dispersion of gaze points elicited by stimuli resembling commonly used calibration animations. The adult group additionally performed body movements during gaze recording that were equivalent to movements infants spontaneously produce during testing. In our results, infants' preference for a particular calibration target did not predict data quality elicited by that stimulus, but targets exhibiting the strongest contrasts in their center or targets with globally distributed complexity resulted in the highest accuracy. Our gaze measures from the adult movement tasks were differentially affected by the type of movement as well as the location where the target appeared on the screen. These heterogeneous effects of movement on measures should be taken into account when planning infant eye-tracking experiments. Additionally, to improve data quality, infants' tolerance for repeated calibrations can be facilitated by alternating between precise calibration targets.

3.2 Introduction

Many insights into infant development are based on the study of gaze behavior. Eye-tracking technology allows an increasingly more detailed analysis of infant gaze behavior and is used to investigate a wide range of phenomena, such as categorization, object and face perception, and social cognition (for reviews see e.g., Aslin, 2007; Gredebäck, Johnson, & von Hofsten, 2009; Oakes, 2012). While the availability of high temporal and spatial measuring resolution expands the possible experimental designs and dependent measures, typical problems that might occur during infant eye-tracking can markedly effect data quality. Therefore, researchers must remain cautious to avoid overestimating its measurement accuracy (Aslin, 2012) and continue to address the inherent challenges of infant eye-tracking (Oakes, 2012).

The major challenges are body movements or inadequate looking behavior during calibration and during the later stages of the experiment. Haith (2004) estimated that an average of 50% of infants recruited for eye-tracking studies did not provide usable data as a result of such failures. In cases where individual infants are not fully excluded from the datasets, rejected trials of otherwise acceptable individual performance increase the proportions of unusable data (for procedures to reduce data loss in post hoc data optimization, see Leppänen, Forssman, Kaatiala, Yrttiaho, & Wass, 2015, for Tobii systems; Renswoude et al., 2018, for EyeLink technology).

A comparison of data quality in infant eye-tracking based on exclusion rates alone is difficult because exclusion criteria are adjusted according to the sensitivity of the phenomena under investigation. For example, psychophysical investigations that are sensitive to stability of gaze might be particularly prone to confounds related to differences in body movement, making more conservative exclusion boundary values necessary (e.g., an average calibration error of <1° or a data yield >80%; Alahyane et al., 2016). Infant studies that include data from adult participants often also employ more conservative exclusion boundaries to facilitate comparisons across differentially behaving participant groups (e.g., a data yield >80%; Morgante, Zolfaghari & Johnson, 2012). Similarly, studies that assess infants' attention to the details of an image depend on high spatial accuracy to produce interpretable results (e.g., Constantino et al., 2017). In contrast, studies that assess attention to larger visual targets that are clearly separated in the visual field can achieve valid data in spite of higher calibration errors or lower proportions of recorded gaze (e.g., Kulke, Atkinson, & Braddick, 2015;

LoBue, Buss, Taber-Thomas, & Pérez-Edgar, 2017). Despite the diverse demands of different experimental paradigms on data resolution, all approaches to infant eye-tracking would benefit from the following: (1) infant participants who are more attentive throughout the experimental session, and (2) enhanced measurement accuracy.

The present study therefore targets the most common pitfalls of infant eye-tracking: the calibration procedure and body movement during remote mode recording. We compared several animated calibration targets for their attractiveness to infants and their ability to direct infants' gaze to their centers. Enhancing the calibration stimuli and procedures used during this essential part of data collection will lead to more reliable recordings. In addition, we systematically investigated the ways in which body and head movements affect the accuracy of gaze recordings. More knowledge about the impact of these factors can help elucidate the best steps to take during and after data recording and adapt experimental procedures accordingly.

3.2.1 Infant Calibration Targets

The accuracy of infant eye-tracking data relies to a large extent on calibration quality (Gredebäck et al., 2009; Oakes, 2012). In standard adult calibration procedures, adults are explicitly instructed to fixate 5 to 13 point-like visual targets as precisely as possible. Infants of course cannot be instructed in this way. Instead, infants' spontaneous attention needs to be captured and held by animated calibration targets. Further, infants are commonly expected to perform calibrations with only 5 to 6 targets because of their limited attention span (Gredebäck et al., 2009). Inattentiveness of an infant during calibration makes repetitions of this procedure necessary, which can lead to annoyance and further inattentiveness. Animations that facilitate infants' attention and result in bundled fixations during calibration should therefore produce more reliable data. Indeed, the design of calibration targets has an impact on fixation stability even for adults, who voluntarily try to keep their gaze still (Thaler, Schütz, Goodale, & Gegenfurtner, 2013).

Determining which features facilitate calibration in infancy is a difficult task. Visual acuity relating to spatial frequency and contrast are not yet as developed in infancy as in adulthood, making less detailed stimuli easier for infants to process. However, patterns that are easy for infants to perceive can become boring when presented too frequently. A family of commonly applied calibration targets therefore consists of looming concentric spheres or rings, which are expected to provoke central fixations. Because concentric forms are not processed in an adult-like way until adolescence (Doucet, Gosselin, Lassonde, Guillemot, &

Lepore, 2005), it is not yet clear how this processing difficulty interacts with infants' attention, especially if the target is additionally flashed, moved, or its contour density is intensified to increase salience (Aslin & Smith, 1988; Zihl & Dutton, 2015). There is reason to suspect that the combination of these features may be problematic because visual patterns that are too stimulating can cause the infant to turn away (Bornstein & Benasich, 1986). Nevertheless, calibration targets must have features that make them sufficiently noticeable when appearing at unexpected locations on the screen because the area covered by the visual field is still increasing during infancy.

Inter-individual variability in the development of the fundamental issues we have raised makes it difficult to rely on theoretical assumptions alone when predicting the impact of calibration targets on infants' gaze behavior. Therefore, a systematic experimental investigation of the applicability and impact on data quality of calibration targets with different features is necessary.

3.2.2 Infant Eye-tracking Accuracy

Several factors that generally lead to a reduction of data quality during eye-tracking are present in infant eye-tracking experiments: movement, sitting position, geometry of the set-up, and the operators' experience with calibration procedures (for an extended discussion of these factors see Holmqvist, Nyström, & Mulvey, 2012). Movement during the recording sequence is particularly challenging because it causes changes in the geometry on which the calibration was based. In addition, the pupils might become partially covered, or move out of the area observable by the eye-tracker's camera, resulting in less robust data recording. Common dependent variables like the number of fixations or response time latencies are systematically influenced by interruptions of contact to the eye-tracking camera (Wass, Smith, & Johnson, 2013).

The circumstances of the infant eye-tracking situation make a more tolerant procedure necessary. Infants sitting on the lap of their caregiver can be expected to move in all spatial dimensions, even if they are interested in the experiment. Although some laboratories successfully use infant seats in eye-tracking studies for certain age groups (e.g., Saez de Urabain, Nuthmann, Johnson, & Smith, 2017), constrictions of movement can be uncomfortable and distracting for infants. Therefore, researchers must account for deviations from a stable position during infant testing. Remote mode eye tracking comes with a moderate spatial tolerance to account for such instability. Some systems also provide the ability to do drift checks to assess whether the measured gaze points have shifted during trial sequences (e.g., EyeLink 1000 Plus). If the reported fixation error is too large, a recalibration procedure should be implemented. A single drift check measurement might not be sufficient if the moment to accept the fixation was poorly chosen or if the infant's saccade towards the validation target was not precise. If the indicated gaze positions on the eye-tracking monitor or on a visual data output give the impression that fixations are systematically displaced, some eye-tracking software offers the possibility to adjust them later during analysis by carefully shifting them to their assumed correct locations (e.g., EyeLink Data Viewer User's Manual, 2002-2015), and researchers have developed procedures for post hoc corrections as well (e.g. Frank, Vul, & Saxe, 2012).

The success of all these factors—the tolerance of the eye-tracking device, drift checks, or subsequent corrections—depend on understanding the effects of movement on the data. The algorithms of the eye-tracker that correct head movements in remote mode might not function properly if participants move too much (Hessels, Cornelissen, Kemner, & Hooge, 2015; Niehorster, Cornelissen, Holmqvist, Hooge, & Hessels, 2017). Additionally, movement might result in blurred camera images leading to noise and a different variance of gaze points (Holmquist et al., 2012; Wass et al., 2014) and changes in the angle of the participant's head in relation to light sources might affect accuracy (Wass, Smith, & Johnson, 2013). Previous investigations of infant eye-tracking described reduced precision as a function of trial number (Hessels, Andersson, Hooge, Nyström, & Kemner, 2015), and high unpredictability of the magnitude or angular direction of inaccurate fixation measurement (Morgante et al., 2012). Therefore, more precise insights into the effects of unstable sitting positions on gaze data are needed.

3.2.3 The Current Study

We compared the impact of different factors on the eye-tracking data quality of infant (8- to 12-month-olds) and adult participants. Our goals in the current study were twofold. First, we compared several different calibration targets for their impact on infants' attention and their ability to guide infants' gaze to their centers. Some of the animated calibration targets we tested were already in regular use in laboratories conducting infant eye-tracking experiments, while two additional novel calibration targets were developed for this study based on the sensitivity of the early visual system and infant perceptual abilities. Second, we systematically assessed effects of certain types of head and body movements during the recording session by asking adults participants to perform movements similar to those typically made by infant participants during fixation sequences.

To our knowledge, this is the first investigation to address attention to different calibration stimuli with infants. The study was conducted in remote mode with the eye-tracking system EyeLink 1000 Plus (SR Research Ltd. 2015). The EyeLink system has been predominantly used with adult participants. Its high sampling rate could enhance the detection of inadequate gaze shifts, but be less robust to unrestricted movement and cause measurement artifacts (Niehorster et al., 2017). Investigations of accuracy and precision with infants were thus far conducted with Tobii eye-tracking technology (Hessels, Andersson, et al., 2015; Morgante et al., 2012; Wass et al., 2013; Wass et al., 2014). The Tobii system assesses fixations on dispersal based algorithms instead of the velocity based algorithm of the EyeLink system, and data quality or dependent variables may be affected in a different manner if another technical system is used (Hessels, Cornelissen, et al., 2015). Moreover, the Tobii system uses different calibration procedures that allow missing calibration points and graphically indicate gaze distance to the calibrated target (Tobii Studio User's Manual, 2016; for a discussion of the procedure see Morgante et al., 2012). In spite of the differences between eye-tracking systems, our investigation of the effects of different calibration targets and movement types on accuracy using EyeLink technology will provide valuable insights for infant eye-tracking studies using other technical systems.

3.3 Method

3.3.1 Participants

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Ethics Committee of the Max Planck Institute for Human Development. The final sample of infant participants recruited from urban and suburban regions of a large European city were 29 healthy, full term infants (age: M = 10 months, 8 days, range = 8 months, 0 days to 12 months, 13 days; 14 female). All infants had normal vision without correction. An additional four infants were recruited but excluded from the final sample because they could not be calibrated due to excessive movement (2 infants), or their eyes were not detected by the eye-tracker (2 infants). We did not assess eye color because it was outside of the scope of the present investigation (for a discussion of eye color affecting infant eye-tracking data quality, see Hessels, Andersson, et al., 2015). The adult sample consisted of 25 participants (age: *M* = 24.9, *SD* = 3.96, *range* = 19 – 34 years; 11 female). All adult participants had normal vision without correction and all adult participants were included in the analysis. Our infant and adult sample sizes were chosen based on those used in similar investigations (e.g., Dalrymple et al., 2018; Morgante et al., 2012; Wass et al., 2013) and to be within the recruiting capabilities of a wide range of infant labs. All participants were recruited from participant databases and tested in the Max Planck Institute for Human Development, Berlin, Germany. Both participant groups received 10 Euros and infants additionally received a participation certificate.

3.3.2 Stimuli

The six calibration targets we tested were animated geometric forms (see Figure 3.8a.). The calibration targets we focused on included: a.) differing concentric forms (spiral, star-like, or circular), b.) blurred contours vs. equally distributed contrasts, and c.) different types of motion around a center (twisting, looming or blinking). We focused on abstract symmetrical forms because stimuli that resembled naturalistic figures (e.g., faces, ducks) were expected to guide infants' gaze to non-central areas of interest (e.g., eyes and mouth of a face, head or tail of an animal). Symmetrical forms equally surround the target's center so that attention is not drawn by irregularities of the silhouette. We therefore sought to compare the gaze elicited by different types of symmetrical forms, some with blurred contours at the outer edges and some without. All of our targets also exhibited some movement to attract infants' attention. The zooming in and out motion gives the impression that the targets are looming towards the participant and receding again. In addition, spirals provide concentric movement effects when they twist. Due to the limitations of infants' attention, we did not parametrically vary all possible feature and movement combinations. Instead, we investigated whether combinations of graphical forms and movement would elicit more central attention.

Contrast and size values were chosen to fit the visual capability of the infant age group (Aslin & Smith, 1988). The calibration targets expanded to a maximum diameter of up to 5° visual angle, and shrank to minimal diameters of between 2.5° and 0.5°, depending on their specific design and behavior. All calibration targets were accompanied by sounds corresponding to their looming and twisting behavior. Video examples of the calibration targets are provided online (https://osf.io/3k8jp/?view_only=e8075dc7bf0e4ab780c5e620b8f4860f). The calibration targets used for the initial calibration procedure were presented on a grey background, while repetitions for validation or as part of the trial sequences were presented on different monochromatic backgrounds of the same luminance level as the grey (see Figure 3.8b and the Section 3.3.4 *experimental design* for further descriptions).

The part of the experiment that was exclusively performed by adults (see Movement Task section below) used the 13 point calibration procedures provided by the manufacturer (SR Research Ltd. 2015). The stimuli that were used during the trial sequences of the adult movement block consisted of small filled circles (Ø = 0.5°) with a crosshair centered on it and a thin blurred circle surrounding the center at Ø = 3° to facilitate peripheral detection (see Figure S3.4). They appeared at 9 different screen locations (see Figure 7a and S1 Method) in randomized order.

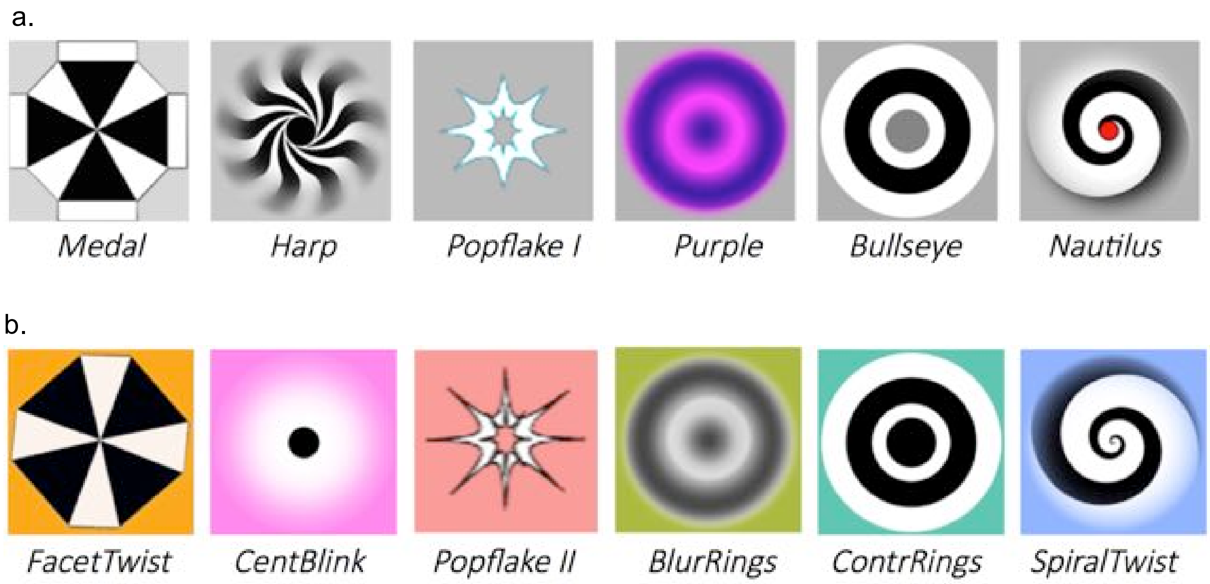


Figure 3.8: Examples of calibration targets and their variants in their fully developed form of appearance.

Top row (a): calibration targets expanding to 5° when presented on the screen. Bottom row (b): Variants of the calibration targets used for the Spread trials, expanding to 17°. The modified calibration targets Popflake II and BlurRings (a variant of Purple) kept their distinct movements. Harp and Nautilus were reduced to CentBlink (a blinking central disc surrounded by a white corona) and SpiralTwist (a twisting spiral). ContrRings resembled Bullseye but lacked the blinking center. FacetTwist was identical to Medal except that it did not show the four white bars. Both ContrRings and FacetTwist kept their contrast in the periphery, while CentBlink and SpiralTwist had a blurred periphery. Two different background colors for each target variant in (b) were equally balanced over the participants. Video examples are provided online (see https://osf.io/3k8jp/?view_only=e8075dc7bf0e4ab780c5e620b8f4860f). Harp, Nautilus and the modifications of the target variants are developed for the present study by the first author, the other calibration targets were kindly provided by other laboratories. We thank Scott Johnson, Gustav Gredebäck, Elika Bergelson and SR Research.

3.3.3 Apparatus

An EyeLink 1000 Plus (SR Research Ltd. 2013 - 2015) eye-tracking system was installed on a host PC with 32bit operating system Intel(R) Core(TM)^2 Duo processor with 2.80GHz and 2Gb Ram. Gaze was recorded using an EyeLink 1000 Plus High-speed Camera with a 16 mm / 1:14 lens and an CL Illuminator TT890. Monocular gaze position was recorded without head stabilization in remote mode. The device has a recording accuracy of 0.25° - 0.5° and a precision (RMS) of < .05 visual angle, as specified by the manufacturer. Pupil and corneal reflection was assessed in a sampling rate of 500 Hz. A target sticker was placed on participants' faces (cheek or forehead) and the camera of the eye-tracker was placed approximately 60 cm in front of the target sticker as recommended by the manufacturer (the possible range is 40 cm - 70 cm for remote mode tracking; EyeLink, 2015). The presentation monitor (Samsung UE50H6470SS, 80 cm by 63 cm, 50" display, with 1280 by 1024 pixel resolution, and 400Hz CMR refresh rate) was set at a distance of 140 cm away from the participants' eyes to approximately fit the trackable area of 32° by 26° visual angle in accordance to the manufacturers suggestion.

3.3.4 Procedure for Infant Experiment

Infants were seated on their caregiver's lap with a small bullseye sticker placed on their forehead that was recognized by the eye-tracking camera. Parents were reminded to sit quietly and not direct their infant's attention during the experiment. Corneal reflection and contrast sensitivity of the eye-tracker were adjusted while an introductory animation clip was shown. The room was dimmed and the eye-tracking device was operated quietly from behind a curtain. The presentation could last up to 9 minutes maximum, but was terminated early if the infant showed fatigue, did not attend to the screen anymore, or if the caregiver requested to end the session.

Experimental design. The infant experiment consisted of six trial sequences. Each trial sequence started with a five point calibration using one of the six calibration targets (see Figure 3.8a). A different calibration target was used for this initial calibration before each of the six trial sequences; the order of the trial sequences was randomized across participants. Calibration success was determined by evaluating the symmetry of the pattern of gaze points shown on the eye-tracking monitor after the infant had attended to all five target locations. Following the instructions provided by the manufacturer, these gaze point locations were of equal distance to each other (EyeLink, 2015). If gaze points were registered at less than five locations, the calibration procedure was not accepted by the eye-tracker and needed to be repeated. We stopped the experiment if three calibration attempts were unsuccessful.

After successful calibration, if the infant still seemed interested in the screen, a five point validation was performed with the same calibration target on a differently colored background. This was done to tentatively assess calibration success during infant eye-tracking, similar to how it is commonly done during adult eye-tracking. If the infant lost interest and started to move during this validation procedure, the validation was stopped immediately and the trial sequence (see below) was initiated so that the accuracy of the calibration would not be impaired through intermediate movement. If the infant already began fidgeting during calibration, the experimenter skipped the validation entirely and went directly on to the trial sequence. After the initial calibration procedure, infants were shown three types of trials in the trial sequence (Example videos for the three trial types are provided online (https://osf.io/3k8jp/?view_only=e8075dc7bf0e4ab780c5e620b8f4860f):

a.) *Preference* trials*:* These trials examined infants' preference for looking at the six different calibration targets (see Figure 3.8a) when they were presented simultaneously on the screen. To do this, the different calibration targets were shown four at a time, evenly spaced in four quadrants of the screen (see Figure S 3.**5**). Infants were shown three different combinations of four calibration targets during one trial, such that each of the six calibration targets appeared twice. The stimuli were shown at the same four screen locations for each of the combinations. Each combination was shown for 8s, resulting in a 24s total duration for the trial. The Preference trials were accompanied by music and occurred only once in each trial sequence.

b.) *Verification* trials*:* In these trials, we assessed the accuracy and precision of infants' gaze elicited by each calibration target (see Figure 3.8a). A calibration target was presented in parallel at three of the five screen locations used in the initial calibration procedure; the configurations across the five possible locations were randomly selected out of several potential combinations and varied across trials to avoid confounds from particular screen locations (see Figure S3.6). The calibration target used in each Verification trial was always different from the target used for the initial calibration procedure. A Verification trial lasted for 12s and was accompanied by one of two rhythmic Marimba sounds. The parallel and synchronous movement of the three identical calibration targets was intended to maintain infants' interest during these trials while their gaze to each of the targets was recorded. Three verification trials occurred in each trial sequence with alternating calibration targets.

c.) *Spread* trials: Here, we compared the accuracy of gaze elicited by variants of the six calibration targets (see Figure 3.8b) during the time course of a trial. We created variants of the calibration targets for these trials in order to understand which visual attributes elicit more accurate gaze (see Figure 3.8 for a precise description of the modifications). A single target was presented at central location on the screen and loomed from a size of 1° to 17° peaking at 2s, and decreased back to 1° until the trial terminated 6s later. In these trials, the target variants were shown one at a time. Three spread trials occurred in a trial sequence; each spread trial showed a different target variant.

Taken together, there were seven trials in each trial sequence (1 *Preference* trial, 3 *Verification* trials, and 3 *Spread* trials) that were shown in randomized order within each of the six trial sequences. Moreover, we randomized the order of the six trial sequences across participants. Finally, two versions of the experiment were alternated to balance the combinations of targets used for the initial calibration procedure and the targets shown in the trial sequence.

3.3.5 Procedure for Adult Experiment

For the adult participants, the same eye-tracker setup was used as with the infants. The adult version of the study lasted approximately 30 minutes. At several pre-defined time points during the experiment, participants were offered a short break.

Experimental design. The adult version of the experiment consisted of four blocks. The first block was a sequence of practice trials consisting of instructions and examples of the respective trials. During this first block, adults were instructed to view the target videos played during the Preference, Verification, and Spread trials freely while keeping their head and body in a central and stable position. Adults were informed that during the movement tasks (see below), they would be asked to perform certain movements at predetermined points in the trials, and that the type of movement would be indicated on the screen. Adults were instructed to look at the targets that appeared during these trials as precisely as possible during or after performing the respective body movements (described in detail below). If necessary, the instructions were explained orally. Adults were also asked to practice the body movements described on the screen with the guidance of the experimenter.

The second block of the experiment was almost identical to the infant version described above, including the five point calibration, except that adults performed two fewer *Preference* trials to reduce the total testing time. The third block investigated the effect of head and body movements on data quality and was unique to the adult version of the experiment. It began with a 13 point calibration followed by four movement sequences. Each sequence started with instruction slides. Participants were asked to perform movement tasks while a static target appeared at one of nine locations distributed grid-like over the screen (see Figure 3.14a and SI S1 Methods for details). The target was a small filled circle (0.5°) in front of a cross-hair pattern (3°); the same target was used throughout the movement sequence. The distance between the target locations was approximately 9° in the horizontal and vertical dimension. The target was presented for 1s at a location, with inter stimulus intervals of 1s. The movement tasks adults were asked to perform were:

a.) *Fix*: Keep their head still and focus on the targets as precisely as possible by only moving their eyes (control condition).

b.) *Head Movement*: Focus on the targets as precisely as possible with the direction of their head following the direction of their eyes. This task mimicked infants' tendency to follow visual stimuli with their head as well as their eyes.

c.) *Side Movement*: Turn their head and upper body out of the area tracked by the eye-tracking camera in the direction indicated by arrows, and then directly return to the central position to fixate precisely on the following targets until the next directional arrow was shown. The arrows appeared three times during the task, pointing to the left, to the right, and upwards. With this movement task, we assessed data quality after the eye-tracking camera had to deal with fast movement and loss of the eyes and the bullseye sticker, as frequently occurs when infants look away from the screen.

d.) *Bend Movement*: Bend about 10 cm (4 inches) forward towards the monitor and stay in this position while directing their gaze on the subsequent visual targets as precisely as possible. Changes in the distance towards the screen are another common occurrence during infant eye-tracking.

The movement sequences consisted of 27 trials.

The final block was the *Calibration-Repetition* block which was also unique to the adult version of the experiment. This block began with another 13 point calibration, then all six calibration targets (see Figure 3.8a) were repeated one at a time in random order at five screen locations identical to those during the five point calibration procedure used with infants. *Calibration-Repetition* was intended to compare our two accuracy measures *Displacement* and *Instability* (described below) for each of the targets. As in the first block, participants were asked to direct their gaze towards the stimuli in a way that reflected their natural interest (free viewing), but not to move their head or body during this part of the experiment. The stimuli were shown on a grey background with their original sound for 6s each.

3.3.6 Data Preparation

Trials were excluded from analysis if the recorded gaze proportion was below 50% of the full trial duration (infants $N = 88$; adults $N = 9$). This exclusion criterion, which may seem liberal for studies comparing infants with adults (see e.g., Morgante et al., 2012), was set because variance in data quality was necessary for the analysis. In addition, if single calibrations during the experiment could not be performed satisfactorily because of temporary movement of the participant (infants $N = 3$) or because of technical problems (infants $N = 1$; adults $N = 6$), that particular trial sequence was excluded.

For saccade detection, a velocity based algorithm was used, with thresholds of velocity 30°/sec, acceleration 8000°/sec^2, and motion 0.1°, and a heuristic filter was applied to reduce velocity noise in favor of saccade detection, as implemented by the manufacturer. Gaze was defined as fixation if it was not recognized as saccade or blink. We used these preinstalled settings because they are the most commonly used criteria and because every change in the thresholds will affect the outcomes (Holmqvist et al., 2011) and would reduce the generalizability of our results. Fixations that were shorter than 50ms, which is one of the post-recording thresholds of the EyeLink software, remained in the analysis because they were considered an indicator of reduced data quality.

We assessed the participants' head distance change after calibration. This was done by subtracting the head camera distance at the moment the calibration was accepted from all other data points of the trial sequence. This measure allowed us to estimate the amount of movement for each participant. The EyeLink 1000 Plus data output provides the distance between the eye tracking camera and the bullseye sticker on the participant's head in millimeters. Note that this measure does not indicate the exact direction of movement[3].

To assess the proportion of recorded gaze, all samples with gaze data were divided by the total number of possible samples during a trial. For inferences about data quality, only points of gaze (POG) within a fixation were used. To further exclude POGs that most likely were not related to a distinct task, areas of interest (AOI) and periods of interest (POI) were defined. The AOIs covered the calibration target and a radial space around it large enough to include misplaced POGs due to inaccurate measurement, but small enough to exclude gaze that was

---

[3] EyeLink 1000 Plus also provides coordinates for sideways or vertical movements, but their units are not clearly defined. EyeLink notes that all values indicating head movement in the data output "are intended for a qualitative indication of subject head position in the camera coordinate. If you need quantitative data output for the head movements and rotation angle, you will need an independent head tracker" (EyeLink Data Viewer User's Manual, 2015, p. 131).

directed at the screen for other reasons, such as gaze at the empty screen center, or intermittent fixations. The POIs began from the first moment when participants' visual attention was directed at one of the targets during our trial sequences. We defined this moment as the first time point when the average of all participants' fixation positions was inside the AOI of the specific trial. The POIs ended when less than the average of all participants' fixation positions were inside the AOI. The POIs excluded orienting and anticipatory fixations at the beginning of a trial. Because POIs were contingent on the AOI of the specific trial, the starting and ending points of POIs differed between the trial types (see Table S 3.14 for a precise description of the AOIs and POIs).

3.3.7 Dependent Variables: Precision and Accuracy Measures

For our study we defined precision in line with Holmqvist et al. (2011) as the ability of the eye-tracker to reproduce a measurement, and spatial accuracy as the offset between the expected and the recorded gaze position. We assessed precision in two ways: first as a root mean square inter-sample distance of POGs (termed *RMS*, Holmqvist et al., 2012) and second as the distance between POG coordinates and their centroid during a fixation, divided by the amount of included POGs (termed *Dispersion*; Komogortsev, Jayarathna, Koh, & Gowda, 2010). Higher values of both precision measures indicate lower precision. During infancy, gaze points during a fixation cover a larger area than during adulthood (Luna, Velanova, & Geier, 2008; Zihl & Dutton, 2015), which must be kept in mind when precision is based on distances between POGs. Nevertheless, impaired precision can affect the proportional looking time to AOIs (Wass et al. 2014).

Accuracy was calculated in two ways as well. For trials following 13 point calibrations during the adult experiment, spatial accuracy of a fixation was assessed as the mean Euclidian distance between all fixational POGs and the stimuli center (termed *Displacement*). In the part of the experiment that was performed by infants and adults and that used animated calibration targets, accuracy was scored differently in order to separate calibration related displacements from gaze spread elicited by the stimuli. We calculated the Euclidean distance between all fixational gaze points occurring during the POI of a trial and their centroid. This score provides an estimate of the spatial spread or density of fixations (termed *Instability*)[4]. Displacement and Instability address distinct characteristics of accuracy. In contrast to

---

[4]Note that Gredebäck et al. (2009) used the same measure but termed as RMS.

Instability, Displacement does not distinguish between fixations that are close together and others that are wide spread if they have a similar distance to the target's center; therefore the two measures might lead to diverging values. To validate the use of Instability as a measure of accuracy, we compared both accuracy measures in the adult Calibration-Repetition task.

The units of all gaze related measures are degrees of visual angle.

3.4 Results

3.4.1 Statistical Analysis

In the part of the study that was performed by both infants and adults, infants successfully completed 761 trials (M_{infant} = 26.2 trials per participant, SD = 10.7, *min* = 8, *max* = 42) and adults completed 976 trials (M_{adult} = 39.4 trials per participant, SD = 3.3, *min* = 25, *max* = 40). In the infant sample, there were no differences between male and female infants in the proportion of the recorded gaze (M_{female} = .88, M_{male} = .88, t = .05, df = 125, p = .95), or in the precision measure Dispersion (M_{female} = .38, M_{male} = .41, t = 1.55, df = 117, p = .12). There was also no correlation between infants' age and Dispersion (*cor* = .24, t = 1.3, df = 27, p = 0.2) or proportion of recorded gaze (*cor* = -.12, t = .61, df = 27, p = 0.5). The covariates age and sex were therefore not included in the main analysis. Further descriptives of the data for the joint infant-adult part of the experiment are provided in the supporting information (SI) Section 1 in S2.

We assessed the effects of our independent variables via linear mixed-effects models using the lme4 package (Version 1.1-12; Bates, Mächler, Bolker, & Walker, 2015) in R (Version 3.3.3). Linear mixed-effects models (LME) are suitable for our study because they tolerate the unequal number of trials provided by our participants (for an application see: Laubrock, Engbert, Rolfs, & Kliegl, 2007). In the models, random slopes were specified for variations of the variable of interest between participants (Pinheiro & Bates, 2000).

3.4.2 Analysis Calibration Targets

Preference trials. To examine which targets attracted participants' attention, we analyzed how long they spent looking at the different calibration targets using dwell time. Total dwell time towards the calibration targets was calculated for the POI of an individual trial by the summing up all gaze points during fixations in the AOI of a target. Dwell time was then transformed by taking its square root to fit the data to a normal distribution. An LME model was conducted to infer how dwell time to a stimulus was explained by the kind of target video presented. The effect of calibration target was taken as random at the participant level, and participant group was included as fixed effect covariate. Calibration target ($F(5) = 47.9$, $p < .001$), participant group ($F(1) = 15.5$, $p < .001$), and their interaction ($F(5) = 8.7$, $p < .001$) substantially contributed to the model, which is confirmed by likelihood ratio tests, indicating that removing video ($\chi^2(5) = 77.8$), group ($\chi^2(1) = 9.5$) or their interaction ($\chi^2(5) = 32.9$) significantly decreased the goodness of fit (all $p < .005$). The estimated random effects accounted for a large part of the variance. Figure 3.9 illustrates dwell times estimated by the model as a function of the calibration videos and the participant groups.

