JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Fachbereich 06
Psychologie und Sportwissenschaft

Looking beyond the average: Human gaze behaviour towards natural scenes in special populations and across the lifespan

Synopsis zur kumulativen Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften
vorgelegt von Marcel Linka

Betreuer und Erstgutachter: Prof. Dr. Benjamin de Haas

Zweitbetreuerin: Prof. Dr. Jutta Billino

Acknowledgments

Over the past few years, many people have been involved in the projects I present here. I want to thank all my collaborators for making these projects possible and for their numerous discussions and valuable input throughout the process.

I feel very fortunate to be part of this wonderful group of enthusiastic researchers from all over the world who are part of our department and associated with it. Thank you all for the great talks, feedback, and for your contagious curiosity and passion for science.

A special thanks goes to:

Ben: Thank you for giving me this opportunity and trusting me to be part of this group from the very first day. Your guidance in teaching me to ask the right questions and your infectious joy for science have been truly inspiring. Thank you for the countless hours of laughter, brainstorming, and exploring research ideas together. For teaching me not to fear mistakes and giving me the freedom to pursue my interests, always being there for help, feedback, and questions—whether related to research or personal matters. All of this has made my PhD a wonderful experience, that has had, and continues to have, a profound impact on me. Thank you for everything!

The Indivisual lab: Thank you, **Diana**, **Max**, and **Petra**, for all the great moments. For hiking through the Dolomites, fun barbecues (and research talks) in Florida, and countless hours of laughter, sharing self-doubt, and doing sport-challenges. For keeping me motivated and being there through the ups and downs. For all the discussions about science and for being the cool group of people you are. It is and has been a great honor working with you.

Mum and Kevin: For being there for me no matter what. For encouraging me to pursue what interests me the most and for teaching me perseverance and love.

Jonah: For being the most beautiful human being you are. For the joy you spread with every part of you and for teaching me every day what really matters.

And last, but most importantly, **Lulu**: Thank you for supporting me in everything I do. For your patience, cleverness, and endless optimism. For your understanding and support during the tough times, your talent for finding clarity in chaos, and your boundless and contagious joy in everything you do. I am incredibly lucky to have you by my side.

Summary

Every day, we encounter an overwhelming amount of visual information. However, our visual system is capable of processing only the central area of the visual field with high acuity. This requires us to constantly move our eyes in order to select regions of interest. Which factors influence where we look? For more than a century, empirical research has investigated this question from various perspectives. The key findings indicate that gaze behaviour is mostly directed by (high-level) features of a scene, our goals, and visuo-spatial biases. A common aspect across most of these studies is that they do not explore potential differences among observers. However, recent findings suggest that gaze is highly individual and different among (age) groups. This thesis focuses on the observer, examining how free viewing behaviour towards natural scenes differs between observers, and investigates the effects of face recognition ability and age on these differences.

To this end, the first study found that free viewing behaviour towards natural scenes is atypical in Super Recognizers (SRs) – individuals with superior face recognition skills. Compared to controls, SRs showed a stronger tendency to focus on faces but looked less towards touched objects and text. When looking at faces, SRs focused less on the mouth but more on a point just below the eyes, which is considered optimal for face identification. Together, these findings suggest that face recognition ability is related to a visual preference for faces and saccadic landing positions close to the theoretical optimum within them. The second study first replicated large and stable individual differences in gaze behaviour towards objects of multiple semantic categories in an ad-hoc sample of adult observers. It then validated a short test, showing that such fixation tendencies can be estimated with a subset of only 40 as opposed to 700 natural scenes. In study 3, we examined the development of semantic gaze biases and applied this short test to probe them in preschoolers. Compared to adults, children looked substantially less towards text, but more towards touched objects and faces, suggesting that text preference may be guided by levels of literacy and reading experience. Finally, to track the development of gaze behaviour across the lifespan, study 4 applied the short test in a large and age diverse sample. We found that the semantic gaze tendencies in children demonstrated in study 3 undergo a protracted development taking ~ 20 years to become adult-like. Moreover, we showed that the central and horizontal biases unfold until early adulthood. Interestingly, while visual exploration showed a protracted increase, the fixation patterns became less idiosyncratic and more canonical across adolescence.

Together, the presented studies show that averages can obscure systematic variation across observers, when studying gaze behaviour, which may inform us about underlying mechanisms. I conclude that gaze towards high-level object attributes is highly individual across the population and atypical in certain cohorts, covering a spectrum of semantic gaze biases. Based on the findings of study 1, I speculate on a mediating role of face recognition ability in shaping fixation biases towards faces. Studies 3 and 4 suggest that gaze behaviour is highly plastic and perhaps affected by changes in visual diet and literacy that mutually reinforce each other. Interestingly, recent findings on cortical recycling in higher visual cortex nicely converge with the protracted development of viewing behaviour presented here. This suggests a link between the development of gaze behaviour, visual diet and cortical recycling that may be addressed by future research.

Contents

I. Synopsis

1. Introduction

Every day, we experience a sheer endless stream of rich visual information. To illustrate this, imagine riding your bike down a busy street alongside a friend. As you proceed, you observe people passing you by, cars, stores and advertisements lining the roadside, and much more. This gives you the impression that you are experiencing a vivid and objective view of the world. Then, your friend points out a stop sign that is right in front of you. To your surprise, you had not noticed it at all.

This reaction might seem familiar to many. We often believe that our perception of the visual world is both complete and objective and it surprises us when we notice it is not, even though we were convinced something was 'in plain sight'. Indeed, there is a strong scientific consensus that human visual perception is limited: For instance, our eyes are sensitive to only a narrow spectrum of electromagnetic radiation (Sliney, 2016); perceptual judgements of object size, color and meaning can be tricked by visual illusions (Coren & Girgus, 2020); and perhaps most related to the initial example, unlike a photographic camera, our eyes do not capture an image with uniform visual detail. Instead, we experience the highest level of acuity in the central region of our visual field, known as the fovea - a region of the retina corresponding to about the central 1.5° of our visual field. This is due to the dense concentration of cone photoreceptors in the fovea, which decreases towards the periphery (Curcio et al., 1990). Consequently, the spatial resolution drops continuously with increasing visual eccentricity. This difference in spatial resolution is not only due to variations in retinal processing but also influenced by the visual cortex's preferential treatment of foveal information (Dougherty et al., 2003; Dumoulin & Wandell, 2008; Harvey & Dumoulin, 2011; Horton & Hoyt, 1991). Therefore, in order to accurately perceive objects of interest, like the stop sign, we need to constantly move our eyes from one region to another. These quick eye movements are called *saccades*. During a *fixation*, the focus remains mainly stationary[1] (Bahill & Troost, 1979; Kowler, 2011). Due to this organization, the visual system is posed with a difficult challenge: Each moment in time it has to decide where to move our eyes next, preferring some areas over others. This raises a natural question: Which factors control where we look at?

1.1. Where we look: eye movements towards natural scenes

With the advent of reliable methods for measuring eye movements in the late 19th century (Huey, 1898; Orschansky, 1899), this question has become central to an increasing body of studies. Buswell (1935) marked the first systematic investigation of gaze behaviour during (artwork) scene perception. He discovered that fixations were consistently attracted by the content of the images, like depicted people. Three decades later, Yarbus (1967) followed up on this, showing that eye movements while viewing a painting depend on the viewer's intent or question in mind.

The foundational work by Yarbus (1967) and Buswell (1935) was followed by a steadily increasing number of studies on gaze behaviour towards natural environments and its influencing factors. Typically, these factors are divided into two broad levels of visual processing: top-down processes and bottom-up processes (Schütz et al., 2011; Theeuwes, 2010). Top-down processes refer to the effect of distinct tasks, goals and instructions on eye movements. Here,

[1] Please note that this thesis primarily focusses on saccades and fixations. However, there are other types of eye movements as well, such as smooth pursuit, micro saccades, and vergence movements (Bahill & Troost, 1979).

interesting lines of research have shown the effect of task demands on eye movements during visual search (Castelhano, 2008; Eckstein, 2011), basic memory tasks (Henderson, 2008) and active everyday activities like making tea (Land et al., 1999), or sports (Hayhoe et al., 2005; Hayhoe & Ballard, 2005; Land & McLeod, 2000). Together, these studies show that eye movements are driven by task demands during goal-oriented activities. Bottom-up processes in contrast emerge from the characteristics of the visual stimulus itself. Here, interesting research programs include the influence of object-context relationship (De Graef et al., 1990; Henderson et al., 1999; Võ et al., 2019) and scene gist (Hillstrom et al., 2012; Oliva, 2005) on eye movements. One influential area of research in this context is low-level visual salience. Salience models are formulated as computational models and are based on the premise that salient, low-level scene characteristics such as the local contrast of color, contrast, and orientation direct our attention and eye movements (Itti et al., 1998; Niebur & Koch, 1995). These models draw inspiration from the selective processing of visual features across the early visual cortex hierarchy. While there is clear evidence that salience models predict gaze behaviour well above chance (Betz et al., 2010; Itti & Koch, 2000; Tatler & Vincent, 2009), more recent studies show that high-level factors like object location (Stoll et al., 2015) and meaning maps (e.g., scene regions rated as informative; Henderson & Hayes, 2017) attain at least comparable predictive accuracy. When considering semantic object attributes, they even outweigh low-level features (Kuemmerer et al., 2017; Xu et al., 2014). Certain objects, such as faces and text (Wang & Pomplun, 2012) have been identified as particular strong predictors of gaze, even overriding task-induced (top-down) factors (Cerf et al., 2009). This thesis has a strong emphasis on such high-level features and their role in directing eye movements. It is important to note, however, that aforementioned top-down and bottom-up factors do not operate in isolation, but rather in an interactive manner. For instance, simple top-down directed visual search tasks (Einhäuser et al., 2008) can fully diminish the influence of low-level, however not high-level salience (Cerf et al., 2009). Moreover, top-down and bottom-up factors do not seem to attract eye movements in the same way across time. While first fixations are believed to be mainly driven by bottom-up factors such as low-level salience, subsequent fixations might be influenced by goal-directed, top-down factors (Anderson et al., 2015; Mackay et al., 2012; Schütt et al., 2019).