For the infant group, Popflake I received the most attention. Popflake I dwell time was higher than for Bullseye (β = 1885ms, $SE = 138.8$, $p < .001$), Nautilus (β = 1742ms, $SE = 174$, $p < .001$), and Purple (β = 1421ms, $SE = 219.7$, $p < .05$). Bullseye was attended to for a shorter time than the other targets (all $t \geq 2.7$, $p < .01$) except Nautilus and Purple (all $t < .9$, *n. s.*).

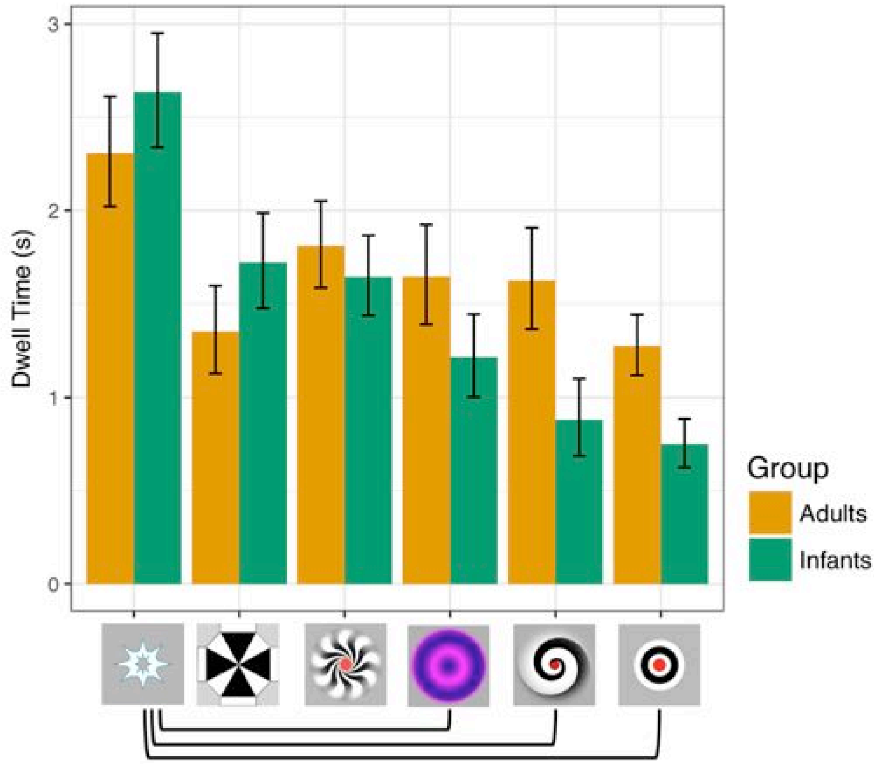


Figure 3.9: Preference for the calibration targets.
Dwell time and 95% *CI* as a function of calibration target. Brackets depict significant differences (p > .05) between the targets. Predicted means in this and in the other plots are estimated and back transformed with the R package predictmeans version 0.99 (Luo, Ganesh & Koolaard, 2014).

Verification trials. To assess the accuracy of infants' gaze elicited by the different calibration targets, we asked if Instability was affected by the calibration target. We also asked if the precision measure Dispersion was affected by the calibration target that was used for the initial calibration procedure of the respective sequence (for means and standard deviations see SI Table S3.17).

The dependent variables (DVs) were log transformed to fit normal distributions. Instability was best explained by the covariate participant group ($F(1) = 109.2$, $p < .001$), the attended calibration target ($F(5) = 12.1$, $p < .001$), head distance change ($F(1) = 10.8$, $p < .001$) and the interaction between group and target ($F(5) = 3.9$, $p < .01$). Removing any of the model terms led to a significant reduction of fit (all *p's* < .01). Intraclass correlation associated with the participants was controlled for by specifying participants as random intercept and target at the participant level as random slope. Instability in the infant group was higher than in the adult group ($\beta = .27°$, $SE = .04$, $t = 5.9$, $p < .001$), and a larger change of head distance after calibration led to higher instability ($\beta = .002°$, $SE = .0006$, $t = 3.4$, $p < .001$). Nautilus elicited the lowest Instability in the infant group, differing from Bullseye with $\beta = -.13°$, $SE = .047$, $t = -2.8$, $p < .01$ (see Figure 3.**10**). No other comparisons were significant.

The usage of a particular target for the initial calibration procedure only marginally predicted the precision measure Dispersion ($F(5) = 1.97$, $p < .10$). Instead, Dispersion was best estimated in an LME model that included participant group ($F(1) = 161.9$, $p < .001$), head distance change ($F(1) = 31$, $p < .001$), and as random slope head distance change at the participant level. Adding initial calibration target changed the model fit by ($\chi^2(5) = 9.8$, $p = .08$, *n.s.*), and removing any of the other variables significantly reduced its fit (all $p < .05$). Infants' fixations had a higher Dispersion than adults' fixations ($\beta = .16°$, $SE = .012$, $t = 12.7$, $p < .001$), and if head - camera distance increased after calibration for 10 mm, Dispersion increased for .013° ($SE = .0005$, $t = 2.6$, $p < .05$).

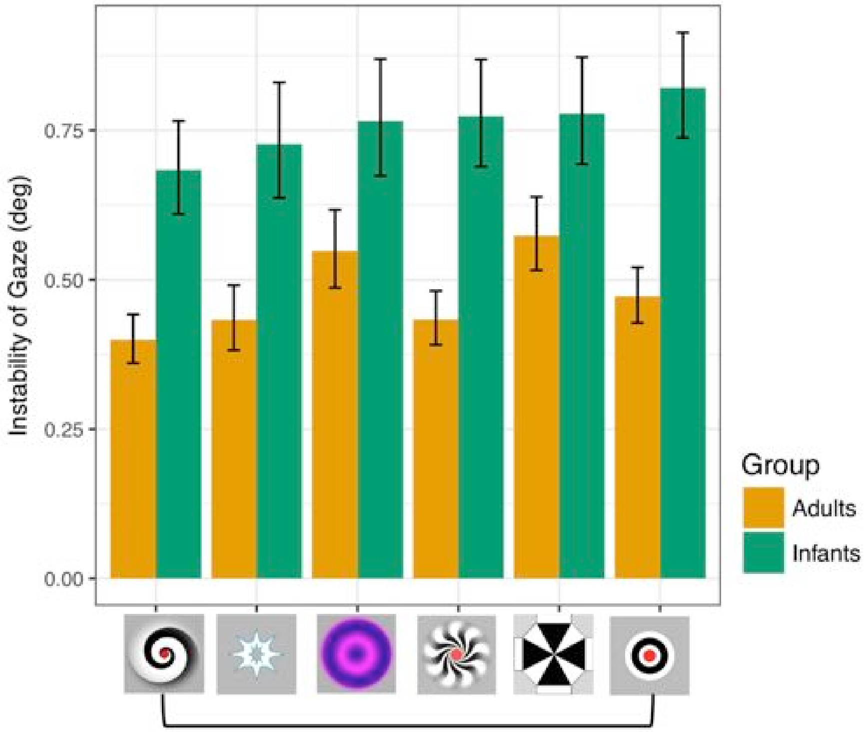


Figure 3.10**: Instability as a function of calibration target and participant group.** Values are back transformed and estimated for a head distance of 13.2 mm, with a CI of 95%. Instability of gaze in the adult group differed from the infant group in that adults attended the videos Medal and Purple with lower accuracy than the infants (all t < 2.1, p < .05). The bracket indicates the difference found in the infant group (*p* < .01).

Spread trials. The 6s time course of the Spread trials was segmented into bins using the following data driven procedure. First, we identified turning points in the slope over which the infant Instability measure developed over time with the R package strucchange (Zeileis, Kleiber, Krämer, & Hornik, 2003). We then defined six bins of approximately similar length around each turning point. Participants' Instability values within a bin, and within the entire POI for a particular target, were then aggregated in order to analyze differences in gaze accuracy towards the target variants over time. Because the targets in the Spread trials increased in size and decreased again over the course of the trial, the six bins also captured gaze toward the target at different sizes.

When looking at the whole POI, Instability was best explained in a model including target variant ($F(5) = 16.3$, $p < .001$), participant group ($F(1) = 12.8$, $p < .001$) and bin ($F(5) = 2.9$, $p < .05$), and their interactions (target variant - bin ($F(25) = 4.6$, $p < .001$; group - bin ($F(5) = 12.5$, $p < .001$). Target variant at the participant level was specified as a random slope. Including the target × group interaction does not improve the fit ($\chi^2(30) = 1.5$, *n.s.).*

Gaze towards the stimuli during the subsequent bins was then analyzed. Within each bin, the effects of target variant and the interaction between participant group and target variant on Instability of gaze were estimated, with participants as a random intercept (Figure 3.**11**a). To account for multiple comparisons, we will only report differences related to infants' instability of gaze towards the stimuli at a significance level $p < .01$ (Figure 3.**11**b).

Infants' gaze became less stable over time and varied by target (Figure 3.**11**a, right panel). In the earliest segment between 0.8 and 1.7s, Bin 1, only CentBlink triggered lower Instability than ContrRings and FacetTwist. Bin 2 between 1.7 and 2.55s, which included the fully expanded stimuli, revealed three target variants with better accuracy than the other target variants. Precisely, CentrBlink and Popflake II led to more stable and central fixations than ContrRings, FacetTwist and BlurRings, and SpiralTwist elicited more accuracy than ContrRings. This same pattern of results occurred within Bin 3, this time showing the largest discrepancies of the entire trial. In Bin 4, between 3.3 and 4.15s, only Popflake II differed from the three lower accuracy target variants. However, in Bin 5 gaze towards Popflake II increased in Instability, and only CentrBlink and SpiralTwist differed from FacetTwist, the latter as well from ContrRings. In Bin 6 all targets were viewed with similar, increasingly high Instability (for coefficients, standard errors and significance values see Table S3.18).

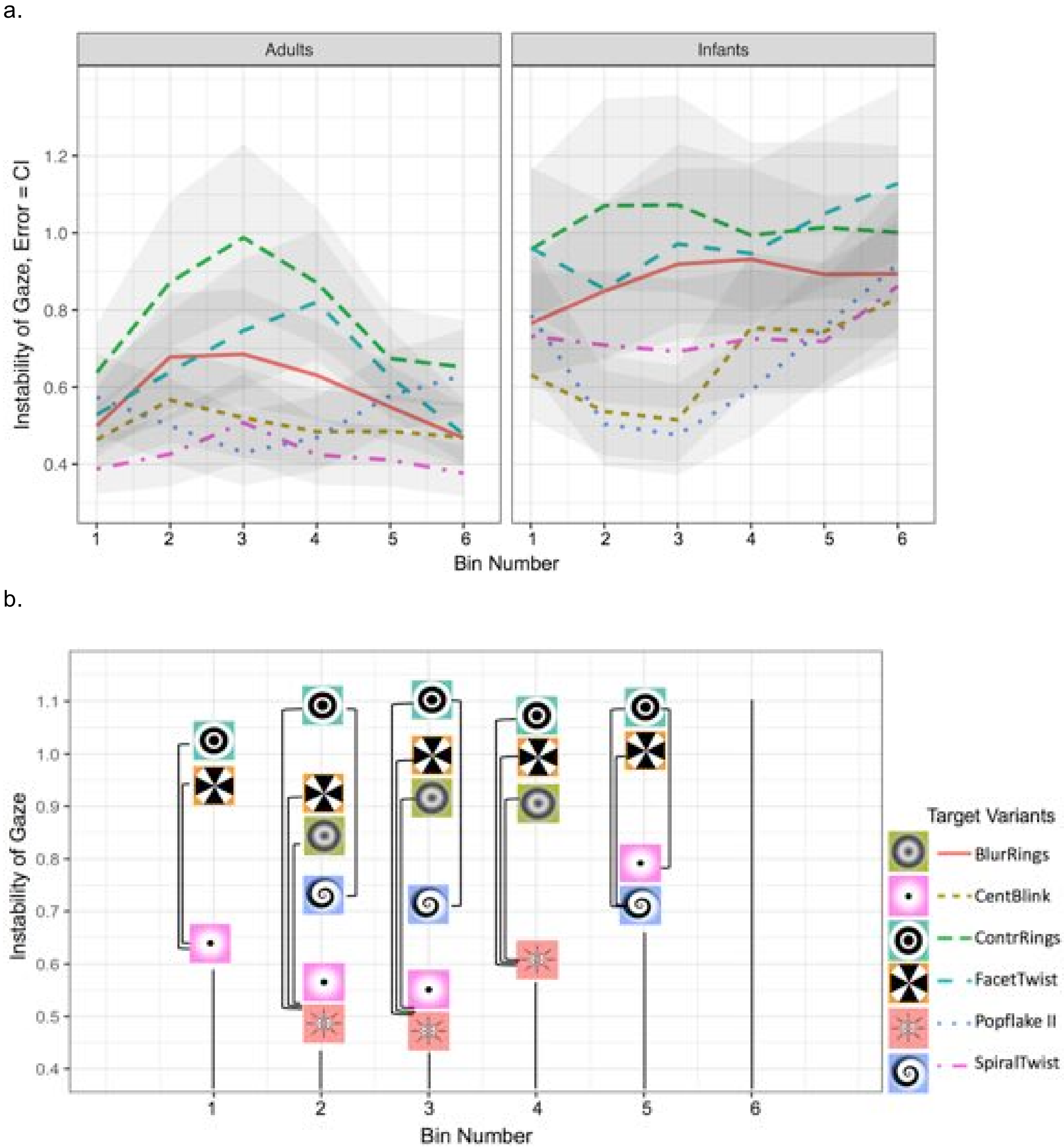


Figure 3.11**: Changes in accuracy during the presentation period of the spread trials.**
(a) Predicted Instability in *deg.* as a function of target variant and participant group, with 95% CI. (b) Infants' Instability as a function of target variants and bins. Brackets indicate differences with significance level p < .01. For coefficients and standard errors see SI Table S3.18. The target variants were most expanded at 2s. Numbers on the x axis represent the 6 time bins.

Calibration-Repetition trials. Next, we assessed adult participants' accuracy scores with our DVs Displacement and Instability. This allowed us to compare the performance of these two accuracy measures.

The LME model that explained Displacement best included the factor calibration target ($F(5) = 4.4$, $p < .01$), the factor target location ($F(1) = 4.6$, $p < .05$), the continuous variable head distance change ($F(1) = 22.1$, $p < .001$), and participant as a random intercept. Removing calibration target ($\chi^2(5) = 18.4$, $p < .01$), target location ($\chi^2(1) = 4.2$, p < .05) or head distance change ($\chi^2(1) = 4.3$, $p < .05$) would significantly decrease in the model's goodness of fit. Displacement increased with calibration targets presented at a peripheral location ($\beta = .04°$, $SE = .019$, $t = 2.1$), and with a larger head distance from the screen ($\beta = .007°$, $SE = .002$, $t = 4.7$). The calibration target Nautilus was attended to with the lowest Displacement and differed from all other videos except Harp (all *t's* < 2.5), while Purple was attended to with the highest Displacement differing from Nautilus and Harp, with all *t's* > 2.5.

Instability was best explained by calibration target ($F(5) = 9.5$, $p < .001$), target location ($F(1) = 17$, $p < .001$) and calibration target at the participant level as random slope. Instability increased with targets shown at a peripheral location ($\beta = .04°$, $SE = .009$, $t = 4.1$). Here as well, Nautilus was attended to with the lowest Instability and differed from all other calibration target except Harp (all *t's* > 2.5). Of those targets with low accuracy it was Purple which led to highest Instability scores, differing from all calibration targets except Bullseye, with all *t*'s > 2.7 (see Figure 3.**12**). The measures Dispersion and Instability were correlated ($r_{df\ 714} = .35$, $t = 10$, $p < .001$), indicating an association of medium effect size between the two measures.

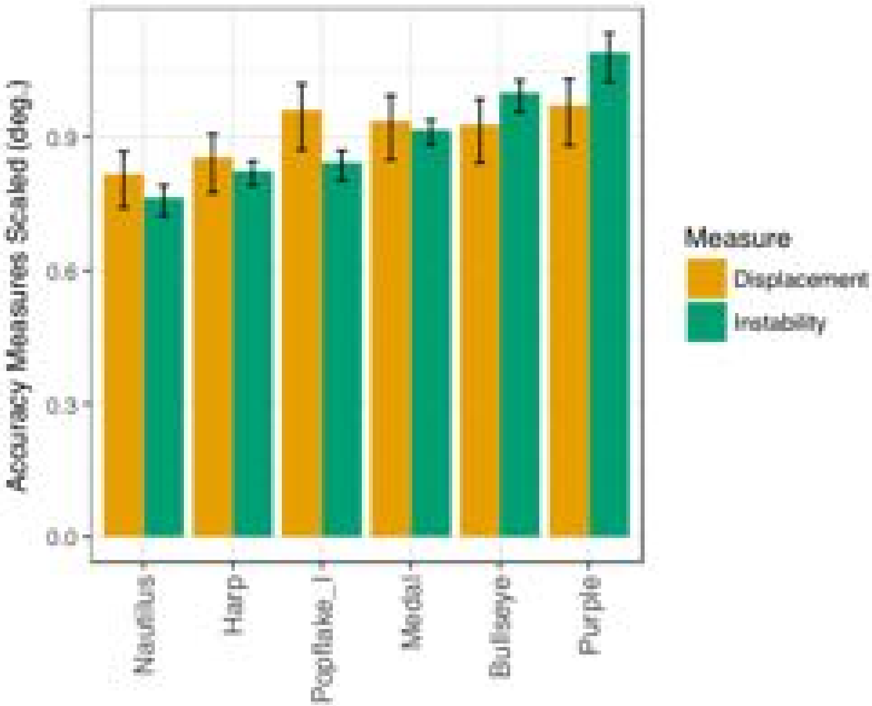


Figure 3.12: **Comparison of the two accuracy measures.**
Scaled accuracy scores of the measures Displacement and Instability as a function of calibration target during the adult task Calibration-Repetition.

3.4.3 Adult Movement Tasks

Finally, using our adult participants, we asked how head and body movements (see section *Procedure for Adult Experiment*) affect accuracy (Displacement) and precision (Dispersion, RMS) compared to recordings without movement (the control condition *Fix*), and if there is an effect of target location on the gaze measurement. The targets appeared at nine screen locations, and were grouped as *Center* (central), *Central-Peripheral* (central on one axis but peripheral at the other axis) or *Peripheral* (all four corners). In all LME models, movement type at the participant level was included as a random slope.

Displacement was best predicted with movement type ($F(3) = 64$, $p < .001$), target location ($F(2) = 88.9$, $p < .001$), and their interaction ($F(6) = 5.4$, $p < .001$; Figure 3.13a). All movement types led to increased Displacement (Side Movement: $\beta = .14°$, $SE = .04$, $t = 3.7$; Head Movement: $\beta = .17°$, $SE = .04$, $t = 3.9$; Bend Movement: $\beta = .55°$, SE = .07 $t = 8.4$), and non-central target locations led to larger Displacement than centrally presented targets (Central-Peripheral: $\beta = .08°$, $SE = .02$, $t = 3.5$; Peripheral: $\beta = .14°$, $SE = .03$, $t = 5.5$). Bending towards the screen significantly increased Displacement at Peripheral locations ($\beta = .13°$, $SE = .04$, $t = 3.6$; all $p < .001$).

Dispersion was predicted by movement type only ($F(3) = 34.8$, $p < .001$). Adding target location to the model did not improve the fit ($\chi^2(2) = 1.3$, $p = .53$), and although the interaction of target location and movement type improved the model fit ($\chi^2(6) = 14$, $p = .03$), we decided against including it for parsimonious reasons and because the interaction without a main effect of target location would not be meaningful here. Head Movement increased Dispersion ($\beta = .046°$, $SE = .008$, $t = 5.6$, while Bend Movement reduced Dispersion ($\beta = -.029°$, SE = .004, $t = 8$; both $p < .001$). Dispersion elicited by Side Movement did not differ from the stable position.

RMS was best predicted by movement type ($F(3) = 63.2$, $p < .001$) and target location ($F(2) = 11.5$, $p < .001$; Figure 3.13b). In a similar pattern as Dispersion, RMS was reduced in Bend Movement ($\beta = -.0023°$, $SE = .0003$, $t = 7.1$), but increased in Head Movement ($\beta = .0025°$, $SE = .0003$, $t = 8.6$) and Side Movement ($\beta = .0003°$, $SE = .0001$, $t = 2.3$). Non-central target locations led to decreased RMS than centrally presented targets (Central-Peripheral: $\beta = -.0004°$, $SE = .0001$, $t = 3.1$; Peripheral: $\beta = -.0006°$, $SE = .0001$, $t = 4.7$; all $p < .05$).

a.

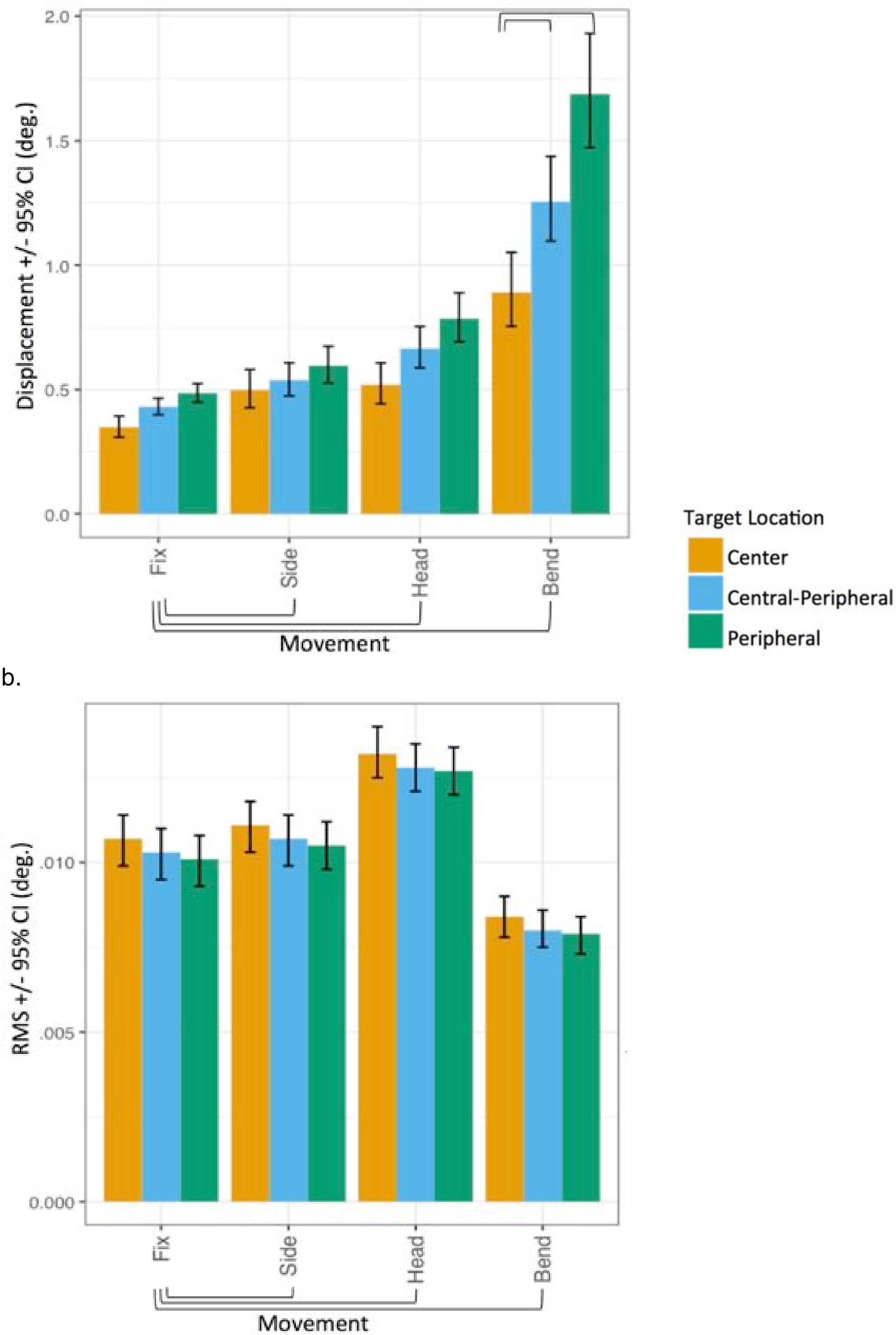
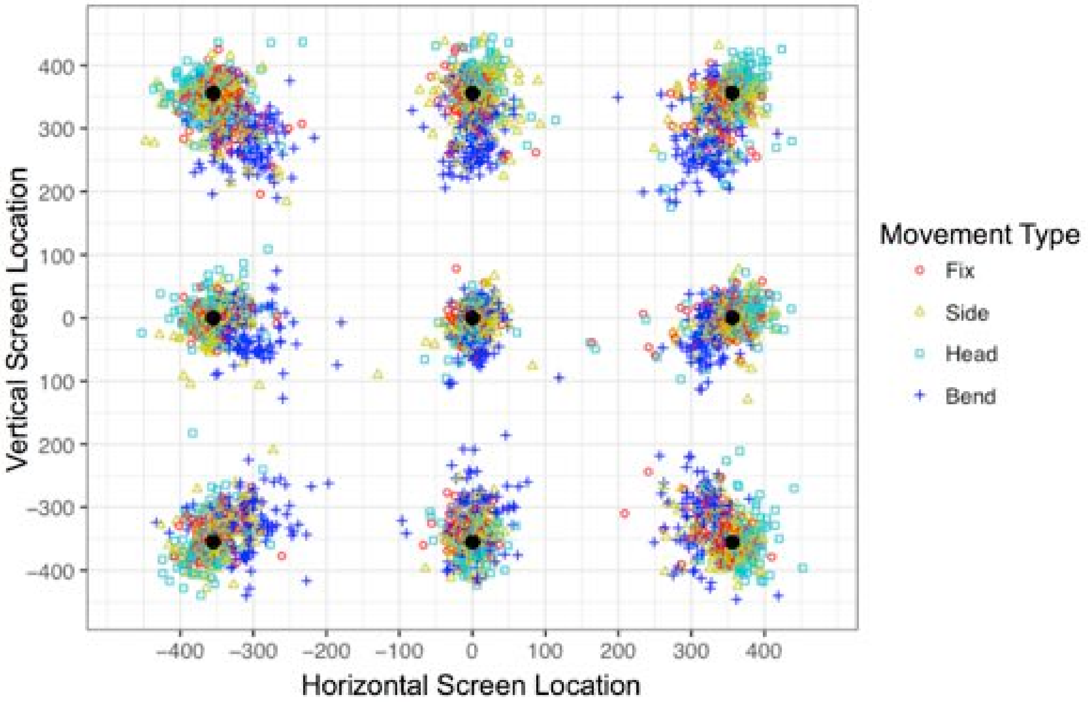


Figure 3.13: Effects of movement type on adult accuracy and precision.

Accuracy (a: displacement) and precision (b: rms) as functions of movement type and target location. note the converse effects for accuracy and precision when approaching the screen during Bend Movement, changing the viewing angle during Head Movement, and when attending peripheral target locations.

a.



b.

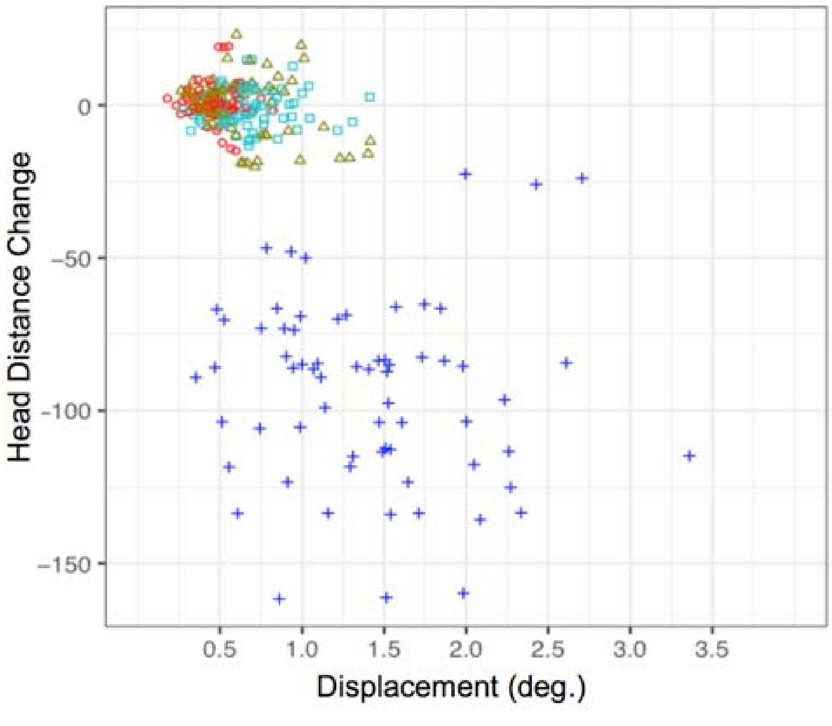


Figure 3.14: Movement type affects the registration of gaze points.
Gaze points during the adult movement tasks plotted on their measured screen locations in pixel coordinates. Black discs indicate the actual target positions, inter target distance was 9° of visual angle (a). Accuracy (x axis) plotted against head–camera distance change after calibration. Negative values indicated reduced distance to the eye-tracking camera in mm (b).

3.5 Discussion

In the present study, we investigated the impact of different infant calibration targets and movements during gaze recording on eye-tracking data quality with infant and adult participants using EyeLink 1000 Plus Remote Mode technology. We found that certain visual attributes of the calibration targets, as well as the duration of their presentation, influenced infants' gaze instability. Targets with interesting centers and low contrast at their periphery resulted in better gaze recording outcomes. Body movement substantially contributed to gaze instability and fixation dispersion. All movement types we tested with adults negatively affected accuracy, as did the eccentricity of a target's location. Movement towards the screen particularly increased peripheral gaze displacement and following a target with head turns resulted in less precise gaze.

3.5.1 Calibration Targets Influence Stability of Gaze

Infants fixated our calibration targets with different gaze stability, demonstrating that some characteristics of an animated graphical form elicited more accurate gaze than others. Interestingly, our results showed that infants' preference to look at a particular calibration target was not predictive of the data quality elicited by that same target in our study. Infants fixated on the target Nautilus for the least amount of time in the Preference trials, but Nautilus nevertheless led to the highest stability of gaze points in the Verification trials. The calibration target which elicited the greatest preference, Popflake I, led to similar gaze stability as Nautilus.

By reducing the attributes of our calibration targets in the Spread trials, we were able to infer which visual characteristics contributed to stable gaze. Our results showed that animations with an interesting center but low contrasts in their periphery (CentrBlink, SpiralTwist), as well as very complex concentric animations (Popflake II), elicit the most stable gaze over time and are therefore better suited for infant calibration. CentrBlink and SpiralTwist share two important attributes with the (not reduced) Nautilus target that performed well in the Verification trials: a blurred periphery and an interesting (blinking and high contrast) center. The target variants leading to less stable gaze consisted of blurred concentric forms without a clear center (as in BlurRings), or point symmetrical patterns with distributed contrast which were not blurred in their eccentric parts (ContrRings and FacetTwist; for a detailed description of all target variants see Figure 3.8).

The decision of when to accept infant's gaze to a target during the calibration procedure is another important criteria for calibration success. Our results from the Spread trials, in which the target variants appeared to loom over the course of the trial, indicated that accuracy dropped similarly for all target variants over time in our infant sample. About four seconds after stimulus onset, infants' gaze started to be less stable even for targets that were fixated more accurately, and after 5 seconds, differences between targets could no longer be found (see Figure 3.11). This is in contrast to adults, who attended to the shrinking targets with increasing gaze stability over time.

To better understand why infants' gaze decreased in stability over time, we compared our Instability measure to the more common accuracy measure of Displacement (also termed "offset" by Hessels, Andersson, et al., 2015) in the adult Calibration-Repetition task. The correlation between Displacement and Instability was of medium effect size in the adult Calibration-Repetition trials, indicating that the two measures were similar but not entirely overlapping. The most obvious difference between the two measures occurred for the visually demanding video Popflake I (see Figure 3.12). Given that the Calibration-Repetition trials were the last block of the experiment, the adult participants were already well acquainted with the targets, and more likely to direct their gaze to details of Popflake I's silhouette as is reflected in the higher Instability score for this target. We therefore interpret the increase of Instability in the later portions of the infant Spread trials as less central gaze, because by this point of the trial, infants became inattentive and increased exploratory gaze around more distributed screen areas. Alternatively, the increase of Instability can also be understood as a loss of interest in the target decreasing in size. These explanations are not mutually exclusive.