Beyond the dichotomy of top-down and bottom-up visual processing, others have demonstrated that gaze behaviour towards scenes is affected by general spatial biases, which are thought to be somewhat independent of image characteristics and task-related (top-down) factors. For example, individuals tend to focus on the central regions of images, a phenomenon referred to as the central bias (Tatler et al., 2005; Tatler & Vincent, 2008). Although the central bias may be partially attributed to a photographer's tendency to place objects of interest at the center of the image (Bindemann, 2010), it cannot be entirely explained by this (Foulsham & Underwood, 2008; Tseng et al., 2009). Furthermore, saccadic eye movements are not evenly distributed in all directions. Rather, individuals tend to make more horizontal saccades than vertical or oblique ones (Foulsham et al., 2008). Whether the horizontal bias is completely explained by the horizontal structure of scenes is not yet fully understood. While the primary gaze direction aligns with the scene's horizon at different scene orientations (Foulsham et al., 2008), a horizontal bias has also been observed in abstract fractal images that lack a conventional orientation (Foulsham & Kingstone, 2010).

Together, the mentioned studies suggest that which regions we fixate in our natural environment is affected by a combination of temporary goals, visual characteristics of scene regions and general spatial biases. All these studies

have one thing in common: They typically aggregate data from multiple adult observers in order to test hypotheses on the very visual and neural processes of interest. As a consequence, they often treat differences among observers as a mere source of measurement error. Thus, while these studies may explain why my friend and I *both* looked at the red car and people passing us by, they do not directly explain why I did not see the stop sign, while my friend did. Of course, it might have just been pure coincidence or driven by some diverging goals at that moment in time. Yet, evidence suggests that some observers show systematic gaze tendencies that largely differ from those of others (e.g. Castelhano & Henderson, 2008; Jones et al., 2008). May thus *stable* observer tendencies be a possible explanation? The further study of such observer-related differences could uncover variations in important neural and cognitive processes related to vision, offering valuable insights into fundamental mechanisms of visual perception. This thesis focusses on the observer, asking the question: What insights can we gain into the visual system by considering the observer as a factor of interest?

1.2. The role of the observer in natural scene viewing

As noted above, for most of the research on eye movements towards complex scenes, the observer factor has played a less important role and mostly been modelled as random 'noise'. It is all the more surprising that even the pioneer work of Buswell (1935) already mentioned substantial differences in gaze behaviour between observers when describing a painting with overlaid scanpaths of multiple observers :

'The fact that the density plots represent composite diagrams from a great many different subjects obviously results in a wider distribution of fixations on account of the varied interests of different individuals in looking at the same picture' – Buswell (1935)

This observation could later be empirically confirmed, showing large and also robust individual differences in basic oculomotor metrics like mean saccadic amplitude and fixation duration during scene viewing (Castelhano & Henderson, 2008; Henderson & Luke, 2014). More recently, de Haas et al. (2019) reported systematic individual differences in gaze behaviour towards objects of multiple semantic dimensions like faces, text, motion, touched objects and those with an implied taste during free viewing of complex scenes. This was also true when only considering the first fixations towards a scene. Interestingly, they could show that this tendency to first fixate a face was associated with enhanced face recognition performance. For gaze towards faces, individuals reliably vary in their preferences when focusing along the vertical axis (Broda & de Haas, 2022; Peterson & Eckstein, 2013), and recent findings by Broda & de Haas (2024) suggest that this preference may generalize to other objects as well. Interestingly, according to a Foveated Ideal Observer model, the optimal landing position of face-directed saccades for processing facial identity is just below the eyes (Peterson & Eckstein, 2012). While observers gravitate towards this optimal point on average, individual deviations along the vertical axis seem to be individually optimal for face recognition performance (Peterson & Eckstein, 2013). This may be attributed to face-specific retinotopic tuning biases or domain-general idiosyncrasies, such as variations in the visual field's geometry (Broda & de Haas 2024; Peterson et al., 2016; Peterson & Eckstein, 2013).

Apart from reliable inter-individual variation in gaze behaviour, there is also evidence that specific groups of individuals differ from others in consistent ways. In recent years, studies reported atypical gaze behaviour in people with neurodevelopmental disorders. For instance, individuals with Autism Spectrum Disorder (ASD) show *reduced* attention to faces and eyes (Constantino et al., 2017; Jones et al., 2008; Klin & Jones, 2008; Wang et al., 2015), whereas those with Williams Syndrome (WS) display *increased* attention to these features (Riby et al., 2013). Further, people with schizophrenia exhibit less visual exploration than healthy controls (Bestelmeyer et al., 2006) and individuals with Developmental Prosopagnosia (DP) show reduced gaze towards faces (Bobak et al., 2017), along with a tendency to focus lower on the face (Peterson et al., 2019). This latter finding, as well as the work previously mentioned by Peterson et al. (2012, 2013) and de Haas et al. (2019), suggests an interesting relationship between visual face preferences (or the lack thereof) and corresponding recognition performance. To this point, only a limited number of studies have investigated eye movements towards natural scenes in special populations, and even fewer have delved into its possible relationship with recognition performance. However, what is common among these studies is that they all suggest a deeply rooted neurobiological basis for attentional biases towards semantic features. This is further corroborated by recent twin studies, which suggest that fixation biases towards social stimuli are evident from an early age and heritable (Constantino et al., 2017; Kennedy et al., 2017).

At the same time, there is also some evidence for developmental changes in gaze behaviour. While a gaze preference for faces is already present in infants (Constantino et al., 2017; Johnson et al., 1991, 2015), children develop further visual preferences towards other types of social stimuli, such as hands, early on (Frank et al., 2011). As children progress into later childhood and adulthood, gaze behaviour continues to evolve. Some work has reported enhanced gaze allocation towards simple, low-level salient features in early childhood that declines with age (Acik et al., 2010). Others showed that more basic oculomotor behaviours also seem to develop. The saccadic amplitude increases, and the duration of fixations decreases in later childhood (Helo et al., 2014). Further, the central bias is less pronounced at age 8 than at age 4 (Krishna et al., 2018) and while a horizontal bias is present in infants, it is less strong than in adults (Van Renswoude et al., 2016). Yet, our understanding of how gaze develops is still very limited. Although recent findings point to a protracted maturation of cortical regions selective to high-level visual features like faces, text and limbs (Kubota et al., 2023; Nordt et al., 2021), it is still an open question whether fixation behaviour towards these features changes in a similar manner. This is also due to the fact that studies often do not include children older than 3 years and usually only compare a limited number of age groups. However recent studies have shown that visual perception seems to take substantially longer to develop than previously thought. Carrasco et al. (2022, 2023) found that visual field asymmetries typically observed in adults are less pronounced in children and adolescents. Specifically, adults show stronger visual field asymmetries compared to children and adolescents, specifically a stronger enhancement of visual performance along the horizontal meridian, and the lower compared to the upper vertical meridian (Carrasco et al., 2022, 2023). The reasons for developmental changes in visual perception could be manifold. Some studies suggest that neural maturation of the ventral stream plays a central role in driving developmental changes in oculomotor behaviour (Helo et al., 2014) and object processing (Nordt et al., 2023). Others speculate that changing visual information—e.g., due to alterations in visual field statistics through changes in body height (Carrasco et al., 2022) or reading acquisition (Kubota et al., 2023) may shape visual perception. Of course,

both views are not mutually exclusive. Please note that the potential neural basis of high-level gaze biases will be explored in greater detail in the discussion section.

Together, research examining the observer role suggests that eye movements are influenced by large and stable inter-individual variation and distinct group-specific biases that may be related to visual processing performance. At the same time, gaze behaviour seems to differ across age groups. Nonetheless, studies that focus on fixations towards high-level object features in special populations are scarce, and our understanding of the link between fixation biases towards faces and face identity processing remains limited. Moreover, it remains an open question whether gaze towards high-level semantic features differ among age groups—and if so, how they and other oculomotor behaviours develop across the lifespan.

1.3. The present studies

This thesis examines the observer factor in free viewing behaviour towards natural scenes from different viewing angles. Specifically, it focusses on individual differences, special populations, and developmental changes to obtain a better understanding of which observer-specific factors influence where we move our eyes. Therefore, the first study investigates free viewing towards natural scenes in a unique group known as Super-Recognizers (SRs) - individuals with superior face processing abilities. Study 2 replicates the stable and large individual differences towards multiple semantic object dimensions reported by de Haas et al. (2019) and validates a short test designed to capture these differences more efficiently. Studies 3 and 4 apply this short test to examine fixation biases in pre-schoolers and to track its developmental trajectories in a large-scale dataset of visitors to a museum. In the following text, I will provide the motivation behind each study included in this thesis, followed by a more detailed summary of the findings in the subsequent chapters. Finally, I will discuss the findings and their possible implications in the last chapter. Please refer to Figure 1A for an outline of the free viewing task, which has been employed across all the

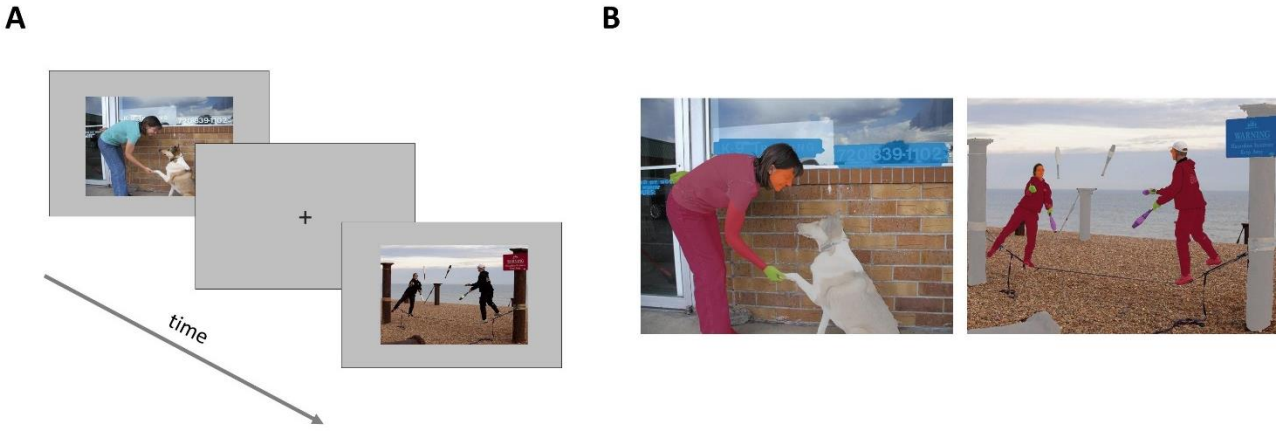

Figure 1. (A) gives an overview of the free viewing task as applied in all studies reported in this thesis. Images were displayed for 2 (study 1) or 3 (study 2-4) seconds each before a self-paced fixation cross appeared. The image size varied across the different studies. Within each study, the order of presented images was identical across groups and participants. (B) shows object pixel masks for two example images, with colors corresponding to the categories faces (orange), text (blue), touched objects (violet), hands (green) and bodies (red; objects outside these categories shown in light gray). These pixel masks come with corresponding semantic object labels and were used for determining which objects were fixated.

studies described here, with only minor differences in the number of images and duration of trials. See Figure 1B for example images including pixel masks used for determining regions of interest.