Our calibration targets clearly differed in how they elicited central gaze, therefore it was surprising that they only marginally modulated later gaze precision when they were used during the initial calibration procedure. This may have occurred for several reasons. First, fixation control develops until early adolescence (Buquet & Charlier, 1996; Ygge, Aring, Han, Bolzani, & Hellström, 2005). Infants' fixations generally cover a larger area than adults' and are less stable (Luna et al., 2008; Zihl & Dutton, 2015), which may have obscured potential differences during fixation and is in line with the significant effect of the covariate participant group (infant vs. adult). Additionally, movement during recording significantly contributed to the variance of our infants' Dispersion scores, leading to a loss of statistical power such that the effects of our calibration videos were only marginal (see SI Section 1 in S3 Discussion for further discussion of this point).

It was difficult to implement a validation of calibration success for our infant participants as it is commonly implemented in adult eye-tracking. The repeated presentation of the identical target at all five calibrated screen locations directly following calibration typically led to infant impatience and inattentiveness. Therefore, in many of the cases we omitted the validation procedure from the trial sequences. Additionally, the generally poor accuracy score reported by the eye-tracker for the attempted validations may not have been attributable to calibration success per se, but instead to the effects of movement due to infant inattentiveness during the validation procedure (see SI, Section 2 in S2 Results for a description of the validation attempts). As a result of these kinds of difficulties, experimenters often skip validation procedures with infants and instead rely on the pictorial pattern of the calibration map to infer calibration success. In future research it would be worth investigating whether the symmetry of the calibration coordinates provided by the EyeLink output can be quantified and included in the statistical analysis (for similar suggestions based on Tobii technology see Dalrymple, Manner, Harmelink, Teska, & Elison, 2018).

3.5.2 Movement Affects Accuracy and Precision in Opposite Directions

The accuracy of gaze measurement in our adult sample was affected by all of the movement types we examined. When the adult's eyes and the bullseye head sticker briefly moved outside of the area registered by the eye-tracking camera—a common occurrence during infant eye-tracking—and returned to central and stable position before the recording started, the target - POG distance systematically increased by .15°. This adds to the findings of Niehorster and colleagues (2017) who performed a similar task and found a right sided insensitivity of the EyeLink system towards the returning gaze. A similarly strong impact of movement on accuracy occurred during head turns towards the target leading to an increased offset of .17°. Turning the head in the direction of a stimulus is also a common movement during infant eye-tracking, since perceptuomotor coordination accompanies attentional strategies and learning in infancy (Gibson, 1969; Yoshida & Smith, 2008).

Movements toward the screen had the strongest effect on gaze accuracy. Displacement not only generally increased by .55° for this movement type, but was further augmented by .13° for targets presented in the four corners of the screen. In fact, our data revealed that during all trials the presentation of non-central targets systematically added between .08° and .14° to the measured gaze - target distance. This finding underscores the importance of using variable target locations during intermittent drift checks to verify calibration accuracy during infant eye-tracking. The full range of drift would not be detected if drift check targets are located only at the screen center. A warped POG map resulting from intermittent movement could also lead to imprecise post hoc adjustment of gaze data if a one directional displacement is assumed.

There was a different pattern of results for precision during the adult movement tasks. Fixation dispersion was unaffected by target location, while non-central target locations reduced RMS values. Head turns decreased precision as assessed by both scores. However, bending toward the screen seemingly increased precision as assessed by Dispersion and RMS. This apparent increase in precision was surprising given that the bend movement led to the lowest gaze accuracy. The reason for this discrepancy, as Figures 7a,b show, is that movement towards the screen after calibration made the POGs drift towards the center of the monitor. This resulted in a reduction of the size of the POG map and in a shrinkage of the inter sample distances. At the same time, the offset of the measured POGs increased, resulting in higher displacement values especially at non-central target locations. This finding also illustrates the necessity of exploring data in multiple ways to avoid misinterpretation—here, better precision scores clearly do not reflect higher data quality.

The high inter sample distance during the Head Movement task may reflect data quality loss originating from the combination of head turns and movements as the adults turned their heads to follow the movement of the target during this task. A change in the angle of the eyes influences the assessed pupil size (Hayes & Petrov, 2016) which again affects the estimation of POGs (Choe, Blake, & Lee, 2016; *EyeLink 1000 Plus User Manual*, 2015; Nyström et al., 2016). Moreover, the bullseye sticker that indicates a participant's head position moves slightly sideways and in its angle during these kinds of movements. Infants usually spontaneously perform a combination of different movements, including more excessive angular positions than adults. Accordingly, these combinations of movement may have caused the considerably higher RMS values for the infant sample than those of the adult sample ($Md_{Infants}$, = .021, *min = .008, max = .063* compared to Md_{Adults} = .011, *min = .006, max = .02*). This finding emphasizes the care that needs to be taken when comparing participant groups of different age, even if no strong distance changes to the eye-tracking camera are obvious (see SI S3 for further discussion).

Taken together, our findings for the movement tasks demonstrate that the consequences of unconstrained recording situations on gaze DVs are difficult to calculate. Specifications given by manufacturers are usually achieved under optimal conditions and differ from the specifications assessed with naturally behaving participants. Our data quality scores were preprocessed (e.g., means of POIs or fixations limited by AOIs) to estimate the variability that may occur during analysis of gaze from participant groups that cannot be restrained. In the user manual of the EyeLink eye-tracker (2013-2015), head movement of 35 cm in vertical and horizontal direction are said to be tolerated without accuracy reduction for a camera distance of 60 cm (EyeLink, 2015). For movements towards the camera, the system reports a warning if the distance exceeds a 20 cm range, outside of which accuracy can not be guaranteed. However, in our study movement within these ranges clearly affected DVs (for further examples including angular movements and recovery of the eye-tracker after loss of the eye, see Hessels, Cornelissen, et al., 2015; Niehorster et al., 2017). Future studies could investigate the usefulness of including the change in head distance registered by the eye-tracker as control variable during Remote Mode infant eye tracking.

3.5.3 Practical Implications

Our results point to several practical steps that infant researchers can take to improve eye-tracking data quality. Of course, the requirements for gaze accuracy depend on the specific context in which eye-tracking data are collected. Therefore, researchers should take into account the demands of their phenomena of interest and of their experimental design when implementing any of our suggestions.

First, we suggest using calibration targets with an interesting center and low contrast in their periphery or globally distributed complexity. Calibration targets with these characteristics—including some kind of movement to attract infants' attention as all of our stimuli did (e.g., looming, twisting, etc.)—elicit more accurate gaze. Even if the differences in accuracy between the types of target used might only seem marginal in some cases, it is nevertheless important to optimize as many aspects of the calibration procedure as possible. Calibration targets that are not controlled in their distribution of contrast or luminance—even if they are provided by some eye-tracking systems—should be avoided. The calibration targets that worked well in our study are available online (see link in the conclusions section below).

Importantly, gaze toward calibration targets during the calibration procedure should be accepted within the first four seconds because attention towards the targets is higher during this phase. To further facilitate infants' attention when repeated calibrations or drift checks are necessary, calibration animations that elicit precise gaze can be alternated. Additionally, the background color of the screen on which the calibration target is shown can be changed to facilitate infants' interest in the display. Because alterations of the display's luminance level would result in changes in pupil size and affect gaze measurement, changes in brightness entering the participant's eye should generally be avoided in eye-tracking experiments. If the background color change is controlled for luminance, it will not interfere with accuracy (EyeLink, 2015). Moreover, depending on the constraints of the experimental conditions, trials can be accompanied by changing sounds or music. In our study, infants were repeatedly confronted with the same six calibration targets during the trial sequences, and we successfully used background color changes and music as described.

Calibration success is crucial for all infant eye-tracking studies, independent of the technology that is used. Poor calibration procedures have a particularly negative effect on infant eye-tracking procedures because the number of trials in these studies is limited by infants' shorter attention spans. Therefore, the risk of a high amount of missing data and incorrect data points can be mitigated by adopting higher quality calibration procedures.

In addition to optimizing calibration targets and procedures, the diverse effects of movement on our gaze measures in the present study should be kept in mind when planning infant eye-tracking studies. Movement towards the screen has an especially high impact on spatial accuracy, and if fixation positions on AOIs are assessed, researchers should expect misplaced POGs with large offsets especially at peripheral screen locations. In such cases, adapting the AOIs accordingly may avoid alterations of the variables of interest (Holmqvist et al., 2012; Orquin, Ashby, & Clarke, 2016). For example, in paradigms that compare attention to multiple areas of the screen, AOIs could be reduced in size, so that misplaced POGs fall into neutral screen areas rather than being falsely attributed to the wrong AOI. A warped POG map, with larger peripheral offsets, could also lead to systematic errors between central and peripheral AOIs.

Experimenters should be attentive to movement throughout the recording sessions and have recalibration procedures prepared if infants exhibit excessive movement of any kind. The measurement of head target - camera distance provided on the EyeLink camera set-up screen as well as a blurred camera image of the eye can both be used as indicators for distance changes even if the eye-tracker does not provide a warning message. Additionally, implementing intermittent drift checks with central and non-central target locations can help to detect shifts of the POGs and possible skewness of the POG map. These checks can occur at regular intervals during the trials. POG shifts can also be assessed via additional software implemented in the experiment (e.g., Dalrymple et al., 2018; Frank et al., 2012).

Studies targeting psychophysical research questions that are more sensitive to fine grained changes in inter-sample distance should be especially aware of the diverse movement effects. If, for example, participant groups differ systematically in their motoric responses, as is the case for comparisons of infants and adults, the resultant systematic distortions in the assessed data could lead to false inferences about group differences. Studies that are particularly sensitive to dispersion of gaze points should consider the inclusion of drift checks and recalibrations at several predetermined intervals during the recording sequence.

Following these practical steps can help to mitigate the problems of infant eye-tracking and increase the quality of measured gaze.

3.6 Conclusion

During infant eye-tracking, uncertainty about calibration success, fussiness caused by the repetition of calibration stimuli, and body movements during testing are frequent constraints on measurement quality. Our systematic investigation of these constraints with infants and adults revealed some characteristics of calibration targets that elicit more reliable data. These calibration targets can be flexibly implemented in different calibration procedure designs and are provided online, together with the necessary information on the adjustment of the background color (https://osf.io/3k8jp/?view_only=e8075dc7bf0e4ab780c5e620b8f4860f). Using EyeLink 1000 Plus technology, we also discovered heterogeneous effects on accuracy and precision as result of movement types which are common during infant eye-tracking. These findings provide some insight into measures that can be taken to improve data quality when conducting infant eye-tracking studies.

3.7 Acknowledgements

We thank our participants and their parents, J. Eichelsdörfer, D. Erdt, Jann Wäscher and the members of the Max Planck Research Group *Naturalistic Social Cognition* for their assistance. We also thank Sam Hutton for supporting information related to the eye-tracking system. During this research, Karola Schlegelmilch was a pre-doctoral fellow of the International Max Planck Research School on the Life Course (LIFE, www.imprs-life.mpg.de). This research was funded by the Max Planck Society.

3.8 Supporting Information

3.8.1 S1 Methods.

3.8.1.1 Comments on the Video Materials that are Available Online
3.8.1.2 Parameters and Design of the Adult Visual Target
3.8.1.3 Parameters of the Infant Video Targets and Background Colors

3.8.2 S2 Results.

3.8.2.1 Descriptives for the Joint Adult - Infant Part of the Experiment
3.8.2.2 Calibration and Validation Error and Number of Validations Completed
3.8.2.3 Calibration Targets (Infant and Adult Part of the Experiment)

3.8.3 S3 Discussion.

3.8.3.1 The Influence of the Initial Calibration Procedure vs. Head Movement on Precision.
3.8.3.2 Low Accuracy and Precision During Angular Movement

3.8.1 S1 Methods.

3.8.1.1 Comments on the Video Materials that are Available Online

The calibration targets used in this study can be downloaded under the URL:
https://osf.io/3k8jp/?view_only=e8075dc7bf0e4ab780c5e620b8f4860f

These calibration targets may be implemented in the calibration procedures of any planned study. Please ensure that the size of the stimuli will be adapted to the monitor size of the planned study so that it will stay in the range of the size that was investigated here (see parameters in S 2 Methods). It should also be taken care that the screen background is changed exactly to the same color and lightness as the background of the calibration target itself. Otherwise, the square of the video might draw participants attention to the periphery of the stimulus. Moreover, the quality of the videos should not suffer from these transformations. The graphs should have clear edges and the blurred parts should not show obvious pixel borders. The videos come with sound that is synchronized to their movement. Some experiment programming software requires separate sound and video image files. The synchrony should then remain as in the original versions.

3.8.1.2 Parameters and Design of the Adult Visual Target

The targets for the adult participants during the movement tasks were presented for 1s with an inter stimulus interval of 1s. They appeared at 9 fixed locations in randomized order with an inter target distance of 9° visual angle. The locations on the screen were (monitor size 1280 × 1024 pixel; horizontal, vertical): [996,157], [996,512], [996,868], [640,157], [640,512], [640,868], [285,512], [285,157], [285,868].

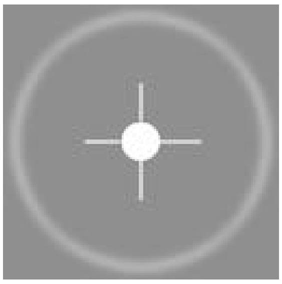


Figure S 3.4: The visual target for adult participants during movement tasks.
The small filled circle measured 0.5° and the crosshair 3° visual angle in diameter.

3.8.1.3 Parameters of the Infant Video Targets and Background Colors

Target Locations. For the initial calibration procedure of each sequence, one of the 6 calibration target videos was shown on a grey background with RGB levels [180,180,180]. The calibration targets were located at the following five positions on a monitor with the size 1280 × 1024 pixel (horizontal, vertical): [640, 512] = center; [285, 868]; [285, 157]; [996,868]; [996,157]. The between target distance of 355 pixel approximates 9° of visual angle horizontally and 8.86° in the vertical dimension.

The calibration targets used in the Preference trials appeared at the four locations in parallel: [306, 306]; [974, 306]; [306, 718]; [974, 718]. A trial consisted of 3 different combinations of calibration targets of 8s each: 1. Medal, Purple, Harp, Bullseye; 2. Popflake, Bullseye, Medal, Nautilus; 3. Harp, Popflake, Nautilus, Purple (see Figure S 3.**5**). To avoid confounds through infants' biases of one side of the screen, two versions of the full trial existed so that each calibration target was shown at all four possible locations. The targets were shown on a grey background [180,180,180].

In the Verification trials, one calibration target appeared three times in parallel at 3 of the 5 locations equal to those of the calibration procedure. The locations were selected so that they covered at least one top and one bottom location as well as one right and one left location to avoid empty halves of the screen (see Figure S3.6). The resulting 6 patterns were randomly selected to avoid confounds through learning of the target locations. The targets were shown on colored backgrounds.

Colors. Repetition of the calibration procedure and presentation of the stimuli variants occurred with different background colors. The luminance level of the changing colors was controlled for by creating them in the CIELAB color space with identical values for luminance (L = 73) and subsequently converting them to the RGB color space for presentation on the screen.

The colors and their RGB levels were:

- blue [147,194,255]
- red [255,146,146]
- turquoise [75,214,178]
- orange [255,168,5]
- green [185,201,54]
- pink [240,168,255]

Video examples of the modified calibration targets used in the Spread Trials and videos documenting the Preference trials and Verification trials can be viewed here:

https://osf.io/3k8jp/?view_only=e8075dc7bf0e4ab780c5e620b8f4860f

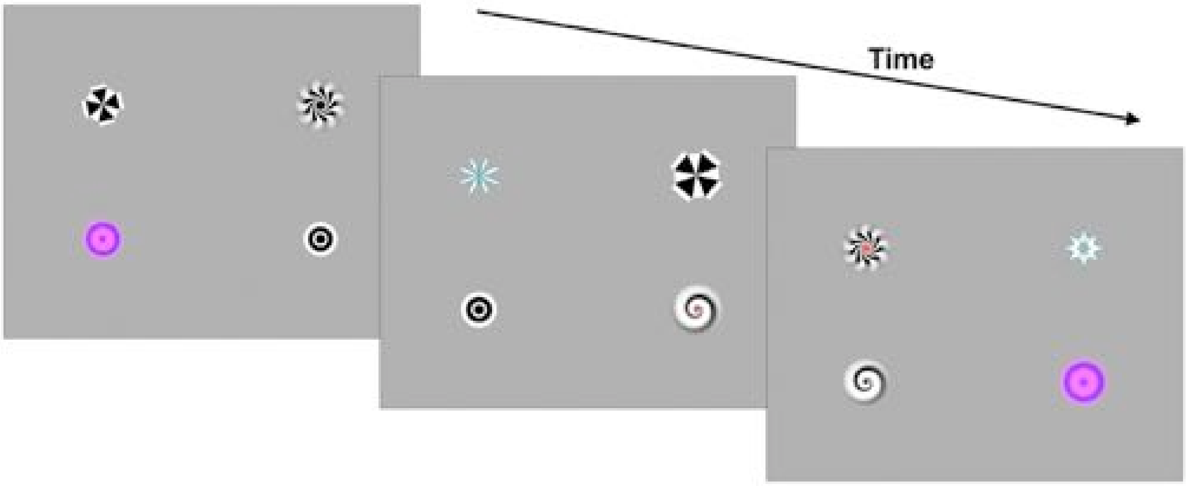


Figure S 3.5: Arrangements of calibration target locations within one Preference trial. The grey squares indicate the screen. Each single target showed the identical movement as during the calibration procedure. The three combinations of calibration targets lasted 8s each, resulting in a total trial duration of 24s. A mirrored version of the full trial alternated with the one shown in the trial sequences, so that each calibration target was presented at all four possible screen locations.

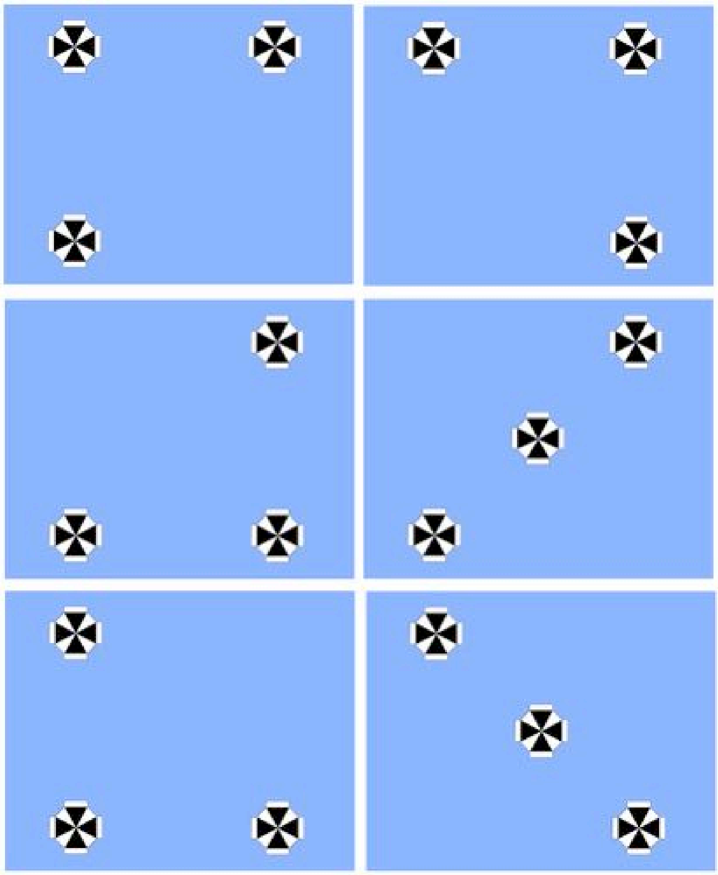


Figure S 3.6**: Arrangements of the target locations in the Verification trials.**
The blue squares indicate the size of the screen in one of the background colors. In each trial, one of the six calibration targets was presented at 3 screen locations synchronously on a colored background. All 6 location combinations that were used in the study in random orders are shown here.

Table S 3.14 The Sizes of Areas of Interest (AOI) and Periods of Interest (POI)

Part of the Experiment	AOI	POI
Preference trials	Four circular areas of 9.2° diameter covering the particular targets. There was a free space between the AOIs of minimum 1° vertically and 7° horizontally.	Identical to the trial length, but separated into three parts that distinguished the distinct selections of stimuli presented simultaneously.
Verification trials	Three circular areas of 10° diameter covering the particular targets which were expanding to maximal 5°.	Started 450ms and ended 4700ms after stimuli onset.
Spread trials	One circular area of 20° diameter, covering the maximum stimuli size of 17°.	Started 800ms after stimulus onset. It ended at 300ms before the trial terminated because there was not enough recorded gaze available to perform an analysis within the infant group.
Adult movement tasks	Circular areas of 10° diameter around the targets.	Started 300ms after stimulus presentation and ended when the trial terminated.
Calibration-Repetition	Radius was the largest stimulus expansion plus .5° leading to a circular area covering 6° in diameter.	Started at 350ms after stimulus presentation and ended when the trial terminated.

Note: The sizes of the AOIs differed between the tasks according to the questions that were expected to be answered. For example, during the adult movement tasks, we expected a larger spread of gaze points (POG) as an effect of movement, that we wanted to capture, while during Calibration-Repetition, we chose narrow AOI for the validation of our accuracy measures. The starting and ending points of POIs differed between the trial types because they were data driven. A POI started when the average of all participants' gaze positions was inside the AOI of the specific trial type and ended when the average of the gaze positions was outside the AOI or when the amount of recorded gaze points was not sufficient for analysis.

3.8.2 S2 Results.

3.8.2.1 Descriptives for the Joint Adult - Infant Part of the Experiment

In the part of the study that was performed by both infants and adults, infants successfully completed 761 trials (M_{infant} = 26.2, SD = 10.7, *min* = 8, *max* = 42) and adults completed 976 trials (M_{adult} = 39.4, SD = 3.3, *min* = 25, *max* = 40).

The two measures of precision, the proportion of recorded gaze, and distance change between the head and the eye tracking camera following a calibration of the adult and infant participants are described in Table S 3.15 for the joint part of the experiment, after exclusion of invalid trials (proportion of recorded gaze within a trial below .5). Figure S 3.**7** visualizes the distribution of the data.

Table S 3.15 Comparison of the Adult and Infant Sample

Measure	Infants		Adults		Difference	
	M	*SD*	*M*	*SD*	*t*-value	*p*
RMS deg.	.022	.008	.011	.002	36.5	<.001
Dispersion deg.	.391	.185	.183	.074	29.2	<.001
Proportion of gaze	.885	.130	.982	.023	-20.3	<.001
Head distance change* mm	27.6	25.8	3.2	3.4	-25.9	<.001
Fixation duration ms	529	875	600	755	-7.3	<.001

* Here, the positive distance change between head and camera after calibration is reported, that was the dependent variable used in the joint adult - infant part of the experiment. In contrast, in the adult movement tasks, we used negative and positive values, indicating reduction and increase of the head camera distance.

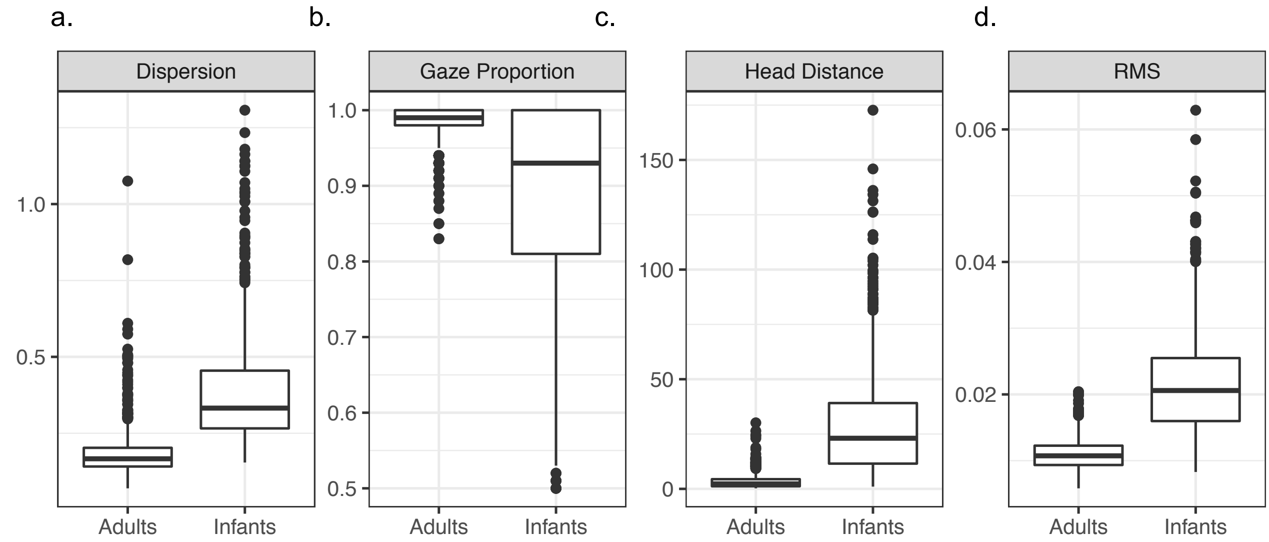


Figure S 3.7: Boxplots for the participant groups in the joint part of the experiment. (a) Dispersion within a fixation. (b) proportion of recorded gaze. (c) mean distance change (mm) between the head and the eye tracking camera after calibration, (d) RMS angular distance of 500Hz eye tracker samples during fixations.

Fixation duration as measured by the eye tracker is shown in Figure S3.8. Note that more very short fixations (< 50ms) were recorded for infants. As planned, these short fixations remained in our analysis because short fixations are seen as possible indicator of lower measuring quality (Wass et al., 2014) and a cutoff can misalign dependent variables (Orquin & Holmqvist, 2017).

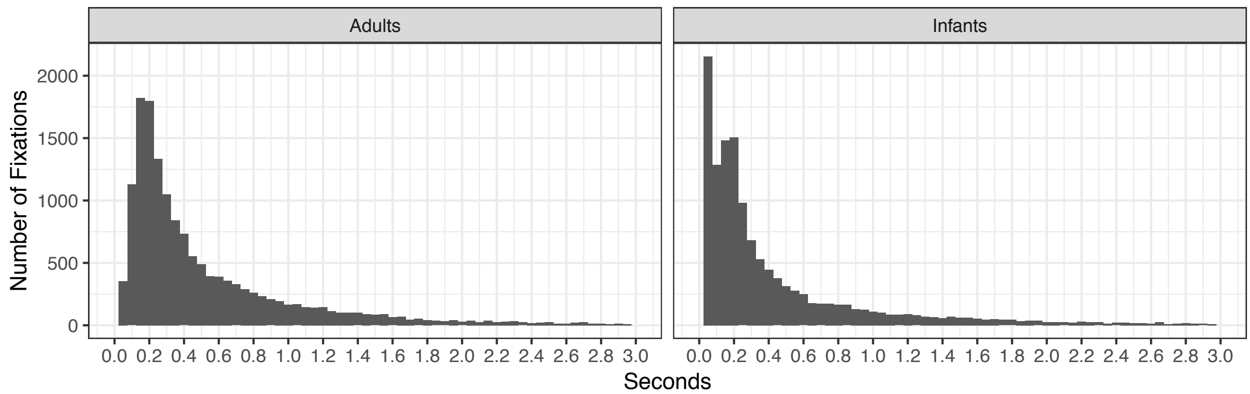


Figure S 3.8: Histogram of fixation duration for adults and infants after exclusion of invalid trials.

3.8.2.2 Calibration and Validation Error and Number of Validations Completed

In the joint part of the experiment, we conducted validations of the five calibrated gaze points after the calibration was successfully completed. For the validation procedure, we used the same calibration target on a differently colored background. Out of the 77 attempts to perform a validation with the infant participants, only one was accepted as good by the eye tracking device. Five validations had moderate error, and 50 were categorized as poor. Twenty-one validations had to be aborted because of inattentiveness and movement of the infant. In 83 cases, we did not attempt the validation because inattentiveness was expected to occur (see Table S3.16). We only proceeded with the trial sequence after a validation if the validation error could be attributed to inattentiveness during validation and not to movement after the calibration procedure. If movement occurred, we recalibrated.

Table S 3.16 Attempts to Validate Calibration Success

	Validation Score [a]					Completed
Infant	Fair	Good	Poor	Aborted	Missing [b]	Sequences [c]
ki02			1	1	1	3
ki03			4	2	2	6
ki04			1	3	2	4
ki05			2		1	3
ki06		1	2	2	3	5
ki07			1	1	5	6
ki08			1		2	3
ki09	2		1		4	5
ki11			1		1	3
ki12			1		5	5
ki13					4	3
ki14			1		3	4
ki15	1		1		2	2
ki16	1		2			3
ki17			6		1	6
ki18			1	3	2	6
ki20			3		2	4
ki21			1	4	1	6
ki22			2		4	5
ki23	1		1		4	6
ki25			1		6	6
ki26			2		7	5
ki27			1	1	5	6
ki28			2		3	5
ki29			1	1	4	4
ki30			3		3	6
ki31			3		3	6
ki32			2	3	1	6
ki33			2	0	2	2
Total	5	1		21	83	134

[a] Good: Errors are generally acceptable. Fair: Errors are moderate, calibration should be improved. Poor: Errors are too high for useful eye tracking (EyeLink® 1000 Plus User Manual, 2015)
[b] Because the infant showed inattentiveness, a validation procedure was not attempted.
[c] Sequences that were starting with a successful calibration and were included in the analysis. The number of sequences is lower than that of the possible validation procedures, because sometimes calibrations and validations had to be repeated to start the trial sequence.

Calibration Targets (Infant and Adult Part of the Experiment)

Table S 3.17. Completed Trials, Means and Standard Deviation of the Accuracy and Precision Scores During the Verification Trials

		Instability[1]				Dispersion[1]		
Group	Stimuli	*N*[2]	*M*	*SD*	Calibration	*N*[2]	*M*	*SD*
Adults	Bullseye	74	0.489	0.204	Bullseye	75	0.175	0.060
	Harp	74	0.455	0.203	Harp	75	0.165	0.046
	Nautilus	73	0.411	0.142	Nautilus	75	0.166	0.050
	Popflake I	73	0.466	0.253	Popflake I	72	0.167	0.043
	Purple	74	0.579	0.258	Purple	72	0.196	0.100
	Medal	73	0.601	0.264	Medal	72	0.163	0.039
Infants	Bullseye	55	0.898	0.327	Bullseye	51	0.350	0.167
	Harp	51	0.852	0.363	Harp	50	0.376	0.199
	Nautilus	51	0.778	0.333	Nautilus	47	0.339	0.114
	Popflake I	52	0.824	0.310	Popflake I	54	0.432	0.239
	Purple	53	0.856	0.325	Purple	59	0.356	0.142
	Medal	56	0.884	0.345	Medal	57	0.394	0.189

Notes: Data was aggregated within trials. *SD* = standard deviation between trials.
[1] Instability was grouped by video stimuli, and the grouping variable for Dispersion was calibration target in the Verification trials. Higher values indicate lower precision or accuracy
[2] Number of trials. Even though the adult sample was smaller, more valid trials were obtained.

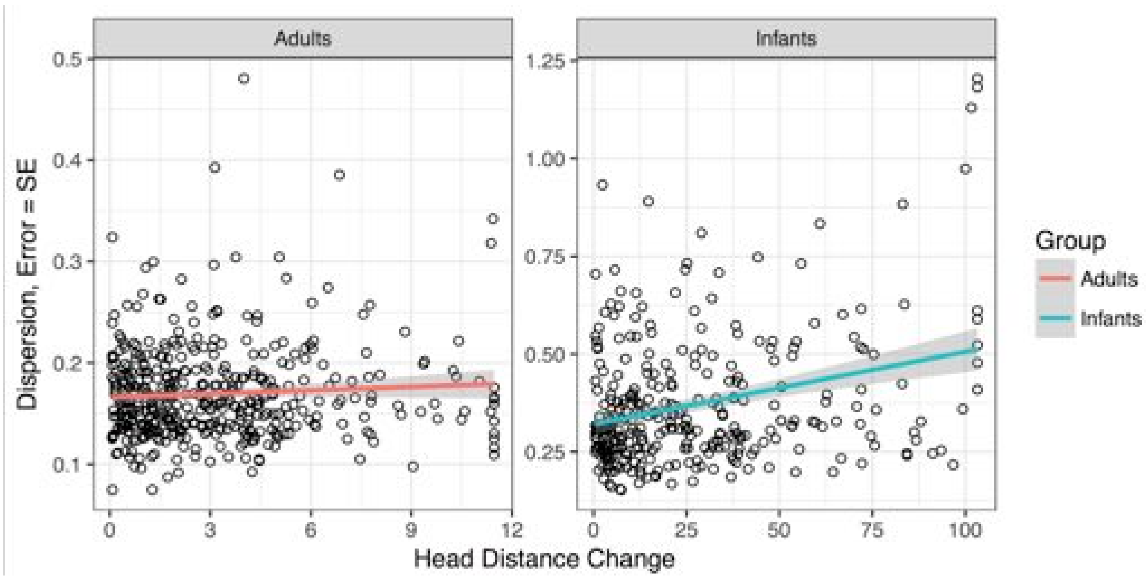


Figure S 3.9: Dispersion as a function of head distance change and participant group in the Verification trials.