Study 1. As mentioned earlier, de Haas et al. (2019) found a positive correlation between the individual tendency to first fixate faces and face recognition performance. Further, people with DP have been shown to look less at faces (Bobak et al., 2017). Interestingly the same study did not find enhanced gaze behavior to faces among SRs. However given this study faces some methodological limitations, using only a small set of just 20 stimuli and identifying SRs through a single face recognition test, further research is necessary to test the relationship between face recognition performance and face preferences. Study 1 further addresses this hypothesis by examining gaze behaviour towards natural scenes among SRs, identified with a more conservative diagnostic framework (Ramon, 2021). We let SRs and controls freely view 700 images and tested whether they differ in fixation biases towards objects of four semantic dimensions: faces, text, objects being touched, and bodies. Moreover, Peterson & Eckstein (2012) showed that the ideal location for facial identity processing is slightly below the eyes. In study 1 we therefore also examined group differences in fixation biases towards faces. In particular, we tested differences in fixations towards the eyes and mouths as well as differences in fixation distance to the theoretically optimal point for face identification.

Study 2. The free viewing paradigm used in study 1, as well as in other studies (de Haas et al., 2019; Xu et al., 2014) included 700 images, equivalent to >1 hour of testing time, making data collection from vulnerable populations like children very difficult. Therefore, it would be desirable to estimate individual gaze biases using a more efficient test. Study 2 had two main objectives. The first was to replicate the consistent individual differences in fixation biases towards objects of several semantic object dimensions and visual exploration reported by de Haas et al. (2019). The second objective was to determine if these differences could be estimated with a shorter test. Therefore, 103 participants underwent two eye-tracking sessions. The first session replicated the original study, employing a free viewing task with 700 images. The second session used a smaller, optimized subset of images to efficiently capture individual differences.

Study 3. Atypical gaze behaviour towards high-level features in clinical populations suggest a deeply rooted neurobiological foundation (e.g. Bestelmeyer et al., 2006; Constantino et al., 2017). However, gaze behavior may also be shaped by visual experiences, such as learning to read. Previous studies have demonstrated that viewing behaviour in older children differs from that of adults in terms of low-level salience (Acik et al., 2010), center-bias (Krishna et al., 2018) and basic oculomotor behaviour (Helo et al., 2014). However, it is unknown whether such differences also exist for fixations towards (high-level) semantic object features. To test this, we applied the short test validated in study 2 to test free viewing behaviour in 5-year-old preschoolers and adults. Specifically, we investigated how children and adults differ in their fixation tendencies towards text, faces, touched objects, hands and bodies.

Study 4. Study 4 continued the developmental focus of study 3. While previous research has examined developmental differences in basic oculomotor behaviour and visuo-spatial biases, study 3 tested developmental differences across the semantic object features text, faces, touched objects and bodies. However, these studies all share one limitation: They typically compare only two or three age groups and rarely examine adolescence making it impossible to

precisely track developmental trajectories of gaze behaviour. Moreover, given their relatively small sample sizes, they often lack statistical power, increasing the likelihood of false negative results. To address this, study 4 aimed to track the developmental trajectories of semantic salience and basic oculomotor biases by collecting a large dataset across a broad age range. Therefore, we collaborated with a local science museum to collect data from over 10,000 participants freely viewing natural scenes in an eye-tracking booth. To examine how gaze behaviour changes throughout the lifespan, we analyzed fixation biases towards objects of various semantic categories (see study 3), alongside basic oculomotor measures such as the central and horizontal bias and visual exploration. Finally, we assessed the development of inter-observer similarities in gaze behaviour.

2. Study 1: Characteristic fixation biases in Super-Recognizers

By Linka, Broda, Alsheimer, de Haas & Ramon (Journal of Vision, 2022)

In study 1, we investigated gaze behaviour towards natural scenes in SRs - individuals with exceptional face identity processing skills. This study had two overarching goals. First, to test gaze biases in SRs towards objects across four semantic dimensions. Secondly, to probe their fixation tendencies within faces.

To this end we let 10 SRs and 43 control participants freely view 700 everyday scenes (Xu et al., 2014) divided into 7 sets of 100 images each. Each image was displayed for 2 seconds. Between trials, participants saw a self-paced fixation cross and could start the next trial by pressing the space bar.

First, we tested differences in the proportion of dwell time spent on objects for the semantic categories faces, text, touched objects, and bodies between SRs and controls. Compared to the control group, SRs allocated a significantly larger proportion of their viewing time to faces, and a smaller proportion to text and objects being touched. No significant differences were observed for the dwell time spent on bodies (Figure 2A). First fixations following the presentation of an image (i.e., the landing position of the first saccade) are believed to be affected mainly by bottom-up processing. Moreover, fixations directed at faces are more difficult to inhibit than those aimed at other types of stimuli (Crouzet et al., 2010; de Haas et al., 2019). Therefore, we also examined group differences in the proportion of first fixations across all four categories. SRs spent a significantly larger fraction of their first fixations towards faces and a lower fraction towards text and objects being touched. Again, no group differences were found for the dimension bodies (Figure 2A). Interestingly, although we found significant difference between SRs and controls, there was considerable overlap between both groups across all dimensions. Finally, we investigated a correlation between face recognition performance and the tendency to initially fixate on faces among SRs, which was not significant. However, this result should be interpreted with caution due to the small size of the SR group (n = 10) and reduced variance within this group, which suggests limited statistical power and an enhanced likelihood of Type II errors.

We than zoomed in on gaze behaviour towards faces, comparing the proportion of dwell time and first fixations on eyes and mouths between SRs and control participants. Our results showed that, in terms of both dwell time and first fixations on faces, SRs focused less on mouths compared to controls, with no significant difference in their gaze towards eyes. Finally, we probed whether SRs fixated closer to the optimal fixation point for identification, as

predicted by an ideal Bayesian Foveated Observer model (Peterson & Eckstein, 2012). Indeed, SRs fixated significantly closer to the estimated optimal fixation point, both when considering all face-directed fixations and only the first one (Figure 2B and C).

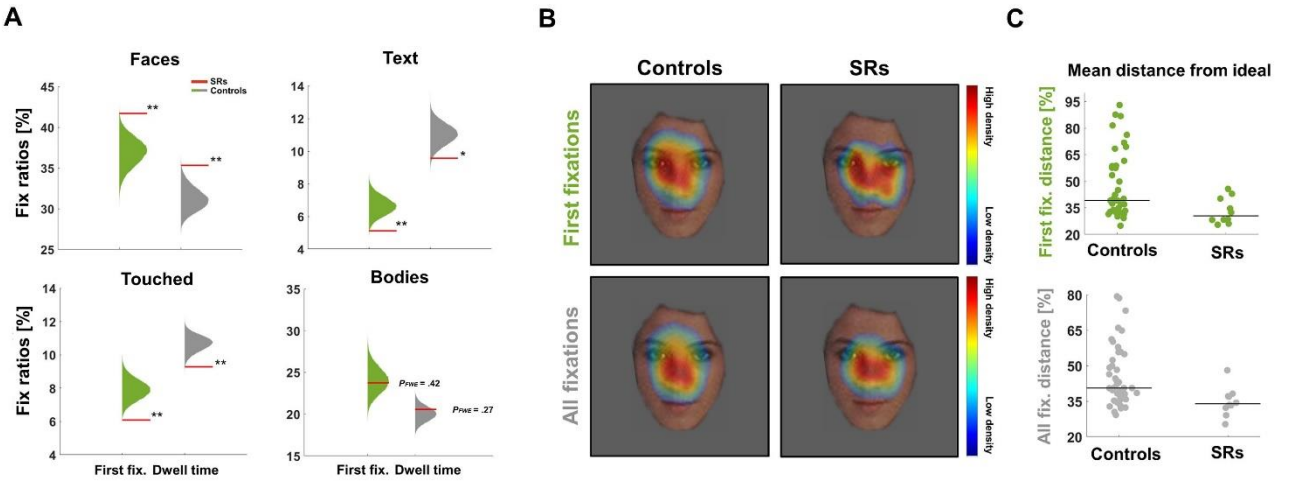

Figure 2. **Group differences in gaze tendency across four semantic dimensions and towards faces**. (A) Bootstrapped null distributions of 10,000 random sample means drawn from all participants (pooled across groups) for percent first fixations (green) and percent of cumulative dwell time (gray). The superimposed red lines refer to the observed mean percent first fixations and mean percent cumulative dwell time of SRs for each given semantic dimension. Significance levels: *p < 0.05, **p < 0.01 (Holm–Bonferroni corrected). (B) Heatmaps show the distribution of first fixations (top row) and all fixations (bottom row) on all observed faces with an eye-to-mouth distance > 2.5 degree visual angle (dva) superimposed over an example image (selected from one stimulus used in the free-viewing task). The heatmap was horizontally compressed to match the example face. Heatmaps on the left-hand side show fixation data from controls, maps on the right-hand side show fixations from SRs. All face fixation coordinates were normalized to the horizontal (X) and vertical (Y) extent of the respective observed face. Warmer colors indicate higher density of fixations. (C) Dot plots on the right-hand side show the average of relative distances from ideal fixation point for first fixations (top row) and all fixations (bottom row) towards faces with an eye to mouth distance >2.5 dva for each observer. Distances from the ideal fixation point just below the eyes (Peterson & Eckstein, 2012) were calculated for each face and fixation, normalized as percentages from the distance from the respective mouth centroid to the ideal fixation point and then averaged for each observer. The resulting mean distances for controls are shown on the left-hand side, those for SRs on the right-hand side.

Together, study 1 showed that when freely viewing complex scenes, SRs compared to controls exhibited increased gaze biases towards faces and reduced tendencies to fixate on touched objects and text. These differences were apparent from the first fixation on. Specifically, within the category of faces, SRs had a stronger tendency than controls to fixate near the theoretical optimum, just below the eyes.

3. Study 2: OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience

By Linka, & de Haas (Journal of Vision, 2020)

Study 2 had two main objectives. The first one was to replicate the finding of large and consistent individual differences in gaze behaviour towards objects of multiple semantic categories and visual exploration, as reported by

de Haas et al., (2019). The second objective was to validate a short test that captures these differences in a more time-efficient manner.