Note the different group levels on the x- and y-axis. The interaction of head distance change by participant group was not predictive for the model (F(1) = .04, n.s.).

Table S 3.18 Differences of Gaze Instability between Target Variants during Spread Trials

		High Instability								
		ContrRings			FacetTwist			BlurRings		
Bin[1]	Low Instability	β	*SE*	*p*	β	*SE*	*p*	β	*SE*	*p*
1	Popflake II			n.s.			n.s.			n.s.
	SpiralTwist			n.s.			n.s.			n.s.
	CentBlink	.33°	.094	< .001	.33°	.095	< .001			n.s.
2	Popflake II	.57°	.10	< .001	.35°	.093	< .001	.34°	.093	< .001
	SpiralTwist	.36°	.12	< .01			n.s.			n.s.
	CentBlink	.53°	.11	< .001	.32°	.095	< .001	.31°	.094	< .001
3	Popflake II	.60°	.11	< .001	.50°	.10	< .001	.44°	.098	< .001
	SpiralTwist	.38°	.12	< .01			n.s.			n.s.
	CentBlink	.56°	.11	< .001	.46°	.10	< .001	.40°	.099	< .001
4	Popflake II	.40°	.10	< .001	.36°	.10	< .001	.34°	.10	< .001
	SpiralTwist			n.s.			n.s.			n.s.
	CentBlink			n.s.			n.s.			n.s.
5	Popflake II			n.s.			n.s.			n.s.
	SpiralTwist	.30°	.11	< .01	.33°	.11	< .01			n.s.
	CentBlink			n.s.	.31°	.12	< .01			n.s.
6	Popflake II			n.s.			n.s.			n.s.
	SpiralTwist			n.s.			n.s.			n.s.
	CentBlink			n.s.			n.s.			n.s.

[1] The six bins are covering the following time segments: 0.8 – 1.7 s (Bin 1), 1.7 – 2.55 s (Bin 2), 2.55 – 3.3s (Bin 3), 3.3– 4.15 s (Bin 4), 4.15 – 4.9 s (Bin 5) and 4.9 – 5.7 s (Bin 6). The whole IP started at 800 ms and ended at 5700 ms.

3.8.3 S3 Discussion.

The Influence of the Initial Calibration Procedure vs. Head Movement on Precision.

We suggested, that movement of the infant participants led to noise in the data and a lack of statistical power, so that the effect of an initial calibration target on precision of the subsequent sequence was only marginal. A tentative post-hoc model selection with the adult participants data supports this explanation: there, calibration video as the exclusive predictor variable ($F(5) = 3.6$, $p = .02$), and head - camera distance as random slope led to the best fit. Secondly, in the model that included both participant groups, an increase in head - camera distance after calibration significantly augmented the Dispersion scores. Interindividual variance in distance change that we estimated in the same model as random slope significantly improved the fit compared to a random intercept model ($\chi^2(2) = 44$, p < .001; Figure S3.9). Head - camera distance is recorded by the EyeLink eye-tracking device and included as a variable in the output files. It only provides a one-dimensional distance score, and does not specify the angle of a movement. In the models referring to the blocks that investigated calibration targets, we therefore transformed head - camera distance to only positive values, indicating general body movement which, al least, can be seen as critical for the camera focus. The precision scores of the verification trials confirm that movement during eye tracking has a weakening effect on the detectability of experimental differences.

Even the minimal movements of the adult participants during the Calibration-Repetition trials influenced the accuracy measure Displacement ($F(1) = 22.1$, $p < .001$). In spite of the instructions they received, adults' postures became more and more unstable during this last block of the experiment. A follow up analysis showed that head distance change was positively related to the duration of this task, correlating with $r_{df\,714} = .27$, $t = 7.4$, $p < .001$. Future studies could investigate the usefulness of including changes in head distance or calibration success as control variables during Remote Mode infant eye tracking.

Interestingly, head distance change did not significantly explain variance of the accuracy measure Instability in the same Calibration-Repetition trials. This confirms our plan to assess accuracy during the joint infant - adult trials mainly independent of interfering movements by using Instability as DV.

3.8.3.1 Low Accuracy and Precision During Angular Movement

We argue that low data quality during the Head Movement task of the adult experiment may be caused by a combination of angular change and movements. When looking at Figure 7, one can notice, that head distance change during Head movement stayed in a lower range than during Side movement or the control task Fix. Still, Displacement scores were significantly stronger affected during this task. In the control task Fix, participants tried to keep their upright and frontal head position while involuntary movement occurred. In contrast, during Head movement participants were asked to include head turns in their movements, and during Side movement, they had to turn their head intermittently and might not have returned to the original frontal position as requested. The resulting offset between target position and measured POG for those two movement types indicates that angular changes in addition to movement may be the cause of the proportionally larger offset between POGs and target centers. Together with the elevated RMS scores during the Head movement task, these results indicate that the eye-tracker was especially affected by angular movement. Future studies could investigate the actual causes for this sensitivity.

Chapter 4

Visual Segmentation of Complex Naturalistic Structures in an Infant Eye-tracking Search Task[5]

(Study 3)

Karola Schlegelmilch and Annie E. Wertz

Max Planck Institute for Human Development, Max Planck Research Group Naturalistic Social Cognition, Berlin, Germany.

[5] Manuscript submitted to PLOS ONE

4.1 Abstract

An infant's everyday visual environment is composed of a complex array of entities, some of which are well integrated into their surroundings. Although infants are already sensitive to some categories in their first year of life, it is not clear which visual information supports their detection of meaningful elements within naturalistic scenes. Here we investigated the impact of image characteristics on 8-month-olds' search performance using a gaze contingent eye-tracking search task in which infants had to detect a target patch on a background image. The stimuli consisted of images taken from three categories: vegetation, non-living natural elements (e.g., stones), and manmade artifacts. Our results showed that infants were better able to detect targets belonging to the same category as the background image, particularly if targets were also high in luminance contrasts. Furthermore, larger target-background differences in scaling invariance and entropy, and also stimulus backgrounds including low pictorial depth, predicted better detection performance independent of target luminance. Taken together, these results suggest that infants use a combination of categorical- and property-related information to parse complex visual stimuli and demonstrate that infants' visual processing of naturalistic scenes offers many questions to be investigated in future research.

4.2 Introduction

During their first year of life, human infants explore their visual environment in an increasingly selective manner (Bronson, 1994; Colombo, 2001). Once they start to be able to grasp and crawl during the second half of their first year, their visual attention becomes directed to the characteristics and spatial layout of the objects around them (Courage et al., 2006; Ruff & Rothbart, 1996, 2001), and infants become able to focus on "tasks at hand" (Colombo & Cheatham, 2006). Infants' attentional deployment is then already modulated by some categorical distinctions (e.g., Mandler & McDonough, 1998a; Quinn, 2011). An infant's environment provides diverse visual scenes, in which some entities are well integrated into their surroundings, like books scattered across a child's colorful carpet, or fallen leaves on the playground's sand. However, since the visual system of young children is still developing and experiences with the variability of objects are sparse compared to that of adults, it is not clear which visual information could support a segmentation of static scenes into distinctive elements in early life (for a discussion of this problem see e.g., Kellman, 2001).

By about six months of age, basic low-level visual capabilities have emerged which enable infants to distinguish visual pattern within their environment (Aslin & Smith, 1988; Braddick & Atkinson, 2011; Kellman & Arterberry, 2007). These include grating acuity (i.e., the finest stripes of varying size which can be resolved; e.g., Lewis & Maurer, 2005), contrast-sensitivity at higher spatial frequencies (i.e., more narrow changes between light and dark regions; Brown & Lindsey, 2009; Pirchio et al., 1978; van den Boomen et al., 2012), and orientation (Braddick & Atkinson, 2011; Morrone & Burr, 1986; Slater et al., 1988). These basic functions become more detailed and refined during infancy and early childhood and are adult-like by around 6 years of age (Almoqbel et al., 2017; Leat et al., 2009; Lewis et al., 2007; Siu & Murphy, 2018).

An infant's visual attention can be drawn by strong luminance contrast, complexity, or novelty of a visual stimulus' attributes. Similar to adults, infants are likely to orient towards locations of a visual scene which stick out from their surrounding due to physically intense cues (e.g., Amso et al., 2014; Itti & Koch, 2001; Kwon et al., 2016). Furthermore, visual information which provides learning opportunities is an important component of infants' attention to objects, visual pattern and events (Fiser & Aslin, 2002; Kirkham et al., 2002). The environment provides opportunities to receive, organize, differentiate, and accumulate visual information (Fantz & Nevis, 1967). Infants actively select their learning experiences

(Oudeyer & Smith, 2016), and learning during visual inspection becomes evident, for example, by increased engagement with more complex stimuli (Courage et al., 2006). Experience with such information is necessary for the development of the basic visual functions mentioned above (Lewis & Maurer, 2005) and provides the basis for higher level visual abilities. These include the integration of contour segments (Putzar et al., 2007), and the perception of fine detail (i.e., letter acuity; Maurer & Lewis, 2013), all of which support visual organization of real-world scenes.

However, infants also disengage from a stimulus or scene if it contains overwhelming levels of intensity or complexity (Aslin & Smith, 1988; Bornstein & Benasich, 1986; Ruff & Rothbart, 2001). Thus, an infant's orientation latency is similarly affected by the significance or familiarity of the current and characteristics of the new stimulus (L. B. Cohen, 1972; Oakes et al., 2002).

4.2.1 Visual categorization during infancy

An infant's attention and behavior can also be affected by categorical information. For example, visual input which is related to specific categories signaling threat or ambiguity—including angry faces, snakes, spiders, or plants—causes arousal or reluctance (e.g., C. Elsner & Wertz, 2019; Hoehl et al., 2017; LoBue & Adolph, 2019). Other research takes advantage of infants' interest in social information and investigates their attention to face stimuli among non-social objects using search tasks (for review see: Leppänen, 2016). Together, these studies indicate that reactions of infants in their first year are increasingly driven by the ecological or cultural relevance of a stimulus' category. Many researchers describe categorization during infancy as relying on perceptually formed representations which are continuously refined and conceptually enriched (Eimas, 1994; Madole & Oakes, 1999; Quinn & Eimas, 2000; Rakison & Yermolayeva, 2010; Westermann & Mareschal, 2012; but see Mandler, 2000). Yet, infants are also able to distinguish between objects which are perceptually similar but belong to different domains (B. Elsner et al., 2013; Mandler & McDonough, 1993; Pauen, 2002). Furthermore, infants also distinguish general categories consisting of instances which are heterogeneous in their characteristics—such as plants versus artifacts and non-living natural objects (e.g., C. Elsner & Wertz, 2019; Mandler & McDonough, 1998b; Wertz & Wynn, 2014) or stimuli differing in the animacy–inanimacy domains (B. Elsner et al., 2006; Opfer & Gelman, 2011; Rakison & Poulin-Dubois, 2001).

Category exemplars in these studies are commonly extracted from their background and presented as a single bounded object. There are only few studies which used real-world scenes to investigate infants' ability to distinguish content-related visual information (e.g., scenes with face-targets: Amso et al., 2014; Frank et al., 2014; Kelly et al., 2019). Moreover, given the many reasons why infants might attend to one visual cue but not another—including exploration and visual learning—research interested in the early detection of ecologically significant visual information might profit from the analysis of visual properties inherent in real-world scenes.

4.2.2 Segmentation of real-world scenes

Naturalistic categories differ in their visual properties. Several statistical properties have been identified which are efficiently processed by the adult visual system (Burton & Moorhead, 1987; Field, 1987) and allow instantaneous segmentation and classification of naturalistic scenes or the detection of scene elements (e.g., Geisler et al., 2001; Oliva & Torralba, 2006). For example, types of scenes differ in the distribution of spatial frequencies (Hansen & Hess, 2006; Tolhurst et al., 1992) and their fractal characteristics (e.g., Isherwood et al., 2017; Ruderman, 1997). Furthermore, higher level characteristics such as symmetry, regularity, or curvature determine the discrimination of ecologically meaningful categories in adults (Baumgartner et al., 2013; Long et al., 2017; Schmidt et al., 2017).

Along with grouping of contour elements (e.g., Elder & Goldberg, 2002), texture segregation (i.e., the effortless segregation of texture patches, for a review see: (i.e., the effortless segregation of texture patches, for a review see: Landy & Graham, 2004) is seen as a major mechanism determining the successful visual organization and identification of scene elements (Kellman, 2001; Marr, 1976; Panis et al., 2008). However, when exposed to artificial stimuli, infants within the first year of life are rather sensitive to differences in luminance, whereas segregation of textures defined by discontinuities in orientation emerges only after the end of the first year of life (Sireteanu & Rieth, 1992). Since categorization and learning of visual regularities is already present in younger infants, this raises the question of whether the segregation of naturalistic textures relies on additional features than those tested with artificial stimuli. For example, Balas and colleagues (2018, 2014) showed that by 9 months of age, infants are sensitive to contrasts between the appearances of naturalistic textures and their statistical transformations (Balas et al., 2018; Balas & Woods, 2014). Moreover, infants are surprisingly proficient at processing signals indicating depth and surface properties provided by graphical stimuli (Kavšek, 2003; Kellman & Arterberry, 2007; Yang et al., 2011). This suggests that infants are able to detect some differences in visual properties beyond luminance contrast. Given the reviewed literature on infants' attention to visual information and their sensitivity to ecologically significant categories, we were interested in identifying how both of these qualities contribute to infants' visual segmentation of complex real-world scenes.

4.2.3 The current investigation

In the current study, we conducted a visual search task with 8-month-olds including images of real-world structures to investigate the effect of different image attributes (e.g., visual properties, categories) on scene segmentation (i.e., the detection of a target structure on a background structure). In contrast to frequently used stimuli in studies on categorization and visual development (i.e., faces, objects, or graphics; for overviews see e.g., Kellman & Arterberry, 2007; Quinn, 2011; but: Balas et al., 2018; Balas & Woods, 2014), we included photographs depicting homogeneous assemblies of natural entities and artifacts. Such visual structures characterize an important proportion of the human environment and were found to include visual properties relevant for adult categorization (e.g., Geisler, 2008; Torralba & Oliva, 2003).

We used photographs of three superordinate categories: vegetation, non-living natural elements such as rock or water surfaces, and artifacts. We chose these categories because they cover important aspects of human environments that are of ecological and social significance, and have been so over evolutionary time. For instance, these three categories are part of either a natural or a manmade world (e.g., Gelman, 1988; Schlegelmilch, 2012; Walther & Shen, 2014), they determine the quality and behavioral affordances of a surrounding (e.g., Adelson, 2001; Schuppli et al., 2016; Smuda, 1986), and they can provide organic or mineral material, represent tools, or provide food (e.g., Carrara & Mingardo, 2013; Wertz, 2019). Moreover, infants typically have visual contact with a variety of instances of each of these categories which provide learning opportunities for some of their aspects.

We used an eye-tracking visual search task in which infants had to find a patch of one type of image presented on a discrepant background image. This task allowed us to test if category membership affects detection. Infants received a reward (i.e., a colorful butterfly and sound) when their gaze landed on the visual target patch. The reward was included to stimulate visual search and appeared when the infant fixated on the target patch. By the age of eight months, infants are able to perform eye movements in order to trigger a reward (Wang et al., 2012), and gaze-contingent rewards motivate infants' search in eye-tracking experiments if there is no clear pop-out effect for the target (Hessels et al., 2016; Jones et al., 2014).

To account for the familiarity and perceptual difficulty of the structures depicted in our stimuli, we included variables which quantified the categorization of the same images by preschool children and adults in a previous study of Schlegelmilch and Wertz (2020). We also assessed visual properties selected from research on adult visual categorization of naturalistic entities (e.g., Geisler, 2008; Heaps & Handel, 1999; Schmidt et al., 2017) which either showed processing advantages in adults, or discriminated statistically between the categories used in our study. Finally, we included a measure of luminance contrast. Taken together, the variables we expected to predict infants' search performance belonged to three groups: (a) content-related visual information, (b) structure-related visual properties, (c) low-level salience (see Table 4.6 for descriptions of the variables).

Table 4.6 Definitions of the Visual Properties

Name	Definition	Relevance
	Content-related variables received from children and adults [a]	
Assigned category	Judgments about to which superordinate category each of the images belonged	Provides estimates of the images' generalizability to any of the categories for different age groups.
Perceived similarity	Judgments of visual similarity between the images that were included as target-background image combinations	Subjective perceptual similarity judgments of different age groups.
	Computational [b]	
Luminance	Mean pixel luminance.	The overall lightness or intensity of a structure.
Alpha	Steepness of the distribution of energy across spatial frequencies ($1/f^{alpha}$), referring to the proportion of larger changes to more narrow changes between light and dark.	In natural scenes, alpha values are found to lie in a typical range (e.g., Hansen & Hess, 2006).

Deviation	Deviation of a spatial frequency distribution from the fitted line defined by *Alpha.* Measure of scaling-invariance (fractality) in a structure's spatial changes between light and dark (e.g., Burton & Moorhead, 1987).	This measure was found to differ between artifacts, plants, and natural scenes (Redies et al., 2007)
Entropy	Shannon entropy of pixel luminance values (Shannon, 1948).	Measure of magnitude and predictability of informational content and differentiation
Skew	Skew of the pixel luminance histogram.	Referring to impressions of shading and lighting (e.g., D. Graham et al., 2016; Motoyoshi et al., 2007).
	Rated [c]	
Curvature	*Angular vs. curved.*	Perceived curvature supports classification between animate and inanimate objects (Long et al., 2017; Schmidt et al., 2017).
Regularity	*Regular vs. chaotic.*	Important characteristic for texture discrimination (Heaps & Handel, 1999; Rao & Lohse, 1996).
Symmetry	*Symmetrical vs. asymmetrical.*	Symmetry attracts attention in natural scenes (e.g., Açık et al., 2009).
Depth	*Plane vs. three-dimensional.*	Indicates spatial arrangement of scene elements.

[a] Assessed in sorting tasks with 4–5-year-olds and adults (Schlegelmilch & Wertz, 2020). Assigned categories were included as the Euclidian distance between the target-image's and background-image's proportional assignments to each of the categories. Similarity judgments were transformed to dissimilarity values.
[b] Computational properties were assessed with functions implemented in Matlab (version R2017b) or provided by literature on image processing (Gonzalez & Woods, 2018)
[c] Rated properties were formulated as opposites and judged on a continuous scale by adult participants.

We expected that computational and rated visual properties could influence infants' search performance in two non-exclusive ways: a) their prominence within a background image might hinder the detection of the target; b) stronger differences between a visual property in the target patch and the background image might increase detectability of the target. We therefore included computationally-assessed properties as difference variables by subtracting the background's variance of the property from its target-plus-background's variance (Section 1 in S1 Methods). Visual properties based on human ratings had been assessed for entire images (Schlegelmich & Wertz, 2020), so these ratings could not validly represent the small regions of the images used as target patch. Consequently, we only included the background's rated properties in our analysis.

Our predictions for the current study were that infants would detect a target patch faster (a) if it depicts a category which is distinct from the background category, rather than belonging to the same category, and (b) if its visual properties differ more strongly from the background properties.

4.3 Methods

4.3.1 Participants

The final sample were *N* =39 eight-month-old infants (age: *M* = 8 months, 11 days; range = 8 months, 0 days to 8 months, 29 days; 18 female), recruited from urban and suburban regions of a large European city. We chose 8-month-olds for the current investigation given their successful performance on gaze-contingent search tasks in previous studies (Forssman et al., 2017; Jones et al., 2014; Wang et al., 2012), and early evidence for distinctions between general categories within the second half of the first year of life (Rakison & Yermolayeva, 2010). An additional two infants were tested but excluded because no data could be assessed due to problems with the eye-tracker. All infant participants had normal vision. They were recruited from our internal participant database and tested in the Max Planck Institute for Human Development, Berlin, Germany. The study was approved by the MPIB Ethics Committee and parents gave written consent for their child’s participation. Participants were compensated with 10 Euros and a participation certificate.

4.3.2 Stimuli

The 27 images which comprised the search stimuli of the current study were selected from a set of 60 greyscale images used in the study of Schlegelmilch and Wertz (2020) that investigated the impact of visual properties on categorization in preschool children and adults. The images depict extracts of real-world structures representing one of the three superordinate categories of either vegetation (e.g., foliage, bark, grass), non-living natural elements (e.g., water surfaces, rocks), or artifacts (e.g., cloth, office supplies; for examples see Figure 4.15a, all images are shown in the supporting information (Figure S 4.11a). Each entity occupies the full size of the image. They were photographed by the first author, or downloaded from license-free online image databases.

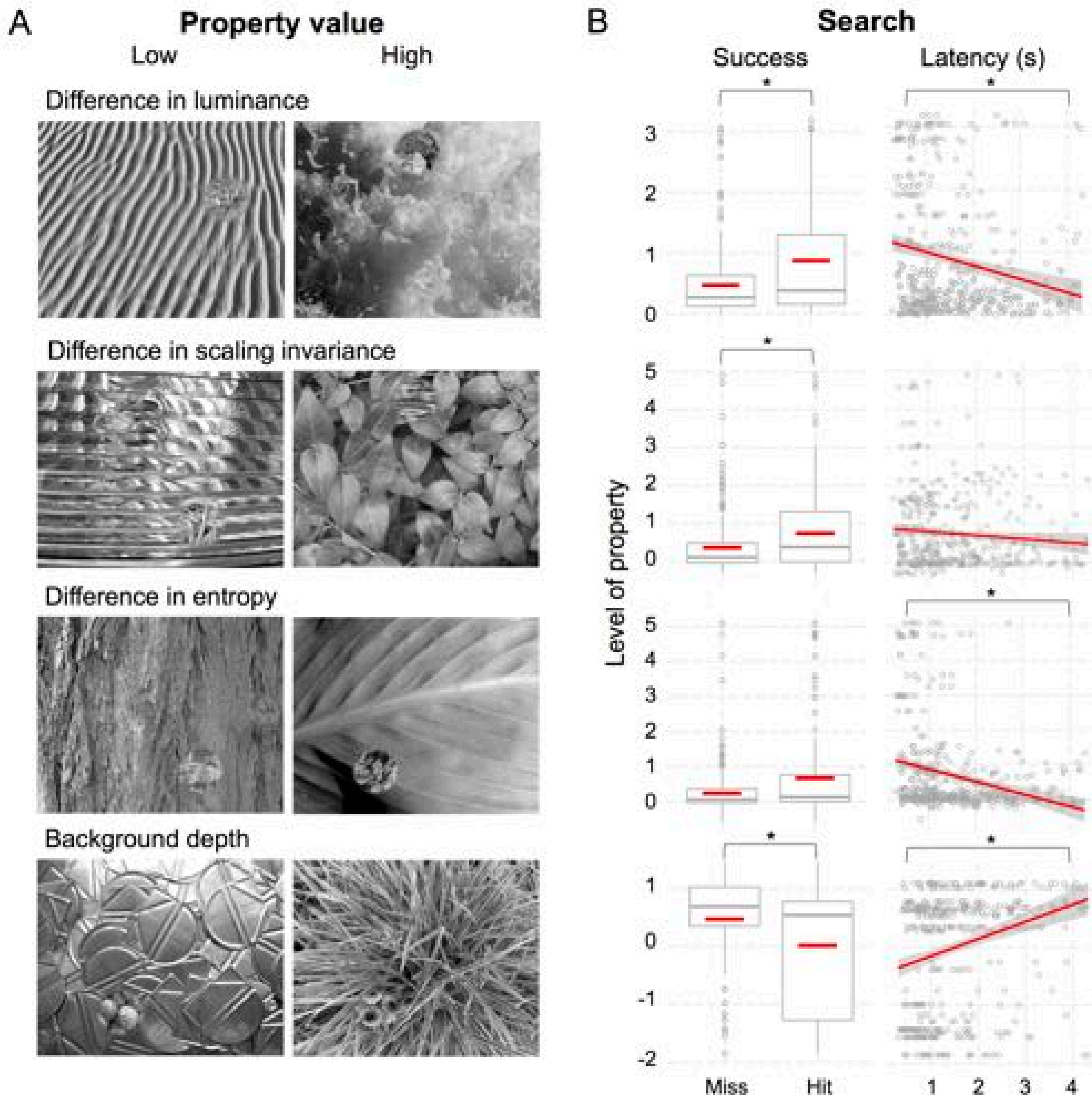


Figure 4.15: Visual properties and their effect on search performance.
(A) Examples of search stimuli with low and high property values, difference in scaling invariance refers to the variable diff_deviation. (B) Properties as functions of search performance. Red lines indicate means (success) or regression lines (latency), error bars: *SE*. Asterisks indicate significant contributions of the variable to the full models, see Table 4.8.

The stimuli for the search task each consisted of one background image into which a circular patch of a different image was inserted as target. The size of the background image measured 1280 × 1024 pixels, leading to 32° × 25.5° of visual angle during presentation (*vis*), the target patch measured 235 pixel (6° *vis*) in diameter. Targets were placed at one of 10 possible locations arranged in a circle with ca. 710 pixel (18° *vis*) distance to the screen center (Figure S 4.10). Each target had a blurred border to prevent that the circular contour of the target was used as cue. With the same intention, a pattern of blurred circles along the outer contours of the 10 possible target locations was included in each background image (for an example, see Figure S 4.10, Figure S 4.11a and b). In order to obtain stimuli with moderate target saliency, we applied a salience algorithm to all possible target-background combinations and locations, using the statistical software Matlab (version R2017b, http://www.mathworks.com). We chose the Graph-Based Visual Saliency algorithm (GBVS; Harel et al., 2007) specified for discontinuities in luminance and orientation. GBVS had reflected infant gaze patterns well (for a discussion of salience applied in infant research see: Kwon et al., 2016). We then chose target-background combinations in which a target was quantified as at least moderately salient, but not as the only salient region of the stimulus (Figure S 4.11c).

Target patches and background images either depicted the same or a different category, leading to stimuli with congruent categories (e.g., vegetation target on vegetation background), or incongruent categories (e.g., artifact target on vegetation background). In the previous study with preschool children and adults, depth cues were an important predictor of categorization decisions. Because depth differed between categories, we balanced the levels of depth in the images we had chosen for the categories, and within the target background combinations. We crossed category-congruency (congruent vs. incongruent) with the control variable depth-congruence (similarly high levels of depth vs. different levels of depth in the target and the background image). This prevented depth being a confound in the analysis of category-congruency.

To make the experiment more engaging for infants, we used three alternating monochromatic colors for the search stimuli. This was done by transforming the greyscale images to HSL color space with the hues: 90° (green), 210° (blue), 330° (red), using the software Adobe Photoshop (Adobe Photoshop CC, Version 2017.0.0). The target and background always shared the same color within a stimulus.

In sum, the target-background combinations of 27 images on 10 possible locations and presented in three different colors led to 260 different stimuli that crossed category congruency and pictorial depth congruency.

4.3.3 Experimental design and procedure

First, a target sticker was placed on the infant's forehead and the infant was seated in a dimmed room in front of the eye-tracker (EyeLink 1000 Plus; SR Research Ltd. 2013 - 2015) either in a baby chair ($N = 37$) with the caregiver right behind, or on their caregiver's lap ($N = 2$). A welcome video was played during the set-up of the eye-tracking camera (EyeLink 1000 Plus High-speed Camera with a 16 mm lens), which was placed approximately 60 cm in front of the target sticker as recommended by the manufacturer (*EyeLink 1000 Plus User Manual*, 2015). Monocular pupil and corneal reflections were assessed in a sampling rate of 500 Hz. The presentation monitor (50” display, with 1280 by 1024 pixel resolution, and 400Hz CMR refresh rate) was set at a distance of 140 cm away from the infants’ eyes to approximately fit the trackable area of 32° *vis* by 26° *vis* in accordance with the manufacturer's suggestion. After set-up, the experimenter stepped behind a curtain from where the infant and caregiver could be monitored on a video screen and started the experiment.

At the beginning of the experiment and after at least each eighth trial, five-point calibrations were conducted with calibration targets alternating in color, form and sound (for a description see Schlegelmilch & Wertz, 2019). The trial started when the average calibration error was below 1° *vis* (see Figure 4.16, and Section 2 in S1 Methods for details).

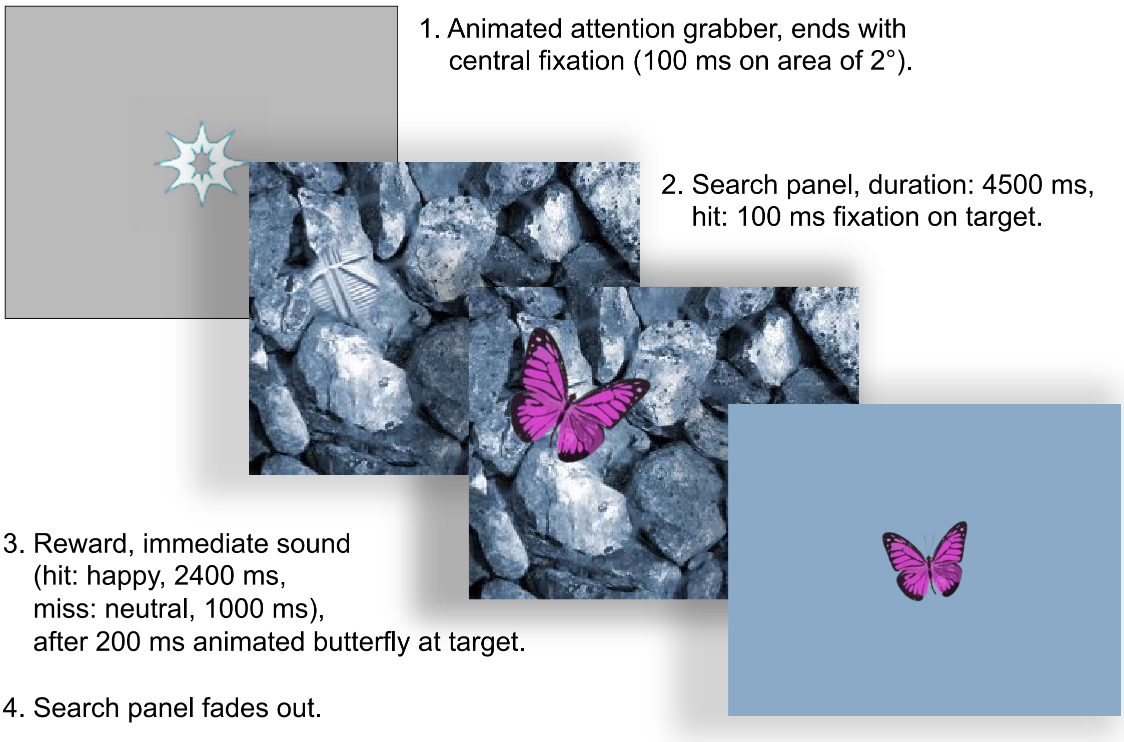


Figure 4.16: **Trial example.**

The first five trials of the experiment were practice trials. They started with an easy-to-detect target patch on a simple background and gradually increased in difficulty. Then, 36 test trials were presented in randomized order. In each trial, the color of the stimulus was randomly altered (green, blue, and red). Additional easy-to-detect practice trials were initiated if it seemed that the infant became unaware of the task after several misses without receiving a reward. If the infant became inattentive, showed fatigue, or if the caregiver requested a break, we paused the experiment for a few minutes, or terminated the experiment prematurely. There were eight versions of the experiment that alternated between participants. Each version included a different selection of 36 target-background image combinations taken from our 260 stimulus variants. No target or background image was included more than twice in one version. To avoid memory effects when an image was repeated, its second occurrence was part of a different target-background image combination and used a different target location and color.

4.4 Results

Infants completed a median of 34 trials (range = 26 to 36 trials), with a median of 88% of gaze recorded by the eye-tracker per trial (range: 1% to 100%). In the following analysis, we included only trials in which infants attended to the stimulus, defined as follows: Trials in which infants detected a target (hit) were accepted if they had a minimum of 80% of recorded gaze. Trials in which a target was not detected (miss) were accepted if they included at least 1240 ms of recorded gaze, which was the median of the hit latency for the whole sample. These criteria led to *Mdn* = 32 valid trials per infant (range = 23 to 36). We did not exclude recorded gaze with low data quality (i.e., fixations with high dispersion of gaze points; Holmqvist et al., 2012), because low-quality fixations for example due to movements can be understood as reactions to the stimuli. Fixation dispersion within all hits was *Mdn* = .17° *vis* (range = 1.03° to 1.3°), whereas in misses it was *Mdn* = .21° *vis*, range = 1.01° to 3.68°. Infants detected a target in 39% of the trials they attended to, range = 17% to 55%.