In total, 103 participants were recruited for two eye-tracking sessions two weeks apart. Similar to study 1 (and in de Haas et al., 2019), in the first session participants viewed 700 images across 7 blocks. Each image was displayed for 3 seconds. In the second session, participants viewed a subset of 200 images. This subset was selected using an iterative search algorithm on the dataset by de Haas et al., 2019. It selected scenes in a way that maximized the predictive validity for the entire stimulus set at each step (i.e., for every additional image from 10 to 200). To replicate the findings by de Haas et al. (2019), we used the data retrieved at the first session. Next, we determined the minimum set size of scenes needed to retrieve reliable estimates, by calculating the re-test correlations between individual gaze tendencies for the full set of images from the first and smaller subsets from the second session.

In a first step we successfully replicated the main findings of de Haas et al. (2019) using all 700 images. We showed that the proportion of cumulative dwell time, as well as that of first fixations on objects across the categories of faces, text, motion, taste, and touched objects exhibited large variation among participants. These variations showed high split-half correlations across 10.000 splits of image sets for each semantic category. Additionally, we found large and consistent individual differences in visual exploration, as indicated by the number of objects fixated. Next, we probed whether these differences could be estimated with smaller image subsets. A correlation of $r \geq .7$ was used as the critical threshold for good test-retest reliability. We found that for cumulative dwell times, a stimulus set of 40 images provided reliable estimates for the dimensions text and faces. Subsets of 100 and 200 images captured individual differences in text, faces, and motion. The dimensions taste and touched objects were approaching the reliability threshold for the 200-image set (Figure 3A). Regarding the proportion of first fixations, a set of 100 images was sufficient to capture reliable individual differences for text and faces while the dimension touched was just below the

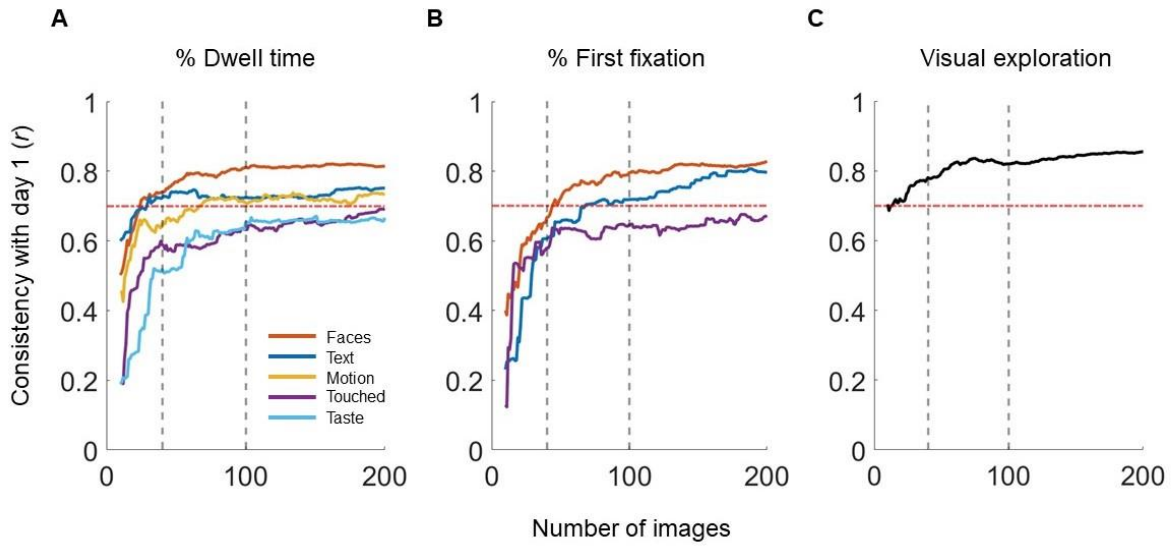

Figure 3. Consistency of fixation ratios based on smaller stimulus sets. (A) Line plots depicting Pearson consistency correlations of individual differences in percent cumulative fixation time, (B) percent first fixations, and (C) visual exploration (number of distinct objects fixated) between the full image set (day 1) and a given subset (10 to 200 images from day 2; x-axis). Dashed gray lines mark stimulus sets of 40 and 100 images. Dashed red lines indicate the consistency threshold of r = 0.7. Colors in (A) and (B) indicate semantic dimensions as shown in the inset.

critical threshold (Figure 3B). A set size of 40 images was sufficient to estimate individual differences in visual exploration (Figure 3C).

Together, study 2 corroborated large and consistent individual differences in gaze behaviour towards objects across various semantic categories and in visual exploration and showed that most of these differences can be estimated through a considerably shorter free viewing task.

4. Study 3: Free viewing biases for complex scenes in preschoolers and adults

By Linka, Sensoy, Karimpur, Schwarzer & de Haas (Scientific Reports, 2023)

Study 3 shifted the perspective to focus on the developmental aspects of gaze behaviour towards natural scenes. The main goal was to examine differences in free viewing behaviour towards semantic object dimensions between preschoolers (5-year-olds) and adults.

We let n = 34 5-year-old preschoolers and n = 42 adults freely view 40 scene images for 3 seconds each (cf. the short test validated in study 2). We then calculated the proportion of cumulative dwell time and the proportion of first fixations devoted to objects within the semantic categories of faces, text, touched objects, bodies, and hands depicted in the scenes.