4.4.1 Statistical analysis of search performance

Was assessed the binarily coded dependent variable (DV) success (hit, miss) and the continuous DV latency, which represented the time until a target was detected if it was a hit. These two DVs covered infants' reactions to aspects of stimulus salience, detection difficulty and background complexity. To account for individual differences, non-normality and unbalanced conditions which are common in infant and eye-tracking data (Kliegl et al., 2011; Valuch et al., 2015), we conducted mixed effect models with the R-package lme4 (Bates et al., 2015). For the generalized linear effect models (GLMM) on the DV success, we used the function glmer and specified a binomial error structure. The units of analysis included as random effects on success were participant, background image, and target location. On the DV latency, linear mixed effect models (LMM) were conducted with the function lmer. For latency, the random effects participant and background image were defined, whereas target location did not improve the model fit and was not included ($\chi^2(1) = 0.04$, *n.s.*). Residual and specification diagnostics were carried out with the R package DHARMa (Hartig, 2020) and by inspection of residual plots. Influential cases were diagnosed with regard to DFBetas (function influence; R-package lme4). The significance of predictors was assessed by comparing the current model with a model reduced by the respective predictor in chi-square likelihood-ratio tests (LRT) with the R-function Anova (package car; Fox & Weisberg, 2019).

To avoid problems of interdependencies between IVs (see e.g., Graham, 2003), we reduced the number of IVs in each comparison by conducting separate models for different research questions (e.g., the impact of computationally assessed visual properties). For these models, we estimated the effect of collinearity by Variance Inflation Factors (VIF; O'brien, 2007) with the function vif (R-package car; Fox & Weisberg, 2019) and only combined IVs in models if VIF values remained below 2.5.

We tested if stimulus color [green, red, blue] generally affected search performance. The comparisons confirmed that color did not predict search performance, success: $\chi^2(2) = 1.6$, *n.s.*; latency: $\chi^2(2) = 3.7$, *n.s.*.

Because movement during remote-mode eye-tracking substantially affects data quality (e.g., Niehorster et al., 2017; Schlegelmilch & Wertz, 2019), we calculated the covariate movement as the maximum of absolute change in head-camera distance within fixations during each presentation of a search stimulus (for details see Section 1 in S2 Results). Movement was included as a covariate in all models.

In our analysis, we were interested in the impact of the predictor variables beyond diff_luminance. We therefore included diff_luminance as fixed effect in all models. We also included the interaction term between diff_luminance and the other predictor variables if it significantly improved the model compared to a model with fixed effects only, as assessed in a LRT (R function anova, package stats; R Core Team, 2019).

4.4.2 The impact of content-related visual information on detection performance

4.4.2.1 Category-congruency

Here, we investigated whether differences between the background category and the target category affected search performance. Category-congruency and depth-congruency were included in the models as predictors. The GLMM on success also included the interaction term between category-congruency and diff_luminance. LRTs on success indicated significant contributions of the fixed effect category-congruency ($\chi^2(1) = 6.2$, p = .013) and the interaction term ($\chi^2(1) = 8.1$, $p = .005$). As a fixed effect, incongruent categories led to a higher probability to detect a target, whereas incongruent categories led to a lower probability to detect a target if combined with higher, but not with lower diff_luminance (Figure 4.17). In the LMM on latency, category-congruency did not contribute to the model, $\chi^2(1) = 3.3$, $p = .068$. The control variable depth-congruency neither improved the model fit on success ($\chi^2(1) = 1.7$, $p = .192$), nor on latency, $\chi^2(1) = 1.7$, $p = .187$, Table 4.7. For an analysis of how category type affected detection of congruent and incongruent category combinations, see S3 Discussion.

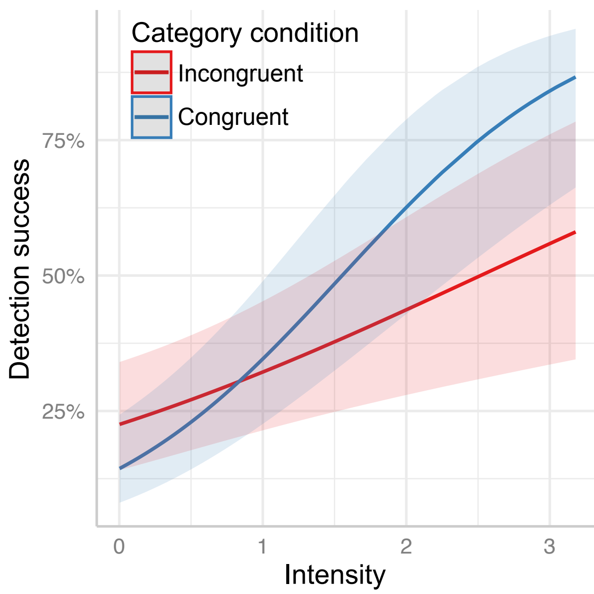


Figure 4.17: Diff_luminance supports the detection of congruent target-background categories.
High diff_luminance is related to higher detection success for congruent categories (blue), compared to incongruent categories (red). The fixed effect found for category-congruency—after including the interaction term between diff_luminance and category-congruency—increased the probability to detect incongruent-category targets by ca. 5%, compared to congruent.

Table 4.7 Category-related Properties Predicting Detection Performance

	GLMM on success				LMM on latency			
Property	Log-Odds	95% *CI*	*z*	*p* [a]	*b* (ms)	95% *CI*	*t*	*p* [a]
	Category-congruency							
Diff_luminance	0.49	[0.18, 0.80]	3.09	.002	-82	[-187, 22]	-1.54	.126
Category-congruency	-0.55	[-0.98, -0.12]	-2.49	.013	-151	[-314, 11]	-1.83	.069
Depth-congruency	0.24	[-0.12, 0.60]	1.31	.192	-121	[-300, 59]	-1.32	.188
Diff_luminance: Category-congruency	0.66	[0.20,1.11]	2.84	.005				
Movement	-2.38	[-2.83, -1.93]	-10.4	< .001	1106	[886, 1326]	9.85	< .001
	Assigned category children							
Diff_luminance	0.68	[0.39, 0.97]	4.64	< .001	-70	[-173, 33]	-1.34	.184
Assigned-category	0.2	[-0.00, 0.41]	1.96	.05	92	[11, 173]	2.22	.027
Image similarity	0.09	[-0.12, 0.29]	0.82	.414	-85	[-169, -1]	-1.99	.047
Diff_luminance: Assigned-category	-0.30	[-052, -0.08]	-2,66	.008				
Diff_luminance: Image similarity	-0.13	[-0.39, 0.14]	-0.92	.356				
Movement	-2.39	[-2.84, -1.94]	-10.4	< .001	1062	[841, 1283]	9.43	< .001
	Assigned category adults							
Diff_luminance	0.69	[0.41, 0.97]	4.83	< .001	-87	[-191, 16]	-1.66	.099
Assigned-category	0.24	[0.03, 0.46]	2.2	.028	68	[-13, 149]	1.65	.101
Image similarity	0.11	[-0.11, 0.32]	0.95	.34	-43	[-131, 44]	-0.97	.331
Diff_luminance: Assigned-category	-0.32	[-0.53, -0.11]	-2.95	.003				
Diff_luminance: Image similarity	-0.05	[-0.30, 0.2]	-0.41	.680				
Movement	-2.37	[-2.82, -1.92]	-10.3	< .001	1101	[881, 1322]	9.79	< .001

Note. See Section 2 in S2 Results for further analysis and discussion.
[a] *P*-values obtained by chi-square likelihood-ratio tests.

4.4.2.2 Previous child and adult categorization variables

To examine whether the category-congruency analysis was influenced by the difficulty of the images, we included variables referring to the distance of category assignments and similarity judgments which were assessed in the previous sorting tasks by children and adults. These variables were analyzed in separate models for the data received from either children or adults. The GLMMs on success also included the interaction terms of category-assignment and similarity-judgment with diff_luminance, respectively. The predictive value of children's image categorization for infants' detection success marginally missed significance as main effect ($\chi^2(1) = 3.8$, $p = .0502$), but contributed as interaction term between diff_luminance and children's category-assignment ($\chi^2(1) = 7.1$, $p = .008$), in that higher, but not lower levels of target diff_luminance related to a better detection success in trials with less distinct category assignments. Moreover, detection latency was predicted by children's category assignments ($\chi^2(1) = 4.9$, $p = .026$), indicating that less distinct category assignments to target and background images decreased the latency to detect the target. Results of children's similarity judgments and adults' variables are provided in Table 4.7.

4.4.3 The Effect of Visual Properties on Detection Performance

Target-background differences in computational properties (i.e., diff_deviation, diff_alpha, diff_entropy, and diff_ skew) and rated background properties (i.e., curvature, depth, regularity and symmetry) were analyzed in separate models. None of the models included interaction terms with diff_luminance. This led to four analyses conducted to assess the impact of visual properties on infants' search performance.

In the GLMM of computational properties on success, diff_luminance contributed to the model fit with $\chi^2(1) = 19.8$, $p < .001$. Of the structure-related predictors, only diff_deviation contributed with ($\chi^2(1) = 22.2$, $p < .001$), in that higher values of both variables lead to a higher probability to detect the target. Latency was predicted by the structure-related property diff_entropy, which contributed to the fit of the LMM with $\chi^2(1) = 8.5$, $p = .004$. Stronger target-background differences in diff_entropy led to a faster detection of the targets (see Figure 4.15 and Table 4.8).

In the GLMMs on success including the rated background properties, diff_luminance contributed to the model with $\chi^2(1) = 27.9$, $p < .001$ and depth affected the model fit with ($\chi^2(1) = 4$, $p = .046$), in that higher values of diff_luminance, but lower values of depth, lead to a higher probability that a target was detected. Depth also contributed to the fit of the LMM on latency ($\chi^2(1) = 11.1$, $p < .001$), with higher values of depth leading to longer detection latencies.

LRTs indicated that no other visual properties affected detection performance, see Table 4.8 for all results and Figure 4.15 for stimuli examples of significant properties.

Table 4.8 Structure-related Properties Predicting Detection Performance

	GLMM on success				**LMM on latency**			
Property	Log-Odds	95% *CI*	*z*	*p* [a]	*b* (ms)	95% *CI*	*t*	*p* [a]
	Computational target-background difference [b]							
Diff_luminance	0.62	[0.35, 0.90]	4.45	< .001	-42	[-148, 65]	-0.76	.446
Diff_alpha	-0.14	[-0.33, 0.05]	-1.4	.160	-2	[-93, 90]	-0.04	.968
Diff_deviation	0.55	[0.32, 0.78]	4.71	< .001	-37	[-124, 49]	-0.84	.402
Diff_entropy	0.19	[-0.06, 0.45]	1.5	.134	-128	[-214, -42]	-2.92	.004
Diff_skew	0.09	[-0.1, 0.28]	0.92	.356	-42	[-118, 34]	-1.08	.283
Movement	-2.4	[-2.85, -1.95]	-10.37	< .001	1071	[850, 1292]	9.50	< .001
	Rated background property							
Diff_luminance	0.70	[0.44, 0.96]	5.28	< .001	-95	[-189, -1]	-1.97	.054
Curvature	0.07	[-0.2, 0.34]	0.51	.611	-11	[-117, 95]	-0.2	.844
Depth	-0.31	[-0.61, -0.01]	-2	.046	187	[77, 297]	3.38	.004
Regularity	0.19	[-0.22, 0.6]	0.9	.37	-66	[-220, 88]	-0.84	.412
Symmetry	0.11	[-0.28, 0.5]	0.54	.59	128	[-20, 276]	1.7	.106
Movement	-2.38	[-2.82, -1.93]	-10.47	< .001	1078	[857, 1298]	9.58	< .001

Note. Visual properties were included together with the covariate movement as fixed effects.
[a] *P*-values obtained by chi-square likelihood-ratio tests.
[b] Assessed as difference between the properties' variance within the background image alone and within the background including the target patch, see Section 1 in S1 Methods.

4.4.4 Were Targets Detected by Coincidence?

In order to investigate whether infants may have fixated on the targets by coincidence, we compared the number of fixations during the presentation of the search-stimulus on each of the 10 possible target locations without a target to the number of fixations on the target. A GLMM specified for count data on the numbers of fixations with the predictor location (the target contrasted to the 10 possible empty target locations) and participant as random effect indicated that there were less fixations to any empty target location than to the target itself, LRT on the IV location ($\chi^2(10) = 599$, $p < .001$), all contrasts $p < .001$. These results indicate that overall, target detection was non-accidental.

4.5 Discussion

Here, we investigated which image characteristics (i.e., category, diff_luminance, structure-related visual property) affected 8-month-olds' ability to detect a discrepant image patch on a complex background image using a gaze-contingent eye-tracking search task. The images depicted one of the three superordinate categories: vegetation, non-living natural elements or artifacts.

Our results indicate that infants attended to combinations of category- and property-related cues to distinguish complex naturalistic patterns. Although—consistent with the previous literature (e.g., Amso et al., 2014)—detection performance was affected to a large extent by diff_luminance, we found that a target's categorical information impacted detection performance if it was supported by a larger target-background difference in luminance. Structure-related visual properties of the images, such as area, entropy and rated depth, were less affected by luminance, and predicted detection performance independently.

In the current study, targets were detected non-accidentally, indicating that infants learned to search for the targets and were able to direct their attention to discontinuities in the appearance of the structures. The current findings differ from earlier eye-tracking search tasks with infants (e.g., Hessels et al., 2016; Kelly et al., 2019) in that targets did not represent a delimited object. Instead, photographs of complex naturalistic surfaces or assemblies of elements alternated as targets and backgrounds. A target was only defined by it being a discrepant structure patch to the background and by leading to a visual reward if looked at. Therefore, our results are relevant for research on image segmentation and visual search (e.g., Aslin, 2011; Bhatt & Quinn, 2011; Sireteanu et al., 2009).

4.5.1 Content-related Information Affected Detection Latency

In contrast to our expectations, we found that the probability of detecting a high- -target was higher if the target belonged to a category which was congruent to the background category, and detection latency was marginally shorter for congruent category combinations. Similarly, less distinct category assignments to the target-background image combinations—as assessed with preschoolers in a recent study by Schlegelmilch and Wertz (2020)—predicted shorter detection latencies in the current experiment. It was only when the interaction terms between diff_luminance and category-related properties were included that detection success became slightly facilitated for more distinct category combinations. We interpret this as a side effect of the strong facilitation of detection success by high diff_luminance for congruent categories (see Figure 4.17, and Section 2 in S2 Results for further discussion). The findings for shorter detection latencies for less distinct categories are difficult to relate to research suggesting that the segmentation of naturalistic scenes in adults is closely linked to categorical information (e.g., Geisler, 2008), because one might expect facilitated segmentation for categories as distinct as those used in the current study.

One explanation for the current result could be that switching attention from one category to another category might have been cognitively effortful for the infants. Elsner and collaborators (2013) came to a similar conclusion in an oddball paradigm with 7-month-olds. EEG recordings revealed that a discrepant-category oddball required more processing capacity compared to processing the same-category oddball (B. Elsner et al., 2013). One can argue that the presentation of single items in an oddball paradigm differs from the current study in that detection of an item is not necessary in the oddball-paradigm. Still, with the current search stimuli, processing difficulty of a target caused by a distinct category might affect covert attention so that saccades are not initiated as easily by the more distinct target category.

Neuroscientific evidence for the existence of category-selective regions in infants provides a further explanation for the delayed detection of more distinct target categories (e.g., Cabral et al., 2019; Deen et al., 2017; Farzin et al., 2012; Kriegeskorte et al., 2008). If category-selective regions associated with the background image are activated, detecting a distinct target category would make a change of the neural activation pattern necessary. The infant's behavioral response (i.e., orientation of gaze) might then also reflect the distance between the neural representations of the distinct categories (Carlson et al., 2014). Distinct locations within the visual processing stream can be presumed for the categories included in the current study (Beeck et al., 2019). The categories used in the current study belong to domains frequently reported in adult neuro-imaging research (e.g., artificial vs. natural objects; Cichy et al., 2014; Haxby et al., 2014). However, category selective brain regions mature into childhood and beyond (M. A. Cohen et al., 2019; Dekker et al., 2011; Gogtay et al., 2004) and, in particular, neural specificity associated with the recognition of human-made objects and tools show protracted development (Cabral et al., 2019; Deen et al., 2017; Farzin et al., 2012), so that detection latency might not only be affected by the distance of the representations, but also by their developmental state.

Post-hoc comparisons of detection latency for all target-background combinations of the images' true categories revealed that artifacts led to longer detection latencies compared to natural categories (i.e., natural elements and vegetation), independent of their being target or background. In contrast, natural categories did not lead to different latencies (see S3 for further discussion and results). Therefore, our results suggest that detection latency for category-related target-background differences in complex visual scenes in infancy is driven by the distance between internal representations of natural and human-made entities (e.g., Carlson et al., 2014), or possibly the later maturation of category-selective brain regions related to artifacts (e.g., Cabral et al., 2019; Deen et al., 2017).

In summary, the facilitated detection of less distinct target categories found in the present investigation might rely on several non-exclusive explanations. These are (i) infants' tendency to detect targets which differed from the background in their level of luminance, supporting sensitivity to category-related image properties, (ii) extra cognitive effort when disengaging from one category and attending to an incongruent category, (iii) the distance or developmental state of neural areas associated with internal representations. Yet, further investigations are necessary to obtain better insight into the current pattern of results (for further discussion of this problem see S2 Results and S3 Discussion).

4.5.2 Structure-related Visual Properties Affected Detection Performance

Target-background differences in area and entropy explained a similar or even higher amount of variance in detection performance than diff_luminance ($R^2_{marginal} = .042$ for diff_deviation vs. $R^2_{marginal} = .019$ for diff_luminance on success; $R^2_{marginal} = .033$ for diff_entropy vs. $R^2_{marginal} = .003$ for diff_luminance on latency; R-function *r.squaredGLMM*, package MuMIn; Bartoń, 2019). This is intriguing because infants' attention at this age is strongly affected by luminance contrast (e.g., Amso et al., 2014). One may ask which aspects of these particular visual properties facilitated infants' detection of the discrepant target structure.

It is possible that differences in the amount of grey-level shades, and in the amount of spatial scales that vary in the properties area and entropy, affected infants' detection performance. A structure defined by high area values (i.e., low scaling invariance) is dominated by only some spatial frequency scales, providing similarly shaded regions of repetitive sizes, whereas a structure defined by low area includes all spatial frequency scales. Structures with high values of entropy include similar numbers of each of the 256 possible luminance values of an 8 bit image, while in low entropy high proportions of some luminance levels lead to less differentiated shading or more monotonous structure regions. Accordingly, higher values of the difference-variables area and entropy—which facilitated infants' target detection—were related to disparities between image regions with more fine-grained, cluttered patterns and differentiated contrasts, versus less detailed, smoother, or repetitive regions. Consequently, infants must have been sensitive to these disparities. This is interesting, because one would expect that infants' immature processing of fine detail and their lower contrast sensitivity would lead to uncertainty and make it difficult for them to detect variability in visual structures that differ in these respective non-matured visual aspects. Such sensitivity to uncertain visual information might nevertheless be beneficial for the infant because it can support strategic behavioral reactions such as avoidance, further exploration, or social referencing (e.g., C. Elsner & Wertz, 2019; Pauen & Hoehl, 2015). Sensitivity to uncertain or deficient visual information might also be reflected in infants' novelty preference (Fantz & Nevis, 1967) and in young children's choice of actions that resolve the greatest amount of uncertainty (Köster et al., 2020; Oudeyer & Smith, 2016; Vygotsky, 1978). Perhaps due to these adaptive behaviors, infants were sensitive to visual properties that varied in their levels of uncertainty and difficulty, and attended to discrepancies within these properties. This would explain how high values of the difference variables area and entropy led infants to detect the discrepant target patch.

However, one may then ask why other difference variables of computational properties did not affect target detection performance. For example, alpha represents statistical aspects of naturalistic scenes, and different levels of alpha have even been found to affect adults' processing speed, recognition, and visual memory (e.g., Hansen & Hess, 2006; Redies et al., 2007; Ruderman, 1997; White et al., 2008). Yet, processing advantages of certain alpha levels during similarity judgments had not been found in younger children (Ellemberg et al., 2012), and in the current study, alpha did not predict infants' detection performance (Table 4.8). It is possible that variations in alpha are much more difficult to perceive, because different levels of alpha vary in the *proportions* of spatial scales, but not in the *range* of the spatial scales included in the structure. In contrast to area, the discrimination of alpha levels makes sensitivity to the full spectrum of spatial scales necessary (Ellemberg et al., 2012).

4.5.3 Depth Cues, But Not Shape Predicted Detection Performance

We found that high rated depth of the background image increased target detection difficulty. The dark regions and high contrast contours which are typical for shading-defined pictorial depth might have diverted infants' gaze and complicated the detection of the target. Infants start to be sensitive to stereoscopic depth and to pictorial depth within a few months after birth (Kavšek et al., 2012; Kellman & Arterberry, 2007). However, a photographic representation of complex three-dimensional arrangements might challenge an infant's perceptual abilities, and could potentially lead to either disengagement from the task or the search for further information (Courage et al., 2006; Ruff & Rothbart, 2001). Thus, an alternative explanation is possible: Infants did not disengage from the depth cues because spatial characteristics of scene elements or their arrangement provided opportunities for further visual exploration (e.g., Bertenthal, 1996) and attentional learning processes (Colombo & Cheatham, 2006; Courage et al., 2006) that were more rewarding than searching for the target.

Interestingly, none of the other rated properties—which all referred to shape characteristics or their arrangement—affected infants' target detection performance. In contrast to computational target-background differences, rated properties were only assessed for the background image but did not quantify the distinctness between target and background. If infants "understood" the task, they could have ignored background properties—as long as these did not interfere with their search by causing difficulty or offering more rewarding (explorative) opportunities. One might argue that the high property levels did not necessarily provide such difficulty or reward, because they were as interesting or difficult for the infant as the opposite low levels, so that all levels of the property similarly affected attention. Yet, this explanation is unlikely, because (i) curved shape is preferred to angular shape very early in life (Fantz & Nevis, 1967), (ii) symmetry is processed in a basic way by 1-year-olds and could attract attention because it provides learning opportunities (Bornstein et al., 1981), and (ii) repeated elements which represent regularity are reliably used as backgrounds in infant search paradigms (e.g., Hessels et al., 2016). As with pictorial depth, these findings were obtained with graphical stimuli. Still, their presence in the complex naturalistic structures included in the current study did not affect attention as much as depth cues did.

The stronger impact of rated depth compared to shape characteristics on search performance can have further reasons: Infants' attention might have been affected by characteristics that are important for segmentation processes. Rated shape, which refers to two-dimensional orientation and form characteristics within the image structure, may not have been as relevant as depth cues for this segmentation purpose. In real-world settings, spatial arrangements and object shape can be explored by actions such as self-generated motion which provides parallax (Atkinson & Braddick, 2013; Kellman & Spelke, 1983). Therefore, visual processing and explorative actions that resolve visual ambiguities are suggested to be closely linked during the first year of life (e.g., Bertenthal, 1996; Hoffman & Singh, 2012; Kovács, 2000). This is also supported by an early dominance of the dorsal pathway (associated with the action related "where" and "how" of vision; Kravitz et al., 2011; Milner & Goodale, 2008) compared to the ventral pathway (associated with the "what"; Atkinson & Braddick, 2013; Colombo & Cheatham, 2006; Hammarrenger et al., 2003). Communication between the dorsal and the ventral visual pathway continues to develop beyond the first postnatal year (Ruff & Rothbart, 2001). Thus, visual categorization of scene elements in infancy that affords segmentation ability most likely activates spatial and action-related visual processing mechanisms, also if it relies on characteristics of the elements' physical shape (e.g., Rosch et al., 1976; see also: Anderson et al., 2013; Bertenthal, 1996; Gibson, 2000).

In contrast to rated depth, the control variable depth-congruency did not predict infants' detection performance. Recall that this variable stated if the level of rated depth in the target image was the same as in the background image, or differed from it. The lack of depth-congruency's predictive power might be explained by an insufficient representation of the depth rating in the target image patch, which was originally made for the entire image of which the target was only one small part.

4.6 Limitations and future questions

It cannot be ruled out that infants used cues to detect the targets beyond those analyzed. For example, they might have learned to associate rewards with round areas of a certain size—despite our efforts to blur the contours of the targets. This might have led them to preferably fixate round salient patches, leading to faster detection if the round patch actually included the target. However, we think that such cues did not strongly alter search performance, since otherwise, rated background curvature would have affected detection significantly, and it did not.

There are several important and promising issues which were out of the scope of the current study, but should be included in future experiments. For example, it would be very useful to further investigate infants' gaze and fixation locations during their search on naturalistic structures. Gaze paths during misses and hit-trials are of great interest, because they might distinguish between movements related to distress or exploration. Additionally, the analysis of fixation locations and durations can offer information about visual characteristics which were preferentially looked at.

In future studies, it might also be useful to add psychophysical search stimuli to the experiment that vary in contrast and fine detail. This would allow to better uncover the relation between the individual infant's low-level visual abilities and the impact of content- or structure-related properties on scene segmentation.

Finally, a search experiment conducted with infants, preschoolers and adults would provide valuable insight into the salience of image characteristics, and how they vary between the age groups. This kind of developmental study could also establish whether the faster within-category compared to between-category target detection is specific to infants or young children.

4.7 Conclusion

The current study revealed that 8-month-olds relied on combinations of luminance contrast, content-related, and structure-related visual properties when searching for discontinuities in photographs depicting naturalistic surfaces or assemblies of elements. Infants were able to learn to search for a discrepant target image patch solely defined by a gaze-contingent reward. The results suggest that infants' gaze was largely guided by visual properties relevant for the current task (i.e., search) and by opportunities for exploration, whereas properties related to the analysis of two-dimensional shape characteristics or details seemed to have less of an impact. These results allow insight into the kind of visual strategies infants employ when confronted with complex visual surroundings. Applied to different age groups, such a paradigm can complement research on adult visual processing of categories as much as on the development of scene perception and categorization.

In congruence with other studies using naturalistic images in controlled laboratory settings (e.g., Balas et al., 2018; Frank et al., 2014; Kelly et al., 2019) we argue that the inclusion of naturalistic images in infant vision research is important and might lead to different results than research with artificial objects or graphic stimuli.

4.8 Acknowledgements

During this research, Karola Schlegelmilch was a pre-doctoral fellow of the International Max Planck Research School on the Life Course (LIFE, www.imprs-life.mpg.de). We thank our participants and their parents, Janek Stahlberg and the members of the Max Planck Research Group Naturalistic Social Cognition for their assistance. We also thank the team of the Department of Developmental Psychology: Infancy and Childhood at the University of Zurich for their support.

4.9 Supporting information

Figure S 4.10**. The arrangement of target locations and blurred contours.**

Figure S 4.11**. Creation of search stimuli from 27 images**.

4.9.1 S1 Methods.

4.9.1.1 Calculation of difference-variables of the computational properties.

4.9.1.2 Calibration procedure and trial sequence.

4.9.2 S2 Results.

4.9.2.1 The effect of movement on detection performance.

4.9.2.2 The impact of category congruency on detection performance.

Table S 4.19 Differences in diff_luminance between more or less alike category combinations.

4.9.3 S3 Discussion. Detection difficulty of individual categories.

Figure S 4.12 Target and background categories affecting detection latency.

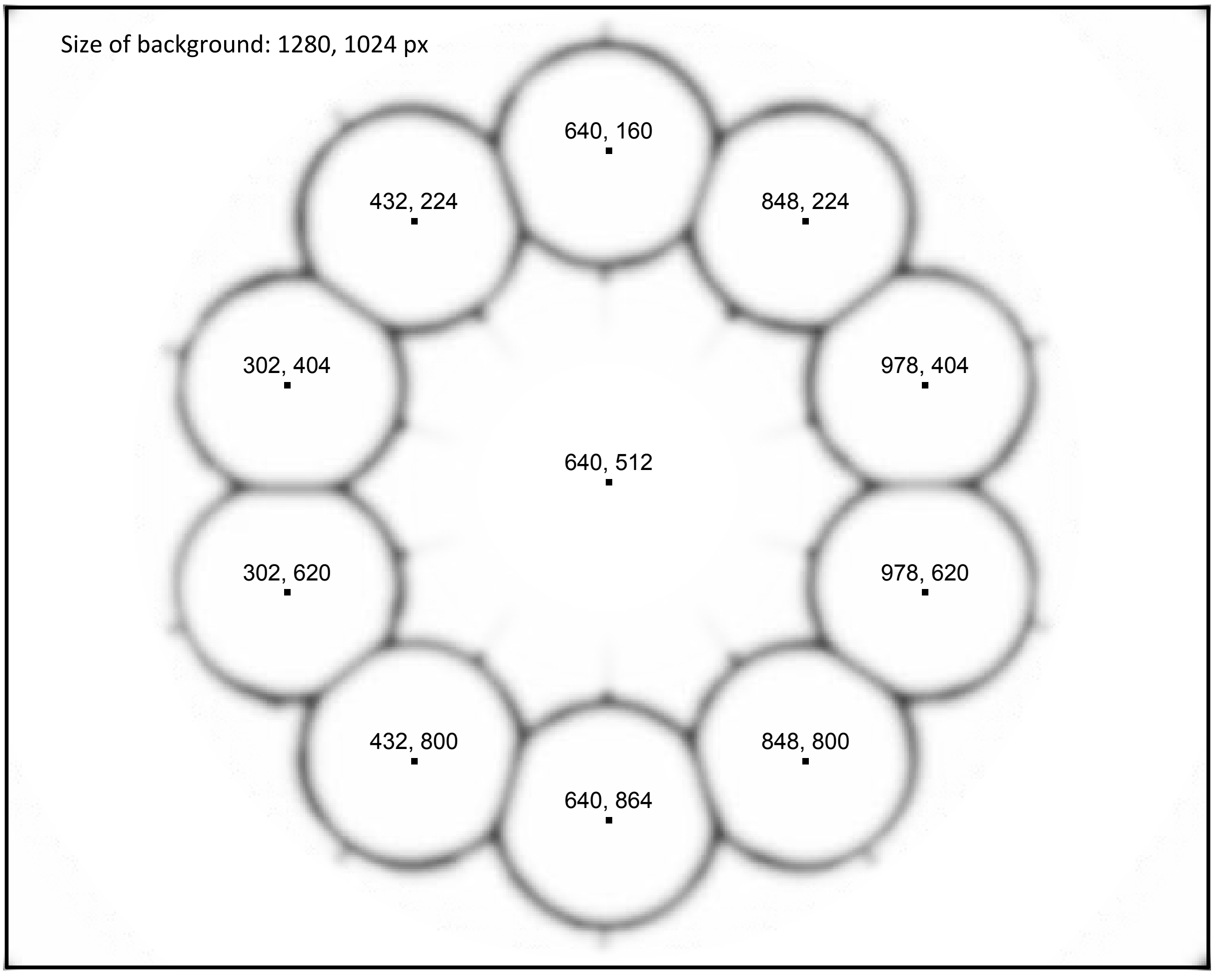
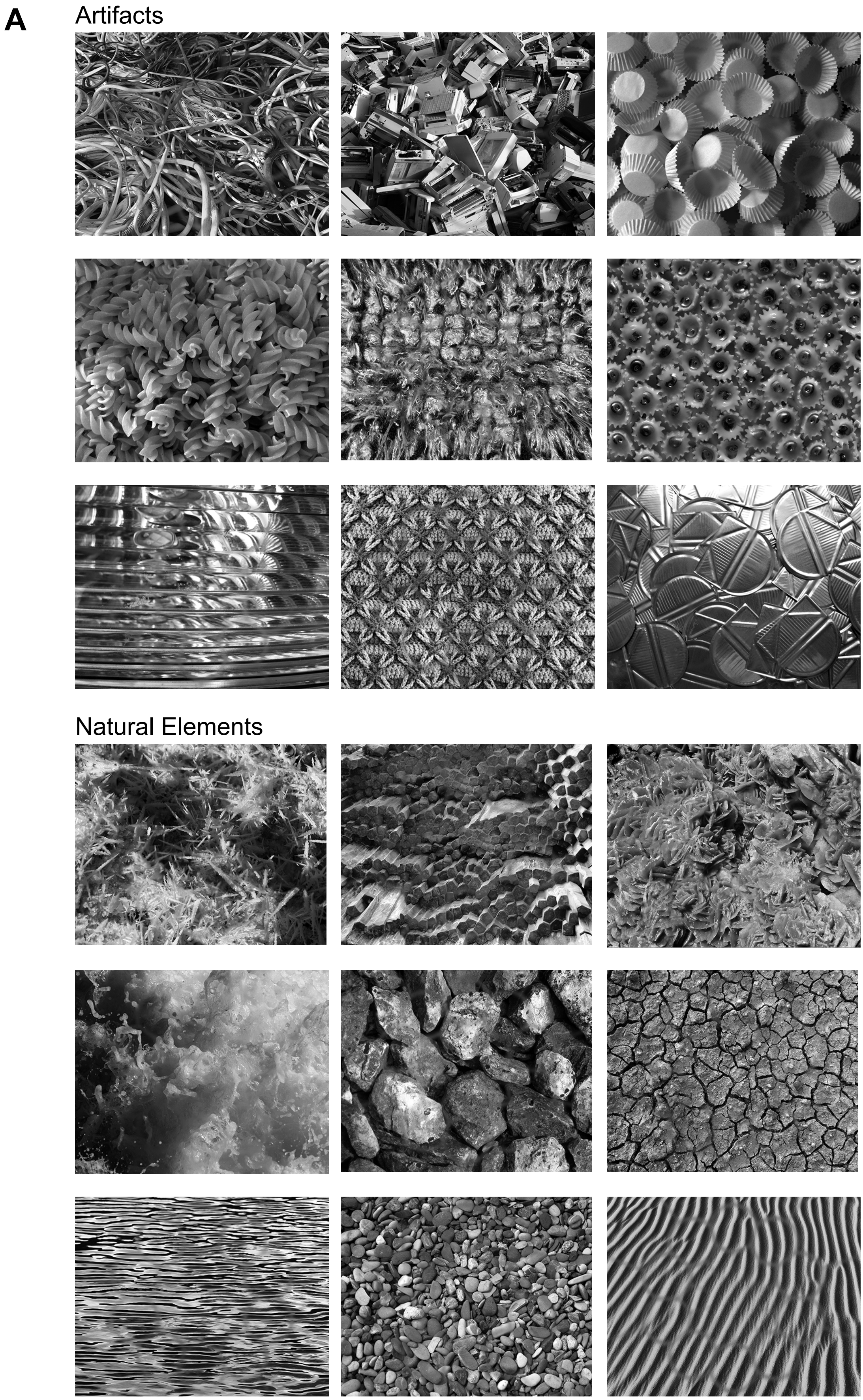
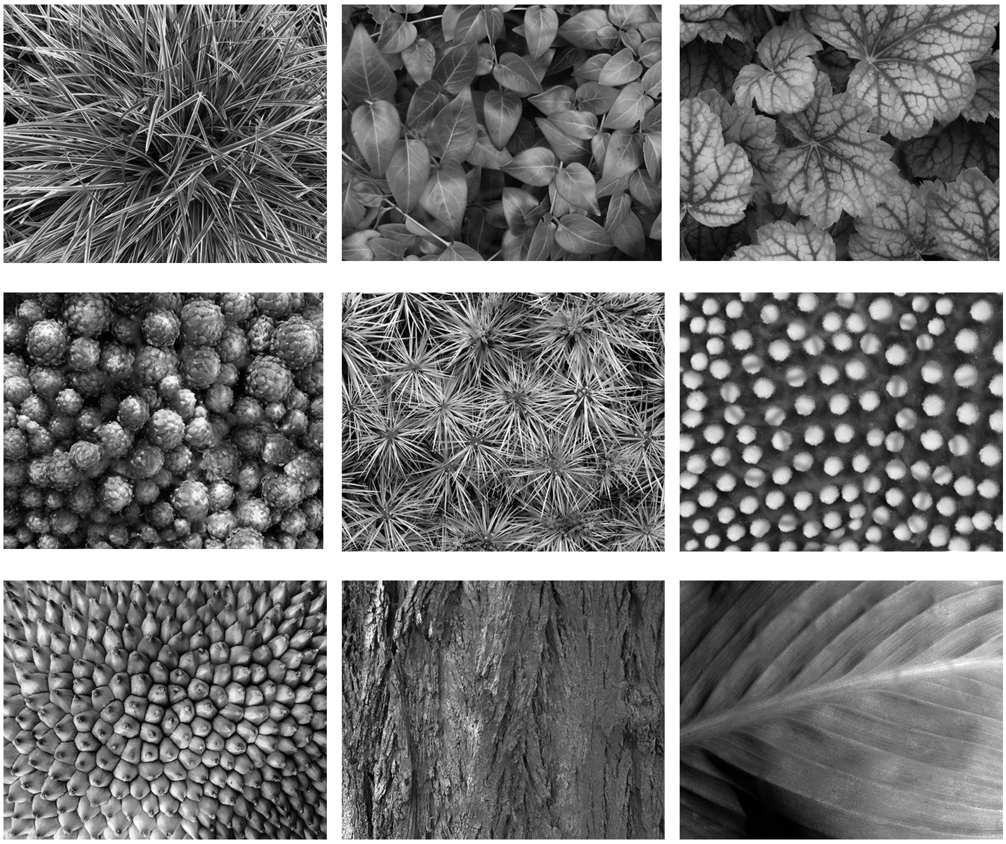


Figure S 4.10: The arrangement of target locations and blurred contours.
Central pixel coordinates of the ten possible target locations, in equal distance to the screen center. The contours around all of the target locations—arranged as circles in a ring— were blurred on each background image, and the target patch was included in one of the circles, respectively.


A
Artifacts
Natural Elements

Vegetation



B

C

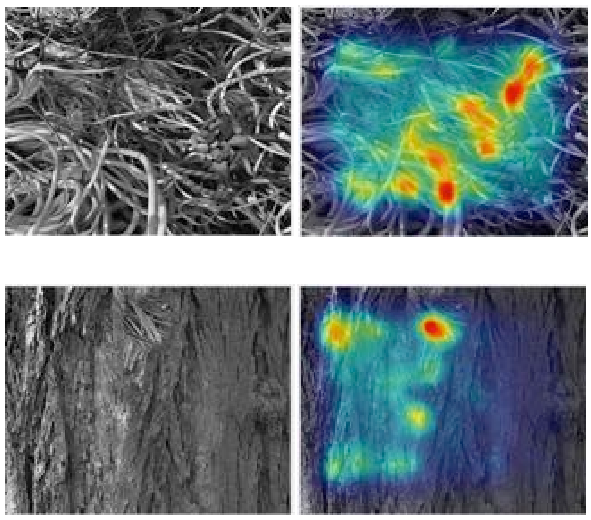


Figure S 4.11: Creation of search stimuli from 27 images

(A) All 27 images of the study are grouped by category in the format in which they were used as backgrounds. Within category they are arranged line-wise left to right according to decreasing levels of rated depth. Backgrounds include blurred contours of possible target locations. (B) Five target patches were sampled from each of the same 27 images. (C) Stimuli with moderate target salience were created by placing different target samples at each of the 10 locations and comparing the results with a salience algorithm (GBVS*;* Harel et al., 2014). Targets which were salient (indicated in the two examples as orange to red overlay) without being the only salient region of the stimulus at the respective location were selected.

4.9.1 S1 Methods.

4.9.1.1 Calculation of Difference-variables of the Computational Properties

We calculated target-background differences of computationally-assessed properties in the following way: We first partitioned the background images without the target patch as well as the background image including the target patch into squares (size = 256 px by 256 px) which fitted the size of the target patch. Next, we calculated a property's variance a) between the partitions of the background, and b) between the partitions of the background including the target patch. Finally, we subtracted the background's variance from the target-plus-background's variance. This procedure was applied to the 260 stimuli covering all target-background image combinations and their target locations included in the study. The obtained difference variables represented the impact of a target property on the variability of this property in the whole stimulus (termed diff_**property-name**). High values of difference variables were obtained if a target exhibited very high or very low levels of a property, which exceeded the range of the background-image's property-levels in either direction. Low values of difference variables resulted from backgrounds in which the levels of a property varied between high and low extremes, so that the property level of the target could not substantially increase the background's variance. Diff_luminance was assessed in the same way with the only difference that we transformed it to its absolute value. If an infant's detection performance was predicted by a difference variable, the infant must have been sensitive to discontinuities of this property—either within the background image (background difficulty) or between background and target (detection facilitation). Note that the impact of a difference variable does not provide information about an infant's respective sensitivity for particular high or low levels of the property.

4.9.1.2 Calibration Procedure and Trial Sequence

At the beginning of the experiment and after at least each eighth trial, five-point calibrations were conducted. Calibrations were accepted if the average error was below 1° *vis*. Each trial included a central attention grabber of 5° *vis* in diameter. As soon as the infant's gaze rested on its central area (2.5° *vis* in diameter) for 100 ms, one of the search stimuli (i.e., a target-background image combination) was shown for a maximum of 4500 ms. If the infant's gaze rested on the target patch before timeout for at least 100 ms, a rewarding music started to play, then a colorful butterfly loomed out of the target's center, and moved to the center of the screen. If a target was missed, the butterfly was only shown for a shorter time, accompanied by a neutral sound. Directly after the butterflies disappeared, a new trial started. Every fifth trial, an attention grabber was shown at a peripheral location in addition to the central location. If the infant's gaze was not recorded within the central region of the attention grabber—for example because of inattentiveness or changes in the distance between the eyes and the eye-tracking camera—an additional calibration was initiated and the camera set-up was corrected if necessary.

4.9.2 S2 Results

4.9.2.1 The Effect of Movement on Detection Performance

Movement of unconstrained participants during remote-mode eye-tracking substantially affects data quality (e.g., Hessels, Andersson, et al., 2015; Niehorster et al., 2017; Schlegelmilch & Wertz, 2019). However, infants' movements might reflect reactions to the complexity or difficulty of a visual stimulus. We therefore assessed the variable *movement* on the basis of values provided by the eye-tracker, calculated as the maximum of absolute change in head-camera distance within fixations during the presentation of the search stimulus. Movement alone predicted success with $\chi^2(1) = 112$, $p < .001$, in that more movement related to a lower probability to detect the target, logit = -2.41, 95% CI = [-2.86, -1.97]. In the LMM of movement on latency however we diagnosed a skewed error structure, which we corrected by reducing values of movement during hit-trials larger than the 99% percentile ($N = 4$) to the value of the 99% percentile. After correction, movement contributed to the LMM on latency ($\chi^2(1) = 97$, $p < .001$) with stronger movement predicting a longer time to detect a target, $\beta = 1.16$, 95% $CI = [0.93, 1.39]$. We included movement as a covariate in all models, and outlier correction of movement was applied to all models conducted on latency.

4.9.2.2 The Impact of Category Congruency on Detection Performance

In sum, six models were conducted to assess the effect of categorization on infants' search performance. They revealed that the impact of categorical information on detection performance cannot be seen independently from the luminance contrast of a target. This is shown by the significant improvement of the GLMMs on success when including interactions between diff_luminance and the category-related variables. Indeed, post-hoc t-tests confirmed that hit trials included higher levels of diff_luminance in closer compared to more distant target-background category combinations, whereas in the full data, there was no difference in diff_luminance between more or less distinct category combinations, see Table S 4.19. Moreover, shorter detection latency was predicted by children's closer category assignments but not by adults' assignments or by category-congruency—in spite of the small luminance difference between children's more and less distinct assignments, as shown by the low effect size of $r = .1$.

We therefore interpret the facilitating effect of trials including less distinct category combinations together with high target luminance contrast as appropriately representing the results. In contrast, the facilitating effects of more distinct category combinations, as reflected in the fixed effects of category-congruency and adults' assignments on success, was weaker and became apparent only in combination with the interaction between diff_luminance and category-related properties. It can therefore probably be seen as a side effect of the strong facilitating effect of diff_luminance on infants' detection performance.

Table S 4.19 Differences in Diff_luminance Between More or Less Alike Category Combinations

Variable	Data	*M* low [a]	*M* high [a]	95% *CI*	*t*	*p*	*r* [b]
Category-congruency	All	0.60	0.67	[-0.16, 0.02]	-1.45	.147	.04
	Hits	0.78	1.09	[-0.52, -0.12]	-3.14	.002	.15
Assigned category children	All	0.61	0.64	[-0.11, 0.07]	-0.47	.636	.01
	Hits	0.78	0.98	[-0.38, -0.03]	-2.24	.026	.1
Assigned category adults	All	0.58	0.67	[-0.17, 0.00]	-1.9	.058	.05
	Hits	0.69	1.08	[-0.5, -0,21]	-4.27	< .001	.2

Note. Comparisons of diff_luminance levels between trials with target-background combinations of low or high category alikeness.
[a] Low: mean of diff_luminance in trials with incongruent category combinations for category-congruency, and with distance values ≥ *Md* for assigned categories; High: trials with congruent category combinations for category-congruency, and with distance values < *Md* for assigned categories.
[b] Effect size: point biserial r equivalent of Cohen's d

4.9.3 S3 Discussion

4.9.3.1 Detection Difficulty of Individual Categories

In order to infer if the pattern of detection latency reflects distances between areas related to neural representations of target and background categories, we compared detection latency for all target-background combinations in a post-hoc analysis. Detection latency should be longest for target-background category combinations between artifacts and the other categories for the natural-artificial distinction, or between vegetation and the remaining categories for the animate-inanimate distinction. We conducted an LMM on latency with target category (*t_ category*), background category (*b_ category*), their interaction, and the control variables depth-congruency and movement. Participants were included as random intercept. LRT revealed significant contributions to the model by the IVs t_ category ($\chi^2(2) = 7.1$, $p = .029$), b_ category ($\chi^2(2) = 8.7$, $p = .013$), and the interaction term ($\chi^2(2) = 9.8$, $p = .045$). The main effects of both target category and background category were motivated by longer latencies for artifacts compared to natural elements, respectively. Furthermore, contrasts between the target categories on the individual background categories showed that congruent target-background combinations were detected faster than incongruent combinations. For example, natural-element targets were detected faster on natural elements than on artifacts, and artifact targets were detected slower on vegetation and natural element backgrounds than the respective category-congruent target (all $p < .05$; for further results see Figure S 4.12). This pattern of results indicates that the inclusion of artifacts in a stimulus led to longer detection latencies in contrast to detection latencies for natural categories (i.e., natural elements and vegetation) which led to more similar latencies. Therefore, our results do not differentiate between the explanations that it might be the distance between internal representations of natural and human-made entities, or the immature developmental state of the artifact category, which increase detection latency in category-incongruent trials.

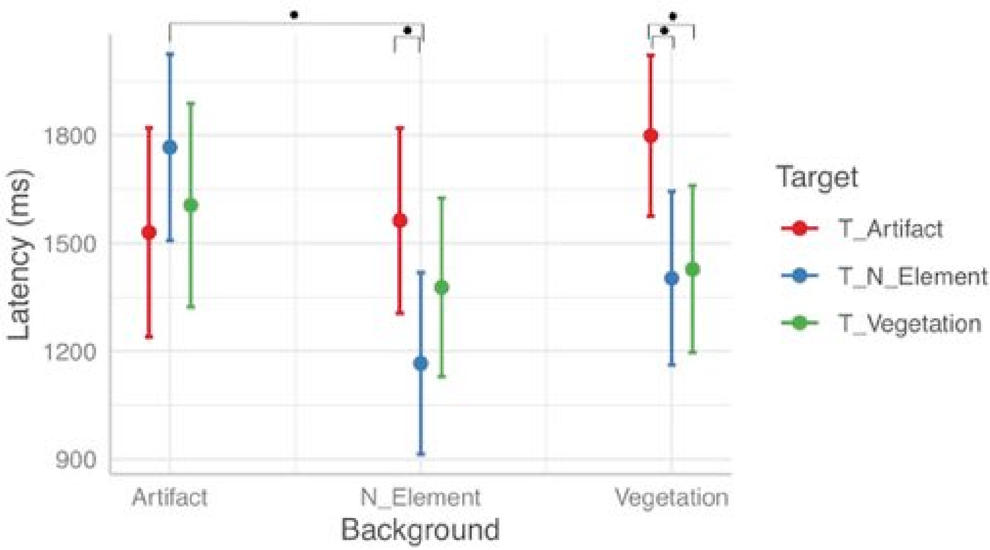


Figure S 4.12: Target and background categories affecting detection latency.
Asterisks indicate interaction effects with *p* values < .05. See main text for further information.

Chapter 5

General Discussion

5.1 Summary of Results

Overall, infants, preschool children, and adults were attending to content-related characteristics and visual properties of naturalistic structures in the sorting tasks and the eye-tracking search task. However, depending on the age group and the kind of categorization (classification, similarity judgments, scene-segmentation), overlapping as well as distinguishable patterns of content- and structure-related attention were revealed between the age groups.

The methodological study (Study 2) showed that calibration targets with interesting centers and low contrast at their periphery, or targets with globally distributed complexity, resulted in better gaze recording outcomes. Body movement substantially contributed to gaze instability and fixation dispersion, and measurement accuracy was negatively affected by the eccentricity of a target's location. Movement toward the screen particularly increased peripheral gaze displacement, and following a target with head turns resulted in less precise gaze recordings. Along with other practices that promoted infants' interest in the display, these findings were carried over into Study 3, where they supported the assessment and analysis of infants' scene-segregation ability.

5.1.1 Category Membership

Study 1 yielded a general sensitivity in children to vegetation, which became apparent in both tasks: the classification of images, and the similarity judgments of images. Children had higher classification sensitivity (assessed with the measure d') to vegetation compared to adults when overall performance differences were accounted for, and they showed equally high classification sensitivity to vegetation and artifacts compared to natural elements, whereas adults were most sensitive to artifacts during classification. During similarity sorting, both children and adults judged images as more similar if they depicted vegetation compared to the other categories, which did not differ.

Infants sensitivity to the categories assessed in the eye-tracking search task (Study 3) became evident only when an interaction term between diff_luminance and the binary variable category-congruency was included in the model. Then, high levels of diff_luminance in combination with congruent categories increased the probability to detect the target, and a main effect of category congruency was qualified by the higher probability to detect category-incongruent targets. The facilitating effect of high diff_luminance in combination with category-congruent targets on detection success resulted in a predominant inclusion of high diff_luminance targets in the analysis of detection latency, in which congruent target-background category combinations were detected faster than incongruent.

Post hoc analysis addressing the effect of each of the respective categories on detection latency revealed that longer detection latencies were rather associated with the presence of the artifact category in a stimulus, compared to the natural categories (i.e., natural elements and vegetation)—independent of the artifact's inclusion as a target or background category.

5.1.2 Visual Properties

In Study 1, preschool children's assumptions about category membership during classification relied less on visual properties than those of adults. Similarly, children's similarity judgments were generally less affected by visual properties than those of adults.

Rated depth was one property that preschool children relied strongly on in both card-sorting tasks. In the classification task, depth contributed to preschoolers inferences during the assignment to categories together with skew, area, and symmetry. Depth also had the highest impact on preschoolers' similarity judgments, followed by skew and regularity, whereas adults judgments were most strongly affected by regularity, followed by symmetry and area. In contrast to their generally weaker reliance on properties during classification, children relied on the property area more strongly in their assignments of images to vegetation. Moreover, a comparison of the participant groups' similarity judgments revealed that they did not differ in their reliance on depth and skew—nor in the equally weak impact of alpha.

In the infant eye-tracking search task, luminance contrast had a strong impact on detection success. However, other visual properties affected detection performance as strongly. Namely, stronger rated depth of the background image led to a lower detection probability and also to longer detection latencies. Target-background differences in area increased detection success, and target-background differences in entropy led to shorter detection latencies.

Preschool children's and adults' category assignments and similarity judgments assessed in Study 1 were included as additional difference-variables in the analysis of infants' search performance, respectively. In combination with diff_luminance, less distinct category assignments by adults and preschoolers increased the probability to detect the respective targets, whereas image similarity did not predict detection success. In contrast, detection latency was shorter when target and background images were assigned to less distinct categories by preschoolers, or whe they were judged as less similar by preschoolers. Yet, latency was not predicted by adults' sorting decisions.

5.2 Did Ecological Significance Affect Categorization?

5.2.1 The Impact of Category Membership

The image categories: vegetation, natural elements and artifacts, varied in the strength in which they respectively predicted or explained the task performance among the different age groups in Study 1 and 3. Figure 5.18 sketches the patterns of impact the particular categories had on task performance for infants, preschoolers and adults. By comparing the relative importance of the three categories, it becomes obvious that the artifact category increases in significance, whereas the natural elements decrease in significance especially during classification for the older age groups. It is possible that the specific roles these categories acquired are reflected in these patterns: Artifact use becomes more important and specialized in everyday life with increasing age, and their classification is a common visual task in adults. However, the visual characteristics of the artifact structures (e.g., the pattern of fabrics, or the form of noodles), are not an essential part of their use and possibly less attended to during similarity judgments. In contrast, everyday experiences with vegetation (e.g., during food preparation or gardening) refer strongly to characteristics such as their texture, consistency and contours—characteristics that allow classification within the plant category, and which may also have guided similarity judgments. A rich representation of plants arises in children by preschool age, and the emerging awareness of their utility, their "liveliness", as well as the importance of classifying them correctly to avoid harm (e.g., Inagaki & Hatano, 1996; Nguyen & Gelman, 2002; Wertz, 2019), may have increased preschoolers' relative attention to this category compared to the other categories.

In contrast, preschool children and adults had a rather low classification sensitivity to the natural element category, and this category only had moderate predictive value for similarity judgments. This could reflect the rather passive role of natural elements in providing background for other tasks (e.g., as part of a scenery or as ground during locomotion). Interestingly, natural elements significantly facilitated target search in infants compared to artifacts. This raises the question whether visual structures belonging to this natural category are already processed more fluently very early in life—in spite of the assumedly lower exposure infants experience to natural elements compared to artifacts, and also in contrast to the findings of Ellemberg et al (2012) on late maturation of sensitivity for typical spatial characteristics of natural scenes, which depicted natural elements and vegetation.

Infants' longer detection latencies for stimuli including artifacts are surprising because one must assume that for infants, who spend the majority of their time inside homes filled with artifacts, human-made objects are very familiar. Yet, vegetation and natural elements have visually dominated the ancestral human environment for a much longer time (e.g., Schick & Toth, 1994), and their perception might be facilitated even if visual abilities are constrained. In addition, categorization of artifacts relies on their purpose or function (Carrara & Mingardo, 2013), making visual properties less diagnostic on the superordinate level.

The differences in relative significance between the categories in infants, children and adults were possibly affected by distinct ways of visual processing. These could have referred to the structures as background or context (e.g., "stuff", Adelson, 2001; natural surroundings, Geisler, 2008), or could have reflected visual processing necessary for categorical inferences (e.g., Binder & Desai, 2011; Mervis & Rosch, 1981; Nosofsky et al., 2018), among others. The different relevance these categories have for the respective age groups, accrual of experiences, and task-specific factors may have influenced which type of category processing occurred.

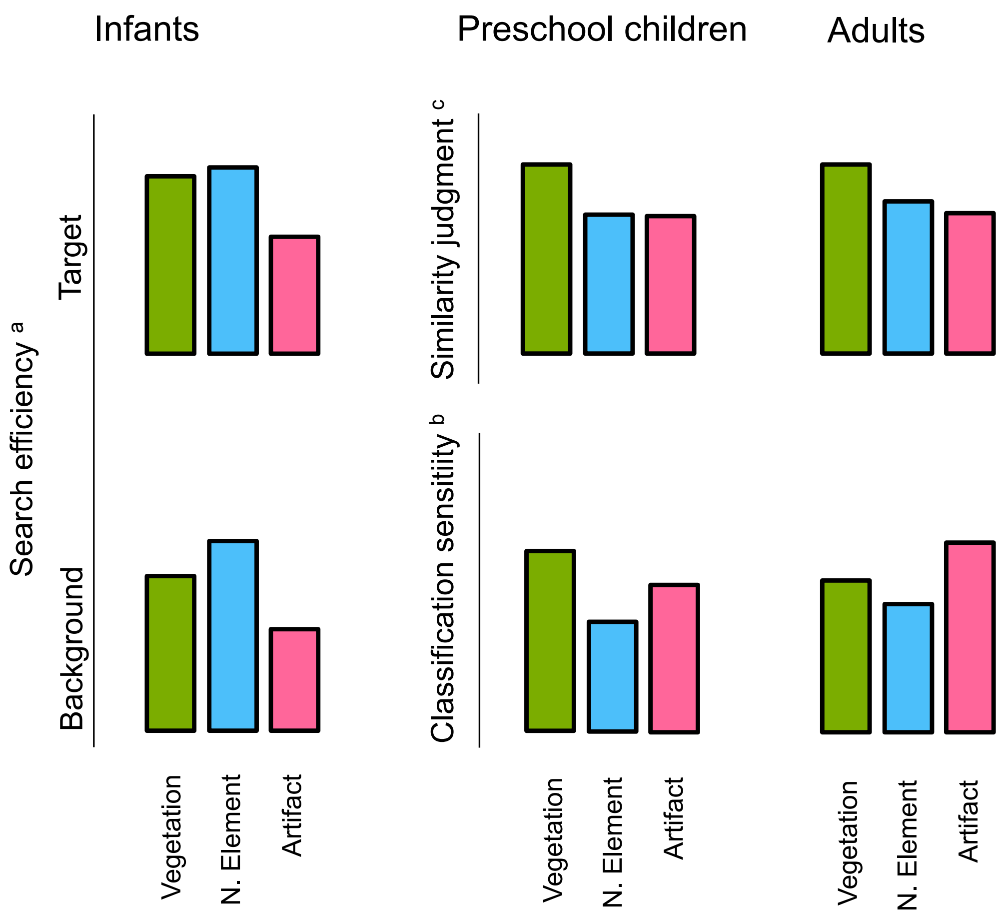


Figure 5.18: Relative importance of categories as Functions of task and age group.
The relative length of the bars are approximations derived from the respective analyses of different dependent variables and therefore do not include comparable units. Overall performance differences between adults and preschool children (i.e., main effects of the factor participant group) are not shown.
[a] Paired contrasts predicted by the LMM on latency. Higher bars indicate shorter latency.
[b] Tuckey's contrasts of the interaction between participant group and true category predicted by the ANOVA on *d'*.
[c] Tuckey's contrasts of the interaction between participant group and assigned category, predicted by the ANOVA on R^2.

5.2.2 The Significance of Depth Cues

The property *rated depth* was of particular significance for infants and children. Yet, the significance of pictorial depth on task performance may have been caused by a variety reasons in the different age groups. The 4- and 5-year-old preschoolers were equally sensitive to depth during classification (i.e., they relied on high levels of depth as indicators for vegetation, and on low levels of depth for artifacts). Preschoolers were also sensitive to depth in their judgment of visual similarity, which indicates a general awareness of spatial characteristics during structure perception. In contrast, infants' sensitivity to rated depth of the background image became obvious, in that it hindered target detection.

Depth engaged the infants differently than other background properties which might have been of similar processing difficulty (e.g., irregular pattern). The literature provides contrasting explanations for why depth may have caused a distraction: Perceptual difficulty and complexity can lead infants to orient away from a stimulus to avoid overstimulation, which is a common regulatory function of attention in older infants (e.g., Rothbart et al., 2011). Yet, space and depth properties are of actual importance for eight-month-olds who start to navigate autonomously. Therefore, engagement in visual inspection and learning opportunities may similarly cause distraction from the search. Relevant experience that infants had with stimuli (e.g., 3-dimensional depth) can lead to longer inspection of a stimulus showing similar properties (e.g., pictorial depth)—the infants' greater expertise may offer more information on the stimulus to be processed (Hurley et al., 2010). The fact that children as young as 4-years old successfully relied on depth during classification, but differed from the older children in their reliance on regularity, symmetry, and curvature (Section 2.8.5 S5), points to an early acquisition of naturalistic pictorial depth and its integration into everyday visual tasks (see also: Kavšek et al., 2012). For example, when children are viewing photographs, the 3-dimensional shape and arrangement of a structure's elements can only be determined if pictorial depth cues are integrated.

The different meanings of depth for infants (perceptual challenge) and children (particular reliance on depth during classification and similarity judgment) might nicely reflect phases within the visual acquisition of functionally significant environmental characteristics. The importance of depth for perceptual organization and physical shape perception during self motion (e.g., Needham, 2000) is seen as a foundation for later object categorization (Atkinson & Braddick, 2013; Bertenthal, 1996). Depth was readily available as a visual cue for the preschoolers in Study 1. A deeper investigation of infants' reactions to the challenge provided by high-rated pictorial depth (e.g., by assessing gaze and movement patterns) in future experiments could offer important insight into the development of sensitivities for visual properties with ecological significance.

5.3 The Inclusion of Visual Properties During Categorization

It became apparent in Study 1, that preschoolers' and adults' inclusion of visual properties was affected by the different tasks (classification vs. similarity grouping). For instance, properties predicting adult classification reflected a similar pattern as that of the images' true categories (qualified by adults' low error rate of 10%), and differed from the properties dominating adults' similarity judgments, which referred mainly to variations in the form and size of structure elements. However, in children, reliance on skew and depth was comparably strong in both tasks. Next to that, children's task specificity became particularly obvious in properties which referred to low-level spatial scales or which made the analysis of shape necessary (Figure 2.**2**Figure 2.**7**).

5.3.1 Perceptual Integration of Naturalistic Structure Components

The strategies used by children and adults in the two categorization tasks may explain the difference in which properties they relied upon. These strategies very likely differed in their attention to particular hierarchical levels of the visual structures (Kimchi, 2015), leading to a more analytic investigation of shape details to solve classification (e.g., Deng & Sloutsky, 2016; Rosch et al., 1976), compared to a more general view on the images during similarity judgments (e.g., configural or shading properties). Additionally, differences in the inclusion of properties could have been caused by the density and distribution of available information, which was lower during classification with one image viewed at a time, compared to similarity sorting with many images presented simultaneously. Children might have been more affected by these differences in organizational demands than adults (Hadad et al., 2010; Hadad & Kimchi, 2006), so a closer look at visual processes involved in the tasks might reveal strategies which enabled children to still solve the tasks.

In complex naturalistic structures, the perception of shape is part of a hierarchical organizational process and relies on several perceptual mechanisms: smaller elements can be grouped, compared, segregated, or perceived in their configural relations, leading to more global elements with which the operations can be repeated (Kimchi, 2015; Wagemans et al., 2012). When viewing the structures, attention can optionally focus more on the detailed elements, or on more general forms. Adults show evidence for global precedence or coarse-to-fine processing hierarchies, but also flexibly integrate hierarchical visual information depending on context, attentional deployment, or the aspect under investigation (Flevaris & Robertson, 2016; Hegdé, 2008). Studies investigating the priority of local versus global image features in children frequently used graphical stimuli, in which local details are operationalized as identically repeated elements (e.g., characters) arranged to resemble a larger bounded form or contour (similar to stimuli introduced in Navon, 1977). With these stimuli, processing advantages are commonly found for local detail until at least 8 years-of-age (e.g., Enns et al., 2000; Kimchi et al., 2005; Scherf et al., 2009), while the integration of local into more global levels starts to develop by 4 years-of-age (Vinter et al., 2010). In a study using photographs (Balas et al., 2020), the younger children who compared materials of bounded objects likewise showed disproportionally higher costs if small details compared to more global features were disrupted, supporting the hypothesis that local superiority still extends over middle childhood even for naturalistic images.

Yet, as suggested above, and similar to adults, children's preference to process global or local graphic features can also be affected by a particular task, the density or number of elements, or whether they are object or non-object (reviewed in: Guy et al., 2013; Kimchi, 2015). Correspondingly, the current results suggest that the preschool children did not follow a clear local preference when confronted with complex naturalistic structures. Instead, they flexibly focused on different levels of the global-local hierarchy, namely: (a) sensitivity to variability in the holistic property regularity increased in importance during similarity grouping compared to classification, (b) distinctions of mid-level shape characteristics between non-symmetrical versus symmetrical forms were present during classification, but reduced during similarity grouping, and (c) sensitivity to the multiple spatial scales differing between low versus high area were more dominant during classification than during similarity grouping.

This flexible uptake of different processing hierarchies might have not only supported inferences to solve the task, but also reflect adjustments to the children's processing capacities or their developing visual abilities. If visual information exceeded the processing capacities of a child, easier properties may have been recruited to proceed in the task. Skew or rated depth—which do not make shape analysis necessary—were recruited more generally in both tasks, whereas mid-level shape features were not integrated into comparisons between several images during similarity grouping. The perception of shape might cause particular difficulties when viewing naturalistic structures, in which hierarchical levels or figure-ground relationships are not clearly defined.

In addition to shapes, children's analysis of fine detail in properties such as alpha seemed to be generally restricted.

5.3.2 The Case of Deviation: Do Young Children Exploit Incomplete Visual Information?

With regard to the restricted inclusion of mid-level shape it is surprising that infants and children distinguished levels of the property area, which makes sensitivity to spatial-frequencies at different scales necessary. The property area assesses spatial scales in a more abstract way—regardless of their integration into any particular higher level characteristic or shape. Therefore, abilities underlying perceptual integration are not as necessary as they are for properties defining shape. Yet, immature acuity and contrast sensitivity in infants and young children sets limits to the perception of pattern, including very high or low spatial frequencies[6] (Atkinson & Braddick, 2013; Ellemberg et al., 1999; Skoczenski & Norcia, 2002).

Infants' sensitivity to variations in area, and preschoolers' reliance on area in their classification suggests that statistical properties, which are rather abstract and processed pre-attentively in adults, might be more central to young children. For example, the visual structures of a natural surrounding might serve as background or context in adults, whereas in young children, variations within these structures offer interesting visual targets (Köster et al., 2017a). These variations within structures can provide experiences which support visual development in infants or the gathering of visual regularities in young children.

Sensitivity to variations in the abstract visual property area (and entropy in Study 3), which include difficult-to-perceive visual information for infants and young children, leads to the question: Can uncertain visual information be integrated into categorical distinctions?

It is possible that image regions of more difficult pattern including fine spatial-frequencies and contrasts draw attention without the necessity to be processed in full detail. Such image regions might function as "perceptual units" as is suggested for proto-objects (homogeneous regions in low-level feature maps; Wischnewski et al., 2009). These regions could serve as targets for task-driven inferences or provoke further action such as exploration or social referencing, even if they are not fully processed—this is supported by theories on novelty seeking and proximal development (e.g., Köster et al., 2020; Vygotsky, 1978) as outlined in Section 1.2 and 4.5.2. In the case of very small detail which is outside the range of spatial acuity, movement towards the stimulus performed by the child would enlarge the details easily and solve prior uncertainty. Such reoccurring inferences may provide models for similar uncertainties (Friston, 2010; Kayhan et al., 2019; O'Regan & Noë, 2001). Additionally, performed actions themselves can serve as representational medium to be drawn upon when solving visual tasks (Rakison & Woodward, 2008). Even in adults, the perceptual impression of a visual surrounding needs to be built on incomplete visual information due to the retinal and neural restrictions of the visual system (e.g., Nassi & Callaway, 2009).

---

[6] Contrast distinction at low (e.g., < .5 cycles per degree; cpd) and high (e.g., > 20 cpd) spatial frequencies are more difficult to perceive than similar contrasts at moderate scales (i.e., 3 – 5 cpd) in all age groups, and 4- to 5-year-olds' sensitivity was lower than adults by a factor of two (Ellemberg et al., 1999). Infants' contrast sensitivity is even more limited at higher spatial frequencies (Norcia et al., 1990).

Preschoolers, whose visual abilities are further developed as those of infants and who are therefore bound to different tasks during their acquisition of environmental regularities might draw on such incomplete regions during categorization. Recall that children's reliance on the more difficult level of area for classifying vegetation exceeded that of adult's, whereas the easier, high level of area significantly predicted children's assignment of artifacts. This makes area additionally a property of ecologically significant category distinctions, which can further increase its inclusion in perceptual inferences.

Undoubtedly, the suggestion that categorization can rely on incomplete or difficult-to-perceive visual property levels needs further investigation. Future developmental studies could, for example, experimentally control levels of detail and perceptual difficulty in categorization tasks with naturalistic structures.

5.4 Relating Infants' Visual Search to Preschoolers' and Adults' Card Sorting

In the infant eye-tracking search task, two distance-variables were included which derived from the card sorting task of preschoolers and adults, respectively. These represented the distance between the categories assigned to two images, and the dissimilarity of two images as assessed by similarity judgments. The variables were included as an alternative to testing a comparison group of adults in the infant eye-tracking task—the originally planned adult data collection had been interrupted by the spring 2020 coronavirus pandemic. By testing a comparison group of adult participant, we intended to validate the balancing of general target salience between conditions. Furthermore, the particular impact of content- and structure related properties on infants' search could have been compared and relativized to the adults' performance. The current alternative variables from the sorting tasks, which refer to the same images as those included in the eye-tracking study, were expected to indicate more general overlapping patterns of visual processing between infants, preschoolers, and adults.

These variables yielded the result that infants detection success was similarly strong predicted by the classification decisions of preschoolers and adults—if targets and backgrounds differed sufficiently in luminance. Yet, infants' detection latency was only predicted by the preschoolers sorting decisions, including their similarity judgments as well as their category assignments. This shows that, if 8-month-olds were able to detect a target, the factors which determined detection latency were overlapping with the preschoolers' perception of the images, but not with the adults. The overlapping predictive value of the property depth in infants and preschoolers cannot explain the results, because depth was only assessed for background images in the search task. Neither can related sensitivities to variations in luminance and contrast have determined the results, since target luminance did not vary in the models on detection latency. Overlapping factors may therefore relate to rather general aspects of vision development which are shared between infants and preschoolers.

So far, the current project suggests three particularities in which young children's processing of complex naturalistic structures differed from adults. These are (i) less efficient perceptual organization or shape processing, (ii) particular attentional priorities to image regions related to variations within the naturalistic scenes which may be of functional relevance, and (iii) sensitivities to difficulty or uncertainty within visual structures, provided by properties such as area or entropy. Recall, due to the innate tendency to detect novel information (e.g., Hunnius et al., 2006), and infants' and young children's attention to visual information which is of functional relevance for explorative purposes (e.g., aspects of three-dimensionality, spatial configurations or surface qualities; Bushnell & Boudreau, 1993), young children will attend to visual cues which may not be obvious for adults during categorization (e.g., Gibson, 2000; Kovács, 2000)7.

Additionally, there might be further overlap in infants and preschoolers because they have less experience with cultural habits transported by language than older children and adults (e.g., Majid et al., 2004), which may also affect scene perception. For example, by school age, children more strongly direct their attention towars targets within a scene that are more coherent with a cultural habit than in younger children (e.g., more fixations at central compared to contextual regions in western cultures; Duffy et al., 2009; Köster et al., 2017a). Likewise, young children's categorization is more affected by visual regularities and experiences during actions, while older children's and adults' classification is increasingly affected by culture and language (Nazzi & Gopnik, 2001; Rakison & Woodward, 2008; Westermann & Mareschal, 2012).

Therefore, several qualitative differences between adult's and young children's categorization and visual attention may provide non-exclusive explanations for the relation of preschoolers' sorting decisions to infant's search latency.

[7] Interestingly, Tatler et al. (2011) suggested that adults' eye movements during natural behavior (in contrast to scene perception in experimental settings) reflect the outcome of long time scales of reinforcement learning—anticipations of the results of explorative behaviors or visually guided actions—during development. The early sensitivity to functionally significant visual information which was present in the current results may be part of this acquisition process.

5.5 Which Visual Cues May Have Affected Infants' Particular Reactions to Plants in Previous Experiments?

One of the main inspirations for the current project were previous findings on 6- to 18-month-olds' reactions to real plants and real-looking artificial plants (i.e., protracted touch, increased social information seeking, identification as food source), compared to novel artifacts, familiar objects and natural objects (C. Elsner & Wertz, 2019; Wertz & Wynn, 2014a, 2014b; Włodarczyk et al., 2018). These previous studies showed that infants' behavioral reactions reflect adaptations to problems faced by humans when having to decide which plants are beneficial and which are dangerous (Wertz, 2019). Can the inferences drawn from the current dissertation help to understand which visual characteristics of the real and artificial plants were affecting infants' particular behavioral reactions? The answer is difficult, because the experiments of the previous studies differed fundamentally in their materials and methods from the current work.

Here, the stimuli were intended to represent vegetation as it occurs in natural environments (i.e., as structure-like regions among other more or less complex structures), and their lack of defined boundaries, three-dimensionality, and color cues differs from the individual plants and the other objects presented in the previous studies. Moreover, Study 3—which 8-month-old participants are equivalent in their age to the youngest infants of the previous studies—was designed to investigate general sensitivity to congruent versus incongruent category combinations, but not to a particular category as was attempted in the previous studies. Still, post-hoc analysis revealed that infants' sensitivity to target and background categories differed between artifacts and the natural categories (i.e., natural elements and vegetation). Of the visual properties investigated in Study 3, depth most strongly affected search performance in that it hindered target detection. The statistical properties area and entropy also affected search performance due to their variations between more or less spatial scales and shades. Infants' sensitivity to depth and to area and entropy may therefore be carefully related to the results of the previous studies.

In the previous studies, characteristics of real-looking plants were moulded onto the novel artifact stimuli (e.g., their dominant upper part, the green color), to understand if some of these characteristics alone triggered similar behavior as plants did. The novel artifacts also differed from plants in other characteristics such as bright colors, artificial materials, or higher-level and conceptual features like the lack of a base, crumpled paper instead of leaves, or a cylinder instead of several stalks (the objects are shown in: Wertz, 2019). Despite these imitations, the novel artifacts did not elicit similar behavior in infants as plants did. The modifications made, though, leave open the possibility that properties on a more abstract level were affecting infants' reactions.

One particular characteristic which was strongly present in the plant stimuli was that their upper part consisted of collections of leaves. These leaves additionally showed faintly shaded patterns. The spatially distributed particles and the low contrasts possibly provided visual processing difficulty to the infants, whose central stereopsis and contrast sensitivity is still developing (Ellemberg et al., 1999; Giaschi et al., 2013). Uncertainty or overwhelming complexity caused by such difficult visual information can lead to reactions similar to those exhibited by the infants in the previous studies, namely social referencing (Pauen & Hoehl, 2015), or avoidance (e.g., by orienting away from the stimulus; Ruff & Rothbart, 2001). Since infants spent more time looking at the plants than the other object types (C. Elsner & Wertz, 2019), the plants' perceptual complexity might have provided an uncertain but interesting counterpart to the infants which increased their attempts to seek for and rely on social information, by hesitating to reach out for them directly—it seems adaptive to treat regions of the environment which have such uncertain spatial characteristics with care.

But there is another of the current findings that could be important here: preschoolers in Study 1 included high levels of the property area in their assignments to artifacts, and had the tendency to relate low area to vegetation. Recall that high levels of area indicate few spatial scales, which can for example be found in an object with an even surface. The current artifact structures were defined by high area, whereas the vegetation and natural elements structures were rather low in area (Figure S 2.**1**). That infants in the previous studies treated natural objects differently than plants may result from the natural entities' more object-like appearance. This can be expected to relate to higher levels of area, which would let them appear more similar to artifact objects and more distinct to vegetation.

The suggestions made here certainly include the possibility that sensitivity to additional visual features not captured in the current study may support the elicitation of adaptive behaviors. These adaptive behaviors might refer to more specific roles of vegetation (e.g., as food source). Color may well be an important additional cue (Fischer, 2012; Gegenfurtner & Rieger, 2000), even though it did not affect search performance in Study 3. Furthermore, a more direct linking of the projects could provide further insight. For example, an assessment of the same visual properties as those investigated here from photographs of the previous stimuli objects, together with an analysis of the properties' impact on infants' reactions on the item level, may help to identify visual cues which affected infants' particular reactions to plants.

5.6 Limitations and Further Directions

As outlined in Study 1 and 3, Sections 2.5.5 and 4.6, choices of materials which were qualified by their applicability in the current experiments also constrain interpretations of the findings. For example, the two-dimensional, monochromatic, and static photographs may lead to different responses in young children than the real entities depicted on them (e.g., T. M. Gerhard et al., 2016). Furthermore, despite the large number of stimuli with high variability in their appearance, the current image set only represents a small excerpt of the variability of real entities, and any image selection can lead to unintended biases which may affect the results (e.g., Yarkoni & Westfall, 2017). Despite these limitations, images of real-world structures are still expected to make valid inferences on visual processing abilities possible, in particular on the proccesing of characteristics like category membership and visual properties, which was investigated here.

It will nevertheless be important to scrutinize and particularize several of the current inferences in future experiments. Deeper insight into some subjects could then provide important contributions to the literature on visual and cognitive development. This suggestion applies in particular to (a) the impact of depth cues on children's solving of the visual tasks; (b) the possible role of incomplete visual information in young children's visual inferences; (c) the seemingly more effortful processing of incongruent relative to congruent categories in infants; and (d) the stronger overlap of infants' processing of visual structures with that of preschool children compared to adults. Follow-up investigations are especially important since the current project is part of a quite novel and seminal field of research.

The methods and materials chosen here also more generally allow to connect the current findings to research in adults which used similar methodological approaches to investigate visual categorization ability (e.g., Heaps & Handel, 1999; Schmidt et al., 2017). It could be very beneficial to relate early visual processing to adult scene perception and categorization to better understand how rapid processing of natural scenery is possible in adults.

5.7 General Conclusions

The current dissertation project investigated the interrelating effects of developing visual abilities, visual properties and ecological significance on categorization in infants, preschool children, and adults. Photographs of naturalistic structures of the categories vegetation, non-living natural elements, and artifacts were used as stimuli. We showed that infants and preschool children were able to process these complex stimuli and adapt their visual inferences to the experimental tasks. All participant groups processed category membership as well as visual properties during their inferences. We also found that infants were sensitive to discontinuities in pattern-related properties independent of the pattern's luminance contrasts when searching for the target structure patch. The results of the categorization studies suggest that classification of visual structures, and attention to particular visual properties, is affected by the functional or ecological significance these categories and properties may have for the respective age groups. A comparison between infants' and preschool children's sensitivity to the categories and visual properties implies, that the period between the second half of the first year and preschool age must have high impact on the acquisition of models on how to identify significant categories. In the current investigation, this achievement was most obvious for vegetation. Early sensitivities to the properties of depth and area may have become the basis for preschool children's ability to draw on these properties in distinguishing vegetation from the other categories. An early more general sensitivity to visual properties also highlights the importance of including naturalistic and structure-like stimuli in developmental research on visual categorization. Such stimuli reveal particular visual competencies in young children, which may not be uncovered by graphic stimuli or images of bounded objects. Examples are the dominance of spatial information over shape-related information, or sensitivity to differences in perceptual complexity of spatial frequency distributions during scene segmentation.

This dissertation provided examples of coherent experimental settings on different aspects of categorization that can be applied to different age groups. Its experimental and explorative inferences were intended to stimulate further research. In particular, investigations with children, who have not yet reached school age, may promote the understanding of important aspects on how regularities of the environment are integrated into visual tasks. Suggestions for a transition from more action-related vision in infancy to an increasing influence of recognition memory and classification ability on scene perception in older children could then be specified (e.g., Bertenthal, 1996; Kovács, 2000). Furthermore, infants' attention directed at photographs and their behavioral reactions to visual information may be strongly affected by natural statistics, because these properties transport fundamental information to solve environmental tasks—without that fine-grained analysis or recognition memory is necessary (e.g., sensitivity to variations in the complexity of entropy and area; see also: Geisler, 2008, emphasizing the importance of building hypotheses on scene statistics). Neuroscientific methods may help to identify differences in the processing of statistical properties due to their respective integration into behavioral strategies or classification-related tasks.

References

Açık, A., Onat, S., Schumann, F., Einhäuser, W., & König, P. (2009). Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. *Vision Research*, *49*(12), 1541–1553. https://doi.org/10.1016/j.visres.2009.03.011

Adelson, E. H. (2001). On seeing stuff: The perception of materials by humans and machines. 4299, 1–12. https://doi.org/10.1117/12.429489

Adolph, K. E. (2000). Specificity of learning: Why infants fall over a veritable cliff. *Psychological Science*, *11*(4), 290–295.

Adolph, K. E., & Tamis‐LeMonda, C. S. (2014). The costs and benefits of development: The transition from crawling to walking. *Child Development Perspectives*, *8*(4), 187–192.

Alahyane, N., Lemoine-Lardennois, C., Tailhefer, C., Collins, T., Fagard, J., & Doré-Mazars, K. (2016). Development and learning of saccadic eye movements in 7- to 42-month-old children. *Journal of Vision*, *16*(1), 6. https://doi.org/10.1167/16.1.6

Almoqbel, F. M., Irving, E. L., & Leat, S. J. (2017). Visual acuity and contrast sensitivity development in children: Sweep visually evoked potential and psychophysics. *Optometry and Vision Science*, *94*(8), 830–837.

Amso, D., & Davidow, J. (2012). The development of implicit learning from infancy to adulthood: Item frequencies, relations, and cognitive flexibility. *Developmental Psychobiology*, *54*(6), 664–673. https://doi.org/10.1002/dev.20587

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PloS One*, *9*(1), e85701.

Anderson, D. I., Campos, J. J., Witherington, D. C., Dahl, A., Rivera, M., He, M., Uchiyama, I., & Barbu-Roth, M. (2013). The role of locomotion in psychological development. *Frontiers in Psychology*, *4*, 440.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Aslin, R. N. (2011). Perceptual organization of visual structure requires a flexible learning mechanism. *Infancy*, *16*(1), 39–44.

Aslin, R. N. (2012). Infant Eyes: A Window on Cognitive Development. *Infancy*, *17*(1), 126–140. https://doi.org/10.1111/j.1532-7078.2011.00097.x

Aslin, R. N., & Smith, L. B. (1988). Perceptual Development. *Annual Review of Psychology*, *39*(1), 435–473.

Atkinson, J., & Braddick, O. (2013). Visual Development. *The Oxford Handbook of Developmental Psychology, Vol. 1*. https://doi.org/10.1093/oxfordhb/9780199958450.013.0010

Backscheider, A. G., Shatz, M., & Gelman, S. A. (1993). Preschoolers' ability to distinguish living kinds as a function of regrowth. *Child Development*, *64*(4), 1242–1257.

Badger, J. R., & Shapiro, L. R. (2012). Evidence of a transition from perceptual to category induction in 3- to 9-year-old children. *Journal of Experimental Child Psychology*, *113*(1), 131–146. https://doi.org/10.1016/j.jecp.2012.03.004

Balas, B. (2017). Children's use of visual summary statistics for material categorization. *Journal of Vision*, *17*(12), 22–22. https://doi.org/10.1167/17.12.22

Balas, B., & Woods, R. (2014). Infant Preference for Natural Texture Statistics is Modulated by Contrast Polarity. *Infancy*, *19*(3), 262–280. https://doi.org/10.1111/infa.12050

Balas, B., Auen, A., Thrash, J., & Lammers, S. (2020). Children's use of local and global visual features for material perception. *Journal of Vision*, *20*(2), 10. https://doi.org/10.1167/jov.20.2.10

Balas, B., Saville, A., & Schmidt, J. (2018). Neural sensitivity to natural texture statistics in infancy. *Developmental Psychobiology*, *60*(7), 765–774.

Barrett, H. C. (2014). The shape of thought: How mental adaptations evolve. Oxford University Press.

Bartoń, K. (2019). *MuMIn: Multi-model inference.* (1.43. 6) [Computer software]. https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(i01).

Baumgartner, E., & Gegenfurtner, K. R. (2016). Image Statistics and the Representation of Material Properties in the Visual Cortex. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01185

Baumgartner, E., Wiebel, C. B., & Gegenfurtner, K. R. (2013). Visual and Haptic Representations of Material Properties. *Multisensory Research*, *26*(5), 429–455. https://doi.org/10.1163/22134808-00002429

Beazley, L. D., Illingworth, D. J., Jahn, A., & Greer, D. V. (1980). Contrast sensitivity in children and adults. *British Journal of Ophthalmology*, *64*(11), 863–866.

Beeck, H. P. O. de, Pillet, I., & Ritchie, J. B. (2019). Factors Determining Where Category-Selective Areas Emerge in Visual Cortex. *Trends in Cognitive Sciences*, *23*(9), 784–797. https://doi.org/10.1016/j.tics.2019.06.006

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Bertenthal, B. I. (1996). Origins and Early Development of Perception, Action, and Representation. *Annual Review of Psychology*, *47*(1), 431–459. https://doi.org/10.1146/annurev.psych.47.1.431

Bhatt, R. S., & Quinn, P. C. (2011). How Does Learning Impact Development in Infancy? The Case of Perceptual Organization: PERCEPTUAL LEARNING IN INFANCY. *Infancy*, *16*(1), 2–38. https://doi.org/10.1111/j.1532-7078.2010.00048.x

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.

Bloom, P. (2002). How children learn the meanings of words. MIT press.

Bondarko, V. M., & Semenov, L. A. (2012). Visual acuity and hyperacuity in 11-to 17-year-old secondary school students. *Human Physiology*, *38*(3), 271–275.

Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. *Developmental Psychology*, *46*(2), 350.

Bornstein, M. H., & Benasich, A. A. (1986). Infant Habituation: Assessments of Individual Differences and Short-Term Reliability at Five Months. *Child Development*, *57*(1), 87–99. https://doi.org/10.2307/1130640

Bornstein, M. H., & Stiles-Davis, J. (1984). Discrimination and memory for symmetry in young children. *Developmental Psychology*, *20*(4), 637.

Bornstein, M. H., Ferdinandsen, K., & Gross, C. G. (1981). Perception of symmetry in infancy. *Developmental Psychology*, *17*(1), 82.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research*, *51*(13), 1588–1609. https://doi.org/10.1016/j.visres.2011.02.018

Brodatz, P. (1966). Textures: A photographic album for artists and designers, by Phil Brodatz. Dover publications.

Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development*, *65*(5), 1243–1261.

Brown, A. M., & Lindsey, D. T. (2009). Contrast Insensitivity: The Critical Immaturity in Infant Visual Performance: *Optometry and Vision Science*, *86*(6), 572–576. https://doi.org/10.1097/OPX.0b013e3181a72980

Buquet, C., & Charlier, J. R. (1996). Evaluation of sensory visual development based on measures of oculomotor responses. In F. Vital-Durand, J. Atkinson, & O. J. Braddick (Eds.), *Infant Vision* (pp. 291–306). Oxford University Press.

Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*(1), 157–170.

Bushnell, E. W., & Boudreau, J. P. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. *Child Development*, *64*(4), 1005–1021.

Cabral, L., Zubiaurre, L., Wild, C., Linke, A., & Cusack, R. (2019). Category-Selective Visual Regions Have Distinctive Signatures of Connectivity in Early Infancy. *BioRxiv*, 675421.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy*, *1*(2), 149–219.

Carey, S. (1988). Conceptual differences between children and adults. *Mind & Language*, *3*(3), 167–181.

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. *Journal of Cognitive Neuroscience*, *26*(1), 132–142.

Carrara, M., & Mingardo, D. (2013). Artifact Categorization. Trends and Problems. *Review of Philosophy and Psychology*, *4*(3), 351–373. https://doi.org/10.1007/s13164-013-0151-6

Casler, K., & Kelemen, D. (2005). Young children's rapid learning about artifacts. *Developmental Science*, *8*(6), 472–480.

Chiang, W.-C., & Wynn, K. (2000). Infants' tracking of objects and collections. *Cognition*, *77*(3), 169–195.

Choe, K. W., Blake, R., & Lee, S.-H. (2016). Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision Research*, *118*, 48–59.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455.

Clausi, D. A. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, *28*(1), 45–62.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, 869–879.

Cohen, M. A., Dilks, D. D., Koldewyn, K., Weigelt, S., Feather, J., Kell, A. J., Keil, B., Fischl, B., Zöllei, L., & Wald, L. (2019). Representational similarity precedes category selectivity in the developing ventral visual pathway. *NeuroImage*, *197*, 565–574.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337–367.

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood.

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., … Jones, W. (2017). Infant viewing of social scenes is under genetic control and atypical in autism. *Nature*, *547*(7663), 340–344.

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Costa, A. F., Humpire-Mamani, G., & Traina, A. J. M. (2012). An efficient algorithm for fractal analysis of textures. *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference On*, 39–46.

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.

Crouzet, S. M. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

Crouzet, S. M., Joubert, O. R., Thorpe, S. J., & Fabre-Thorpe, M. (2012). Animal Detection Precedes Access to Scene Category. *PLOS ONE*, *7*(12), e51471. https://doi.org/10.1371/journal.pone.0051471

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An Examination of Recording Accuracy and Precision From Eye Tracking Data From Toddlerhood to Adulthood. *Frontiers in Psychology, 9.*

Daw, N. W. (2014). *Visual Development*. Springer US. https://doi.org/10.1007/978-1-4614-9059-3

De Leeuw, J., & Mair, P. (2011). Multidimensional scaling using majorization: SMACOF in R.

De Vries-Khoe, L. H., & Spekreijse, H. (1982). Maturation of luminance and pattern EPs in man. *Doc Ophthalmol Proc Ser*, *31*, 461–475.

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Kanwisher, N., & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, *8*, 13995.

Dekker, T. M., Ban, H., van der Velde, B., Sereno, M. I., Welchman, A. E., & Nardini, M. (2015). Late Development of Cue Integration Is Linked to Sensory Fusion in Cortex. *Current Biology*, *25*(21), 2856–2861. https://doi.org/10.1016/j.cub.2015.09.043

Dekker, T. M., 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.

Dekker, T. M., Schwarzkopf, D. S., de Haas, B., Nardini, M., & Sereno, M. I. (2019). Population receptive field tuning properties of visual cortex during childhood. *Developmental Cognitive Neuroscience*, *37*, 100614. https://doi.org/10.1016/j.dcn.2019.01.001

DeLoache, J. S. (2011). Early development of the understanding and use of symbolic artifacts. In U. C. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (Vol. 2, pp. 312–336). Blackwell Publishers Ltd.

DeLoache, J. S., & LoBue, V. (2009). The narrow fellow in the grass: Human infants associate snakes and fear. *Developmental Science*, *12*(1), 201–207.

Deng, W. (Sophia), & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24–62. https://doi.org/10.1016/j.cogpsych.2016.09.002

Doucet, M.-E., Gosselin, F., Lassonde, M., Guillemot, J.-P., & Lepore, F. (2005). Development of visual-evoked potentials to radially modulated concentric patterns. *Neuroreport*, *16*(16), 1753–1756.

Duffy, S., Toriyama, R., Itakura, S., & Kitayama, S. (2009). Development of cultural strategies of attention in North American and Japanese children. *Journal of Experimental Child Psychology*, *102*(3), 351–359.

Eimas, P. D. (1994). Categorization in early infancy and the continuity of development. *Cognition*, *50*(1–3), 83–93.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*(4), 5–5.

Ellemberg, D., Hansen, B. C., & Johnson, A. (2012). The developing visual system is not optimally sensitive to the spatial statistics of natural images. *Vision Research*, *67*, 1–7. https://doi.org/10.1016/j.visres.2012.06.018

Ellemberg, D., Johnson, A., & Hansen, B. (2009). The development of contrast sensitivity for gratings and natural images: Revisiting the golden standard. *Journal of Vision*, *9*(8), 984–984. https://doi.org/10.1167/9.8.984

Ellemberg, D., Lewis, T. L., Liu, C. H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research*, *39*(14), 2325–2333.

Elsner, B., Jeschonek, S., & Pauen, S. (2013). Event-related potentials for 7-month-olds' processing of animals and furniture items. *Developmental Cognitive Neuroscience*, *3*, 53–60. https://doi.org/10.1016/j.dcn.2012.09.002

Elsner, B., Pauen, S., & Jeschonek, S. (2006). Physiological and behavioral parameters of infants' categorization: Changes in heart rate and duration of examining across trials. *Developmental Science*, *9*(6), 551–556. https://doi.org/10.1111/j.1467-7687.2006.00532.x

Elsner, C., & Wertz, A. E. (2019). The seeds of social learning: Infants exhibit more social looking for plants than other object types. *Cognition*, *183*, 244–255. https://doi.org/10.1016/j.cognition.2018.09.016

Enns, J. T., Burack, J. A., Iarocci, G., & Randolph, B. (2000). The orthogenetic principle in the perception of "forests" and "trees"? *Journal of Adult Development*, *7*(1), 41–48.

EyeLink Data Viewer User's Manual. (2015) (Version 2.3.1). Mississauga, Ontario, Canada: SR Research Ltd.

EyeLink 1000 Plus User Manual (Version 1.0.6). (2015). SR Research Ltd.

Fantz, R. L., & Nevis, S. (1967). Pattern preferences and perceptual-cognitive development in early infancy. *Merrill-Palmer Quarterly of Behavior and Development*, *13*(1), 77–108.

Farzin, F., Hou, C., & Norcia, A. M. (2012). Piecing it together: Infants' neural responses to face and object structure. *Journal of Vision*, *12*(13), 6–6. https://doi.org/10.1167/12.13.6

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379–2394. https://doi.org/10.1364/JOSAA.4.002379

Fischer, S. (2012). Ursprung der emotionalen Semantik von kongruenten Farben und Tönen: Assoziationen und Emotionen Erwachsener sowie Blickpräferenzen in der frühen Kindheit. https://pub.uni-bielefeld.de/publication/2551430

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Fisher, A. V., Godwin, K. E., Matlen, B. J., & Unger, L. (2015). Development of Category-Based Induction and Semantic Knowledge. *Child Development*, *86*(1), 48–62. https://doi.org/10.1111/cdev.12277

Fleming, R. W. (2015). Visual Development: Learning Not to See. *Current Biology*, *25*(24), R1166–R1168. https://doi.org/10.1016/j.cub.2015.11.012

Fleming, R. W. (2017). Material Perception. *Annual Review of Vision Science*, *3*(1), 365–388. https://doi.org/10.1146/annurev-vision-102016-061429

Fleming, R. W., Wiebel, C., & Gegenfurtner, K. (2013). Perceptual qualities and material classes. *Journal of Vision*, *13*(8), 9–9.

Flevaris, A. V., & Robertson, L. C. (2016). Spatial frequency selection and integration of global and local information in visual processing: A selective review and tribute to Shlomo Bentin. *Neuropsychologia*, *83*, 192–200. https://doi.org/10.1016/j.neuropsychologia.2015.10.024

Fornari, E., Rytsar, R., & Knyazeva, M. G. (2014). Development of spatial integration depends on top-down and interhemispheric connections that can be perturbed in migraine: A DCM analysis. *Neurological Sciences*, *35*(1), 215–224. https://doi.org/10.1007/s10072-014-1777-6

Forssman, L., Ashorn, P., Ashorn, U., Maleta, K., Matchado, A., Kortekangas, E., & Leppänen, J. M. (2017). Eye-tracking-based assessment of cognitive function in low-resource settings. *Archives of Disease in Childhood*, *102*(4), 301–302. https://doi.org/10.1136/archdischild-2016-310525

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage publications. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fox, J., & Weisberg, S. (2019). Using car Functions in Other Functions.

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology*, *118*, 13–26.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the Development of Social Attention Using Free-Viewing. *Infancy*, *17*(4), 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Frazor, R. A., & Geisler, W. S. (2006). Local luminance and contrast in natural images. *Vision Research*, *46*(10), 1585–1598.

Freud, E., & Behrmann, M. (2017). The life-span trajectory of visual perception of 3D objects. *Scientific Reports*, *7*(1), 1–11.

Friesen, M. C., Shortreed, S. M., Wheeler, D. C., Burstyn, I., Vermeulen, R., Pronk, A., Colt, J. S., Baris, D., Karagas, M. R., & Schwenn, M. (2015). Using hierarchical cluster models to systematically identify groups of jobs with similar occupational questionnaire response patterns to assist rule-based expert exposure assessment in population-based studies. *Annals Of Occupational Hygiene*, *59*(4), 455–466.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gathers, A. D., Bhatt, R., Corbly, C. R., Farley, A. B., & Joseph, J. E. (2004). Developmental shifts in cortical loci for face and object recognition. *Neuroreport*, *15*(10), 1549–1553.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805–808. https://doi.org/10.1016/S0960-9822(00)00563-7

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.*, *59*, 167–192.

Geisler, W. S., & Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1420), 419–448.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*(6), 711–724.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409. https://doi.org/10.1016/j.tics.2004.07.001

Gerhard, H. E., Wichmann, F. A., & Bethge, M. (2013). How Sensitive Is the Human Visual System to the Local Statistics of Natural Images? *PLOS Computational Biology*, *9*(1), e1002873. https://doi.org/10.1371/journal.pcbi.1002873

Gerhard, T. M., Culham, J. C., & Schwarzer, G. (2016). Distinct visual processing of real objects and pictures of those objects in 7-to 9-month-old infants. *Frontiers in Psychology*, *7*, 827.

Giaschi, D., Narasimhan, S., Solski, A., Harrison, E., & Wilcox, L. M. (2013). On the typical development of stereopsis: Fine and coarse processing. *Vision Research*, *89*, 65–71. https://doi.org/10.1016/j.visres.2013.07.011

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice-Hall.

Gibson, E. J. (2000). Perceptual learning in development: Some basic concepts. *Ecological Psychology*, *12*(4), 295–302.

Gibson, E. J., & Walk, R. D. (1960). The" visual cliff". *Scientific American*, *202*(4), 64–71.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., & Toga, A. W. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*(2), 125–157.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In *Perceptual organization in vision* (pp. 245–290). Psychology Press.

Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing*. Pearson.

Graham, D., Schwarz, B., Chatterjee, A., & Leder, H. (2016). Preference for luminance histogram regularities in natural scenes. *Vision Research*, *120*, 11–21. https://doi.org/10.1016/j.visres.2015.03.018

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*(11), 2809–2815.

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye Tracking in Infancy Research. *Developmental Neuropsychology*, *35*(1), 1–19. https://doi.org/10.1080/87565640903325758

Groll, A., & Tutz, G. (2014). Variable selection for generalized linear mixed models by L 1-penalized estimation. *Statistics and Computing*, *24*(2), 137–154.

Guy, M. W., Reynolds, G. D., & Zhang, D. (2013). Visual attention to global and local stimulus properties in 6‐month‐old infants: Individual differences and event‐related potentials. *Child Development*, *84*(4), 1392–1406.

Hadad, B., & Kimchi, R. (2006). Developmental trends in utilizing perceptual closure for grouping of shape: Effects of spatial proximity and collinearity. *Perception & Psychophysics*, *68*(8), 1264–1273.

Hadad, B., Maurer, D., & Lewis, T. L. (2010). The effects of spatial proximity and collinearity on contour integration in adults and children. *Vision Research*, *50*(8), 772–778. https://doi.org/10.1016/j.visres.2010.01.021

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (Vol. 5). Prentice hall Upper Saddle River, NJ.

Haith, M. M. (2004). Progress and Standardization in Eye Movement Work With Human Infants. *Infancy*, *6*(2), 257–265. https://doi.org/10.1207/s15327078in0602_6

Hammarrenger, B., Leporé, F., Lippé, S., Labrosse, M., Guillemot, J.-P., & Roy, M.-S. (2003). Magnocellular and parvocellular developmental course in infants during the first year of life. *Documenta Ophthalmologica*, *107*(3), 225–233.

Hansen, B. C., & Hess, R. F. (2006). Discrimination of amplitude spectrum slope in the fovea and parafovea and the local amplitude distributions of natural scene imagery. *Journal of Vision*, *6*(7), 3. https://doi.org/10.1167/6.7.3

Hardy, K. (2018). Plant use in the Lower and Middle Palaeolithic: Food, medicine and raw materials. *Quaternary Science Reviews*, *191*, 393–405.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 545–552.

Hartig, F. (2020). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models (Version 0.2.7) [R]. https://CRAN.R-project.org/package=DHARMa

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456.

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*(2), 510–527.

Heaps, C., & Handel, S. (1999). Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 299–320. https://doi.org/10.1037/0096-1523.25.2.299

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439.

Hespos, S. J., & VanMarle, K. (2012). Physics for infants: Characterizing the origins of knowledge about objects, substances, and number. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(1), 19–27.

Hespos, S. J., Ferry, A. L., & Rips, L. J. (2009). Five-month-old infants have different expectations for solids and liquids. *Psychological Science*, *20*(5), 603–611.

Hespos, S. J., Ferry, A. L., Anderson, E. M., Hollenbeck, E. N., & Rips, L. J. (2016). Five-month-old infants have general knowledge of how nonsolid substances behave and interact. *Psychological Science*, *27*(2), 244–256.

Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., & Kemner, C. (2015). Consequences of Eye Color, Positioning, and Head Movement for Eye-Tracking Data Quality in Infant Research. *Infancy*, *20*(6), 601–633. https://doi.org/10.1111/infa.12093

Hessels, R. S., Cornelissen, T. H. W., Kemner, C., & Hooge, I. T. C. (2015). Qualitative tests of remote eyetracker recovery and performance during head rotation. *Behavior Research Methods*, *47*(3), 848–859. https://doi.org/10.3758/s13428-014-0507-6

Hessels, R. S., Hooge, I. T. C., & Kemner, C. (2016). An in-depth look at saccadic search in infancy. *Journal of Vision*, *16*(8), 10–10. https://doi.org/10.1167/16.8.10

Hickling, A. K., & Gelman, S. A. (1995). How does your garden grow? Early conceptualization of seeds and their place in the plant growth cycle. *Child Development*, *66*(3), 856–876.

Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *NeuroImage*, *57*(2), 482–494. https://doi.org/10.1016/j.neuroimage.2011.04.056

Hoehl, S. (2016). The development of category specificity in infancy – What can we learn from electrophysiology? *Neuropsychologia*, *83*, 114–122. https://doi.org/10.1016/j.neuropsychologia.2015.08.021

Hoehl, S., Hellmer, K., Johansson, M., & Gredebäck, G. (2017). Itsy Bitsy Spider…: Infants React with Increased Arousal to Spiders and Snakes. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01710

Hoffman, D. D., & Singh, M. (2012). Computational Evolutionary Perception. *Perception*, *41*(9), 1073–1091. https://doi.org/10.1068/p7275

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye Tracker Data Quality: What It is and How to Measure It. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 45–52. https://doi.org/10.1145/2168556.2168563

Hunnius, S., Geuze, R. H., & van Geert, P. (2006). Associations between the developmental trajectories of visual scanning and disengagement of attention in infants. *Infant Behavior and Development*, *29*(1), 108–125. https://doi.org/10.1016/j.infbeh.2005.08.007

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of discrete entities. *Cognition*, *85*(3), 203–221. https://doi.org/10.1016/S0010-0277(02)00088-4

Hurley, K. B., Kovack-Lesh, K. A., & Oakes, L. M. (2010). The influence of pets on infants' processing of cat and dog images. *Infant Behavior and Development*, *33*(4), 619–628. https://doi.org/10.1016/j.infbeh.2010.07.015

Hyvärinen, A., Hurri, J., & Hoyer, P. O. (2009). *Natural image statistics: A probabilistic approach to early computational vision.* (Vol. 39). Springer Science & Business Media.

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development*, *67*(6), 2823–2840.

Isherwood, Z. J., Schira, M. M., & Spehar, B. (2017). The tuning of human visual cortex to variations in the 1/f α amplitude spectra and fractal properties of synthetic noise images. *NeuroImage*, *146*, 642–657. https://doi.org/10.1016/j.neuroimage.2016.10.013

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, *15*(4), 496–505.

Jones, P. R., Kalwarowsky, S., Atkinson, J., Braddick, O. J., & Nardini, M. (2014). Automated measurement of resolution acuity in infants using remote eye-tracking. *Investigative Ophthalmology & Visual Science*, *55*(12), 8102–8110.

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia*, *83*, 201–226.

Julesz, B. (1981). A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics*, *41*(2), 131–138.

Kastner, S., De Weerd, P., & Ungerleider, L. G. (2000). Texture segregation in the human visual cortex: A functional MRI study. *Journal of Neurophysiology*, *83*(4), 2453–2457.

Kavšek, M. J. (2003). Infants' perception of directional alignment of texture elements on a spherical surface. *Infant and Child Development*, *12*(3), 279–292. https://doi.org/10.1002/icd.295

Kavšek, M. J., Yonas, A., & Granrud, C. E. (2012). Infants' sensitivity to pictorial depth cues: A review and meta-analysis of looking studies. *Infant Behavior and Development*, *35*(1), 109–128. https://doi.org/10.1016/j.infbeh.2011.08.003

Kayhan, E., Hunnius, S., O'Reilly, J. X., & Bekkering, H. (2019). Infants differentially update their internal models of a dynamic environment. *Cognition*, *186*, 139–146.

Kayser, C., Körding, K. P., & König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current Opinion in Neurobiology*, *14*(4), 468–473.

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. https://doi.org/10.1016/j.visres.2006.02.003

Kellman, P. J. (2001). Separating processes in object perception. *Journal of Experimental Child Psychology*, *78*(1), 84–97.

Kellman, P. J., & Arterberry, M. E. (2007). Infant Visual Perception. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0203

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*(2), 141–221.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*(4), 483–524.

Kelly, D. J., Duarte, S., Meary, D., Bindemann, M., & Pascalis, O. (2019). Infants rapidly detect human faces in complex naturalistic visual scenes. *Developmental Science*, *22*(6). https://doi.org/10.1111/desc.12829

Kimchi, R. (2015). The perception of hierarchical structure. *Oxford Handbook of Perceptual Organization*, 129–149.

Kimchi, R., Hadad, B., Behrmann, M., & Palmer, S. E. (2005). Microgenesis and ontogenesis of perceptual organization: Evidence from global and local processing of hierarchical patterns. *Psychological Science*, *16*(4), 282–290.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*, 238.

Knill, D. C., Field, D., & Kerstent, D. (1990). Human discrimination of fractal images. *JOSA A*, *7*(6), 1113–1123. https://doi.org/10.1364/JOSAA.7.001113

Knyazeva, M. G. (2013). Splenium of corpus callosum: Patterns of interhemispheric interaction in children and adults. *Neural Plasticity*, *2013*.

Köster, M., Castel, J., Gruber, T., & Kärtner, J. (2017a). Visual cortical networks align with behavioral measures of context-sensitivity in early childhood. *NeuroImage*, *163*, 413–418. https://doi.org/10.1016/j.neuroimage.2017.08.008

Köster, M., Haese, A., & Czernochowski, D. (2017b). Neuronal oscillations reveal the processes underlying intentional compared to incidental learning in children and young adults. *PLoS ONE*, *12*(8). https://doi.org/10.1371/journal.pone.0182540

Köster, M., Kayhan, E., Langeloh, M., & Hoehl, S. (2020). Making Sense of the World: Infant Learning From a Predictive Processing Perspective. *Perspectives on Psychological Science*, *15*(3), 562–571. https://doi.org/10.1177/1745691619895071

Komogortsev, O. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms. In *Proceedings of the 2010 Symposium on eye-tracking research & applications* (pp. 65–68). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1743682

Kovács, I. (2000). Human development of perceptual organization. *Vision Research*, *40*(10–12), 1301–1310.

Kovács, I., Kozma, P., Fehér, Á., & Benedek, G. (1999). Late maturation of visual spatial integration in humans. *Proceedings of the National Academy of Sciences*, *96*(21), 12204–12209.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Kulke, L., Atkinson, J., & Braddick, O. (2015). Automatic Detection of Attention Shifts in Infancy: Eye Tracking in the Fixation Shift Paradigm. *PLOS ONE*, *10*(12), e0142505.

Kwon, M.-K., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, *52*(4), 537–555. https://doi.org/10.1037/dev0000080

Läge, D., Egli, S., Riedel, M., Strauss, A., & Möller, H.-J. (2011). Combining the categorical and the dimensional perspective in a diagnostic map of psychotic disorders. *European Archives of Psychiatry and Clinical Neuroscience*, *261*(1), 3–10. https://doi.org/10.1007/s00406-010-0125-y

Landy, M., & Graham, N. (2004). Visual perception of texture. In *The visual neurociences* (pp. 1106–1118). MIT Press.

Laubrock, J., Engbert, R., Rolfs, M., & Kliegl, R. (2007). Microsaccades are an index of covert attention: Commentary on Horowitz, Fine, Fencsik, Yurgenson, and Wolfe (2007). *Psychological Science*, *18*(4), 364–366. https://doi.org/10.1111/j.1467-9280.2007.01904.x

Lawrence, M. A. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments* (Version 4.4-0) [Computer software]. https://CRAN.R-project.org/package=ez

Leat, S. J., Yadav, N. K., & Irving, E. L. (2009). Development of Visual Acuity and Contrast Sensitivity in Children. *Journal of Optometry*, *2*(1), 19–26. https://doi.org/10.3921/joptom.2009.19

Leppänen, J. M., Forssman, L., Kaatiala, J., Yrttiaho, S., & Wass, S. (2015). Widely applicable MATLAB routines for automated analysis of saccadic reaction times. *Behavior Research Methods*, *47*(2), 538–548.

Leppänen, J. M. (2016). Using eye tracking to understand infants' attentional bias for faces. *Child Development Perspectives*, *10*(3), 161–165.

Lewis, T. L., & Maurer, D. (2005). Multiple sensitive periods in human visual development: Evidence from visually deprived children. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *46*(3), 163–183.

Lewis, T. L., Kingdon, A., Ellemberg, D., & Maurer, D. (2007). Orientation discrimination in 5-year-olds and adults tested with luminance-modulated and contrast-modulated gratings. *Journal of Vision*, *7*(4), 9–9.

Liben, L. S. (2003). Beyond point and shoot: Children's developing understanding of photographs as spatial and expressive representations. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 1–43). Elsevier Science.

LoBue, V. (2014). Deconstructing the snake: The relative roles of perception, cognition, and emotion on threat detection. *Emotion*, *14*(4), 701–711. https://doi.org/10.1037/a0035898

LoBue, V., & Adolph, K. E. (2019). Fear in infancy: Lessons from snakes, spiders, heights, and strangers. *Developmental Psychology*, *55*(9), 1889–1907. https://doi.org/10.1037/dev0000675

LoBue, V., & DeLoache, J. S. (2008). Detecting the Snake in the Grass: Attention to Fear-Relevant Stimuli by Adults and Young Children. *Psychological Science*, *19*(3), 284–289. https://doi.org/10.1111/j.1467-9280.2008.02081.x

LoBue, V., & DeLoache, J. S. (2010). Superior detection of threat-relevant stimuli in infancy. *Developmental Science*, *13*(1), 221–228. https://doi.org/10.1111/j.1467-7687.2009.00872.x

LoBue, V., Rakison, D. H., & DeLoache, J. S. (2010). Threat Perception Across the Life Span: Evidence for Multiple Converging Pathways. *Current Directions in Psychological Science*, *19*(6), 375–379. https://doi.org/10.1177/0963721410388801

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, *17*(6), 20–20. https://doi.org/10.1167/17.6.20

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293–308. https://doi.org/10.1016/j.bandc.2008.08.019

Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, *19*(2), 263–296.

Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-Fit Assessment in Multidimensional Scaling and Unfolding. *Multivariate Behavioral Research*, *51*(6), 772–789. https://doi.org/10.1080/00273171.2016.1235966

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114. https://doi.org/10.1016/j.tics.2004.01.003

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*(1), 3–36.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, *8*(3), 291–318.

Mandler, J. M., & McDonough, L. (1998a). On developing a knowledge base in infancy. *Developmental Psychology*, *34*(6), 1274.

Mandler, J. M., & McDonough, L. (1998b). Studies in inductive inference in infancy. *Cognitive Psychology*, *37*(1), 60–96.

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*(10), 443–450.

Markman, E. M. (1989). Categorization and naming in children: Problems of induction. Mit Press.

Markman, E. M., Cox, B., & Machida, S. (1981). The standard object-sorting task as a measure of conceptual organization. *Developmental Psychology*, *17*(1), 115.

Marr, D. (1976). Early processing of visual information. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 275(942), 483–519.

Mash, C. (2006). Multidimensional shape similarity in the development of visual object classification. *Journal of Experimental Child Psychology*, *95*(2), 128–152.

Matan, A., & Carey, S. (2001). Developmental changes within the core of artifact concepts. *Cognition*, *78*(1), 1–26.

Maurer, D., & Lewis, T. L. (2013). Sensitive Periods in Visual Development. *The Oxford Handbook of Developmental Psychology, Vol. 1*. https://doi.org/10.1093/oxfordhb/9780199958450.013.0008

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*(1), 89–115.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785. https://doi.org/10.1016/j.neuropsychologia.2007.10.005

Mondloch, C. J., Geldart, S., Maurer, D., & de Schonen, S. (2003). Developmental changes in the processing of hierarchical shapes continue into adolescence. *Journal of Experimental Child Psychology*, *84*(1), 20–40. https://doi.org/10.1016/S0022-0965(02)00161-3

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science*, *10*(5), 419–422.

Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (2012). A Critical Test of Temporal and Spatial Accuracy of the Tobii T60XL Eye Tracker: *Infancy*, *17*(1), 9–32. https://doi.org/10.1111/j.1532-7078.2011.00089.x

Morrone, M. C., & Burr, D. C. (1986). Evidence for the existence and development of visual inhibition in humans. *Nature*, *321*(6067), 235–237.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*(7141), 206–209. https://doi.org/10.1038/nature05724

Murtagh, F., & Legendre, P. (2011). Ward's hierarchical clustering method: Clustering criterion and agglomerative algorithm. *ArXiv Preprint ArXiv:1111.6285*.

Nassi, J. J., & Callaway, E. M. (2009). Parallel Processing Strategies of the Primate Visual System. *Nature Reviews. Neuroscience*, *10*(5), 360–372. https://doi.org/10.1038/nrn2619

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383.

Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition*, *80*(3), B11–B20.

Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development*, *1*(2), 131–156.

Nguyen, S. P., & Gelman, S. A. (2002). Four and 6-year olds' biological concept of death: The case of plants. *British Journal of Developmental Psychology*, *20*(4), 495–513.

Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2017). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, 1–15. https://doi.org/10.3758/s13428-017-0863-0

Nieuwenhuis, R., Te Grotenhuis, H. F., & Pelzer, B. J. (2012). Influence. ME: tools for detecting influential data in mixed effects models.

Norcia, A. M., Tyler, C. W., & Hamer, R. D. (1990). Development of contrast sensitivity in the human infant. *Vision Research*, *30*(10), 1475–1486.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018). Tests of an exemplar-memory model of classification learning in a high-dimensional natural-science category domain. *Journal of Experimental Psychology: General*, *147*(3), 328–353. https://doi.org/10.1037/xge0000369

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2017). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, 1–27.

Nyström, M., Hooge, I., & Andersson, R. (2016). Pupil size influences the eye-tracker signal during saccades. *Vision Research*, *121*, 95–103. https://doi.org/10.1016/j.visres.2016.01.009

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939.

Oakes, L. M. (2012). Advances in Eye Tracking in Infancy Research. *Infancy*, *17*(1), 1–8. https://doi.org/10.1111/j.1532-7078.2011.00101.x

Oakes, L. M., Kannass, K. N., & Shaddy, D. J. (2002). Developmental changes in endogenous control of attention: The role of target familiarity on infants' distraction latency. *Child Development*, *73*(6), 1644–1655.

Ofen, N., & Shing, Y. L. (2013). From perception to memory: Changes in memory systems across the lifespan. *Neuroscience & Biobehavioral Reviews*, *37*(9), 2258–2267.

Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proceedings of the National Academy of Sciences*, *112*(4), E351–E360. https://doi.org/10.1073/pnas.1415146112

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.

Olshausen, B. A., & Field, D. J. (1996). Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, *7*(2), 333–339.

Opfer, J. E. (2002). Identifying living and sentient kinds from dynamic information: The case of goal-directed versus aimless autonomous movement in conceptual change. *Cognition*, *86*(2), 97–122.

Opfer, J. E., & Gelman, S. A. (2011). Development of the animate-inanimate distinction. In U. C. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (Vol. 2, pp. 213–238). Blackwell Publishers Ltd.

Orquin, J. L., Ashby, N. J. S., & Clarke, A. D. F. (2016). Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research. *Journal of Behavioral Decision Making*, *29*(2–3), 103–115. https://doi.org/10.1002/bdm.1867

Oudeyer, P.-Y., & Smith, L. B. (2016). How Evolution May Work Through Curiosity-Driven Developmental Process. *Topics in Cognitive Science*, *8*(2), 492–502. https://doi.org/10.1111/tops.12196

Panis, S., De Winter, J., Vandekerckhove, J., & Wagemans, J. (2008). Identification of everyday objects on the basis of fragmented outline versions. *Perception*, *37*(2), 271–289.

Pauen, S. (2002). Evidence for knowledge–based category discrimination in infancy. *Child Development*, *73*(4), 1016–1033.

Pauen, S., & Hoehl, S. (2015). Preparedness to Learn About the World: Evidence from Infant Research. In T. Breyer (Ed.), *Epistemological Dimensions of Evolutionary Psychology* (pp. 159–173). Springer New York. https://doi.org/10.1007/978-1-4939-1387-9_8

Pearson, J., & Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends in Cognitive Sciences*, *12*(9), 334–341.

Pelz, J. B., & Rothkopf, C. (2007). Oculomotor behavior in natural and man-made environments. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements* (pp. 661–676). Elsevier.

Perrinet, L. U., & Bednar, J. A. (2015). Edge co-occurrences can account for rapid categorization of natural versus animal images. *Scientific Reports*, *5*, 11400.

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-Effects Models in S and S-Plus*, 3–56.

Pirchio, M., Spinelli, D., Fiorentini, A., & Maffei, L. (1978). Infant contrast sensitivity evaluated by evoked potentials. *Brain Research*.

Portilla, J., & Simoncelli, E. P. (2000). A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*, *40*(1), 49–70. https://doi.org/10.1023/A:1026553619983

Potts, R. (2012). Evolution and Environmental Change in Early Human Prehistory. *Annual Review of Anthropology*, *41*(1), 151–167. https://doi.org/10.1146/annurev-anthro-092611-145754

Putzar, L., Hötting, K., Rösler, F., & Röder, B. (2007). The development of visual feature binding processes after visual deprivation in early infancy. *Vision Research*, *47*(20), 2616–2626.

Quinn, P. C. (2011). Born to categorize. In U. C. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 129–152). Blackwell Publishers Ltd.

Quinn, P. C., & Eimas, P. D. (2000). The emergence of category representations during infancy: Are separate perceptual and conceptual processes required? *Journal of Cognition and Development*, *1*(1), 55–61.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rakison, D. H., & Derringer, J. (2008). Do infants possess an evolved spider-detection mechanism? *Cognition*, *107*(1), 381–393. https://doi.org/10.1016/j.cognition.2007.07.022

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological Bulletin*, *127*(2), 209.

Rakison, D. H., & Woodward, A. L. (2008). New perspectives on the effects of action on perceptual and cognitive development. *Developmental Psychology*, *44*(5), 1209–1213. https://doi.org/10.1037/a0012999

Rakison, D. H., & Yermolayeva, Y. (2010). Infant categorization. *WIREs Cognitive Science*, *1*(6), 894–905. https://doi.org/10.1002/wcs.81

Rao, A. R., & Lohse, G. L. (1993). Identifying high level features of texture perception. *CVGIP: Graphical Models and Image Processing*, *55*(3), 218–233.

Rao, A. R., & Lohse, G. L. (1996). Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research*, *36*(11), 1649–1669.

Redies, C., Hasenstein, J., & Denzler, J. (2007). Fractal-like image statistics in visual art: Similarity to natural scenes. *Spatial Vision*, *21*(1–2), 137–148.

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, *44*(19), 2301–2311.

Renswoude, D. R. van, Raijmakers, M. E. J., Koornneef, A., Johnson, S. P., Hunnius, S., & Visser, I. (2018). Gazepath: An eye-tracking analysis tool that accounts for individual differences and data quality. *Behavior Research Methods*, *50*(2), 834–852. https://doi.org/10.3758/s13428-017-0909-3

Reynolds, G. D., Courage, M. L., & Richards, J. E. (2013). The Development of Attention. *The Oxford Handbook of Cognitive Psychology*. https://doi.org/10.1093/oxfordhb/9780195376746.013.0063

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*(3), 244–274. https://doi.org/10.1016/j.cogpsych.2009.05.001

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Rothbart, M. K., Sheese, B. E., Rueda, M. R., & Posner, M. I. (2011). Developing mechanisms of self-regulation in early life. *Emotion Review*, *3*(2), 207–213.

Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research*, *37*(23), 3385–3398.

Ruff, H. A., & Rothbart, M. K. (1996). *Attention in early development: Themes and variations*. Oxford University Press.

Ruff, H. A., & Rothbart, M. K. (2001). *Attention in Early Development*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195136326.001.0001

Saez de Urabain, I. R., Nuthmann, A., Johnson, M. H., & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research*, *134*, 43–59. https://doi.org/10.1016/j.visres.2016.10.015

Scherf, K. S., Behrmann, M., Kimchi, R., & Luna, B. (2009). Emergence of global shape processing continues through adolescence. *Child Development*, *80*(1), 162–177.

Schick, K. D., & Toth, N. P. (1994). Making silent stones speak: Human evolution and the dawn of technology. Simon and Schuster.

Schlegelmilch, K. (2012). Architectures of Particularities. In K. Pinther, L. Förster, & C. Hanussek (Eds.), *Afropolis: City media art* (English edition). Jacana.

Schlegelmilch, K., & Wertz, A. E. (2019). The effects of calibration target, screen location, and movement type on infant eye-tracking data quality. *Infancy*, *24*(4), 636–662.

Schlegelmilch, K., & Wertz, A. E. (2020). Grass and Gravel: Investigating visual properties preschool children and adults use when distinguishing naturalistic images. *PsyArXiv*. https://doi.org/10.31234/osf.io/tgmd3

Schlegelmilch, K., & Wertz, A. E. (2021). Visual segmentation of complex naturalistic structures in an infant eye-tracking search task [manuscript in preparation].

Schmidt, F., Hegele, M., & Fleming, R. W. (2017). Perceiving animacy from shape. *Journal of Vision*, *17*(11), 10–10. https://doi.org/10.1167/17.11.10

Schuppli, C., Graber, S. M., Isler, K., & van Schaik, C. P. (2016). Life history, cognition and the evolution of complex foraging niches. *Journal of Human Evolution*, *92*, 91–100. https://doi.org/10.1016/j.jhevol.2015.11.007

Şerban, P., Wilson, J. R. U., Vamosi, J. C., & Richardson, D. M. (2008). Plant Diversity in the Human Diet: Weak Phylogenetic Signal Indicates Breadth. *BioScience*, *58*(2), 151–159. https://doi.org/10.1641/B580209

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive Neuroscience*, *27*(4), 665–678. https://doi.org/10.1162/jocn_a_00733

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Accuracy and speed of material categorization in real-world images. *Journal of Vision*, *14*(9), 12–12. https://doi.org/10.1167/14.9.12

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In *The adapted mind: Evolutionary psychology and the generation of culture.* (pp. 495–532). Oxford University Press.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*(1), 1193–1216.

Sireteanu, R., & Rieth, C. (1992). Texture segregation in infants and children. *Behavioural Brain Research*, *49*(1), 133–139.

Sireteanu, R., Encke, I., & Bachert, I. (2005). Saliency and context play a role in infants' texture segmentation. *Vision Research*, *45*(16), 2161–2176. https://doi.org/10.1016/j.visres.2005.02.003

Sireteanu, R., Rettenbach, R., & Wagner, M. (2009). Transient preferences for repetitive visual stimuli in human infancy. *Vision Research*, *49*(19), 2344–2352. https://doi.org/10.1016/j.visres.2008.08.006

Siu, C., & Murphy, K. (2018). The development of human visual cortex and clinical implications. *Eye and Brain*, *Volume 10*, 25–36. https://doi.org/10.2147/EB.S130893

Skoczenski, A. M., & Norcia, A. M. (2002). Late maturation of visual hyperacuity. *Psychological Science*, *13*(6), 537–541.

Slater, A., Morison, V., & Somers, M. (1988). Orientation discrimination and cortical function in the human newborn. *Perception*, *17*(5), 597–602.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*(6), 246–251.

Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24–62.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166.

Smith, L. B. (1979). Perceptual development and category generalization. *Child Development*, 705–715.

Smuda, M. (1986). *Landschaft* (Vol. 2069). Suhrkamp.

Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, *13*(7), 310–318.

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research*, *76*, 31–42. https://doi.org/10.1016/j.visres.2012.10.012

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), 5–5. https://doi.org/10.1167/11.5.5

Taylor, G., Hipp, D., Moser, A., Dickerson, K., & Gerhardstein, P. (2014). The development of contour processing: Evidence from physiology and psychophysics. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00719

Tenenbaum, J. B. (1995). Learning the structure of similarity. In D. Touretzky, M. C. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems 8* (pp. 3–9).

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Tobii Studio User's Manual (2016). Version 3.4.5. Copyright Tobii Technology AB.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, *12*(2), 229–232.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391–412.

Truxaw, D., Krasnow, M. M., Woods, C., & German, T. P. (2006). Conditions under which function information attenuates name extension via shape. *Psychological Science*, *17*(5), 367–371.

Tummeltshammer, K. S., Mareschal, D., & Kirkham, N. Z. (2014). Infants' Selective Attention to Reliable Visual Cues in the Presence of Salient Distractors. *Child Development*, *85*(5), 1981–1994. https://doi.org/10.1111/cdev.12239

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*(2), 169.

Valuch, C., Pflüger, L. S., Wallner, B., Laeng, B., & Ansorge, U. (2015). Using eye tracking to test for individual differences in attention to attractive faces. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00042

Van den Boomen, C., Lamme, V. A. F., & Kemner, C. (2014). Parallel development of ERP and behavioural measurements of visual segmentation. *Developmental Science*, *17*(1), 1–10. https://doi.org/10.1111/desc.12093

van den Boomen, C., van der Smagt, M. J., & Kemner, C. (2012). Keep your eyes on development: The behavioral and neurophysiological development of visual mechanisms underlying form processing. *Frontiers in Psychiatry*, *3*, 16.

Van Gompel, R. P. (2007). Eye movements: A window on mind and brain. Elsevier.

VanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, *14*(5), 498–504.

Vinter, A., Puspitawati, I., & Witt, A. (2010). Children's spatial analysis of hierarchical patterns: Construction and perception. *Developmental Psychology*, *46*(6), 1621–1631. https://doi.org/10.1037/a0020615

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34–41.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172.

Wallis, T. S. A., Funke, C. M., Ecker, A. S., Gatys, L. A., Wichmann, F. A., & Bethge, M. (2017). A parametric texture model based on deep convolutional features closely matches texture appearance for humans. *Journal of Vision*, *17*(12), 5. https://doi.org/10.1167/17.12.5

Walther, D. B., & Shen, D. (2014). Nonaccidental Properties Underlie Human Categorization of Complex Natural Scenes. *Psychological Science*, *25*(4), 851–860. https://doi.org/10.1177/0956797613512662

Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in Control: Rapid Anticipation of Action Outcomes in a Gaze-Contingent Paradigm. *PLoS ONE*, *7*(2), e30884. https://doi.org/10.1371/journal.pone.0030884

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge: Further fractionations and an attempted integration. *Brain*, *110*(5), 1273–1296.

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and Precision: How Data Quality May Influence Key Dependent Variables in Infant Eye-Tracker Analyses. *Infancy*, *19*(5), 427–460. https://doi.org/10.1111/infa.12055

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*(1), 229–250. https://doi.org/10.3758/s13428-012-0245-6

Wertz, A. E. (2019). How plants shape the mind. *Trends in Cognitive Sciences*, *23*(7), 528–531.

Wertz, A. E., & Wynn, K. (2014). Thyme to touch: Infants possess strategies that protect them from dangers posed by plants. *Cognition*, *130*(1), 44–49. https://doi.org/10.1016/j.cognition.2013.09.002

Wertz, A. E., & Wynn, K. (2014a). Thyme to touch: Infants possess strategies that protect them from dangers posed by plants. *Cognition*, *130*(1), 44–49. https://doi.org/10.1016/j.cognition.2013.09.002

Wertz, A. E., & Wynn, K. (2014b). Selective Social Learning of Plant Edibility in 6- and 18-Month-Old Infants. *Psychological Science*, *25*(4), 874–882. https://doi.org/10.1177/0956797613516145

Wertz, A. E., & Wynn, K. (2019). Can I eat that too? 18-month-olds generalize social information about edibility to similar looking plants. *Appetite*, *138*, 127–135. https://doi.org/10.1016/j.appet.2019.02.013

Westermann, G., & Mareschal, D. (2012). Mechanisms of developmental change in infant categorization. *Cognitive Development*, *27*(4), 367–382. https://doi.org/10.1016/j.cogdev.2012.08.004

White, B. J., Stritzke, M., & Gegenfurtner, K. R. (2008). Saccadic Facilitation in Natural Backgrounds. *Current Biology*, *18*(2), 124–128. https://doi.org/10.1016/j.cub.2007.12.027

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press, USA.

Wilcox, T., & Chapa, C. (2004). Priming infants to attend to color and pattern information in an individuation task. *Cognition*, *90*(3), 265–302. https://doi.org/10.1016/S0010-0277(03)00147-1

Wischnewski, M., Steil, J. J., Kehrer, L., & Schneider, W. X. (2009). Integrating inhomogeneous processing and proto-object formation in a computational model of visual attention. In *Human centered robot systems* (pp. 93–102). Springer.

Włodarczyk, A., Elsner, C., Schmitterer, A., & Wertz, A. E. (2018). Every rose has its thorn: Infants' responses to pointed shapes in naturalistic contexts. *Evolution and Human Behavior*, *39*(6), 583–593.

Yang, J., Kanazawa, S., Yamaguchi, M. K., & Motoyoshi, I. (2015). Pre-constancy Vision in Infants. *Current Biology*, *25*(24), 3209–3212. https://doi.org/10.1016/j.cub.2015.10.053

Yang, J., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., & Motoyoshi, I. (2011). Perception of surface glossiness by infants aged 5 to 8 months. *Perception*, *40*(12), 1491–1502.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.

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*(1), 480–483. https://doi.org/10.1196/annals.1325.049

Younger, B., & Gotlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories? *Developmental Psychology*, *24*(5), 611.

Yoshida, H., & Smith, L. B. (2008). What's in View for Toddlers? Using a Head Camera to Study Visual Experience. *Infancy : The Official Journal of the International Society on Infant Studies*, *13*(3), 229–248. https://doi.org/10.1080/15250000802004437

Zachariou, V., Del Giacco, A. C., Ungerleider, L. G., & Yue, X. (2018). Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *Journal of Vision*, *18*(12), 3–3.

Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, *44*(1), 109–123. https://doi.org/10.1016/S0167-9473(03)00030-6

Zhu, S.-C., Guo, C., Wang, Y., & Xu, Z. (2005). What are Textons? *International Journal of Computer Vision*, *62*(1–2), 121–143. https://doi.org/10.1023/B:VISI.0000046592.70770.61

Zihl, J., & Dutton, G. N. (2015). Development and Neurobiological Foundations of Visual Perception. In Cerebral Visual Impairment in Children (pp. 11–49). Springer Vienna. https://doi.org/10.1007/978-3-7091-1815-3_2

Zujovic, J., Pappas, T. N., & Neuhoff, D. L. (2013). Structural texture similarity metrics for image analysis and retrieval. *IEEE Transactions on Image Processing*, *22*(7), 2545–2558.