Our findings revealed that compared to adults, preschoolers spent more dwell time on faces (though the effect was marginally significant), touched objects and hands, while substantially less on text. No difference was observed for the dimension bodies (Figure 4A). Consistent with the findings for dwell time, preschoolers allocated a larger proportion of first fixations to touched objects and hands, and a significantly smaller proportion to text than adults. Interestingly, however, children directed more first fixations towards faces and bodies compared to adults (Figure 4B).

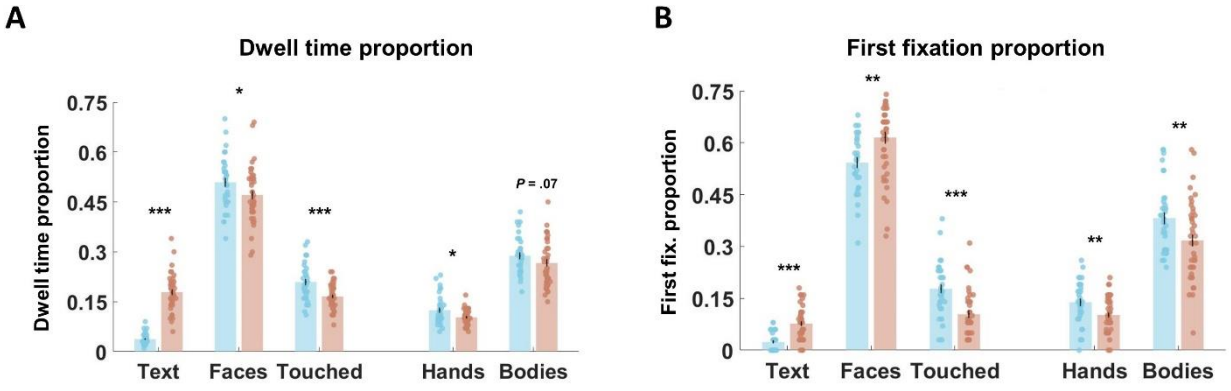

Figure 4. **Group differences between adults and children in gaze behaviour along five semantic dimensions**. (A) Bar plots showing the mean values as well as data points of cumulative dwell time proportion and (B) the proportion of first fixations towards objects of the dimensions faces, text, touched, hands and bodies respectively for children (blue) and adults (red). Data points indicate the individual mean values for a given dimension and group. Errorbars indicate +/- 1 standard error of the mean SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (Holm-Bonferroni corrected).

Next, we were interested in whether the reduced gaze in adults towards faces, touched objects and hands might be due to their strong bias towards text. To address this question, while also controlling for potential effects of lower-

level object features such as (object) size, eccentricity (object distance from the image centroid), and low-level visual salience (*Graph-Based Visual Saliency* ; Harel et al., 2006), we conducted generalized linear mixed models. The results showed that children only showed increased gaze preferences for person-related dimensions (except faces) when text was removed from the analysis. This suggests that the differences in gaze between children and adults are largely attributable to children's reduced focus on text.

In sum, we reported differences in gaze biases towards semantic object features between preschool children and adults. Follow-up analyses indicated that most body-related fixation biases in children can be attributed to the absence of competition with text observed in adults.

5. Study 4: Protracted development of gaze behaviour

By Linka, Karimpur & de Haas (PsyArXiv, 2024)

Study 4 followed up on the developmental approach of study 3. However, while study 3 looked at differences in gaze behaviour between two age groups (preschoolers and adults), here we asked: How does gaze behaviour towards natural scenes develop across the lifespan?

To answer this question, we collaborated with a local science museum and gathered gaze data from n = 10,632 participants between ages 5 to 72. Participants sat in an eyetracking booth and freely viewed 40 natural scenes, each for 3 seconds (cf. the stimulus set validated in study 2). Three main goals were of interest. First, following up on the findings of study 3, we wanted to trace the developmental trajectory of gaze biases towards objects of the categories faces, text, touched objects and bodies. Second, given evidence on developmental changes in basic visuo-spatial and oculomotor metrics such as the center bias (Krishna et al., 2018), the horizontal bias (Van Renswoude et al., 2016) and fixation frequency (Helo et al., 2014), we aimed to pinpoint their developmental trajectories more precisely. Finally, we wanted to test the development of inter-individual differences in gaze behaviour.

We found that the gaze behaviour towards objects of several semantic categories was surprisingly protracted. Figure 5 shows the developmental trajectories for dwell time proportion (A) and first fixation proportions (B). The proportion of dwell time and first fixations directed towards text continuously increased before stabilizing during the third decade of life. At the same time the tendency to fixate touched objects continuously decreased until early adulthood at age 19-20 for the proportion of dwell time and until age 17-18 for the proportion of first fixation. Gaze towards faces developed over the first two decades of life. Interestingly, in line with the findings of study 3, while the proportion of dwell time spent on faces showed a steady *decrease* until age 15-16, the proportion of first fixation continuously *increased* until age 17-18. We did not find any developmental changes for gaze behaviour towards bodies.

Apart from high-level aspects of gaze, we also found visuo-spatial biases to continuously develop until adulthood. Specifically, we showed an increase in horizontal bias until age 15-16 (Figure 6A). As reading typically goes along with small horizontal saccades, we tested a possible correlation between fixation bias towards text and the horizontal bias, which we could not confirm when controlling for age. Moreover, we observed a continuous decline of the

central bias until age 20-21 and a parallel increase in fixation frequency and visual exploration, indicated by the number of objects fixated, that plateaued at age 22-23.

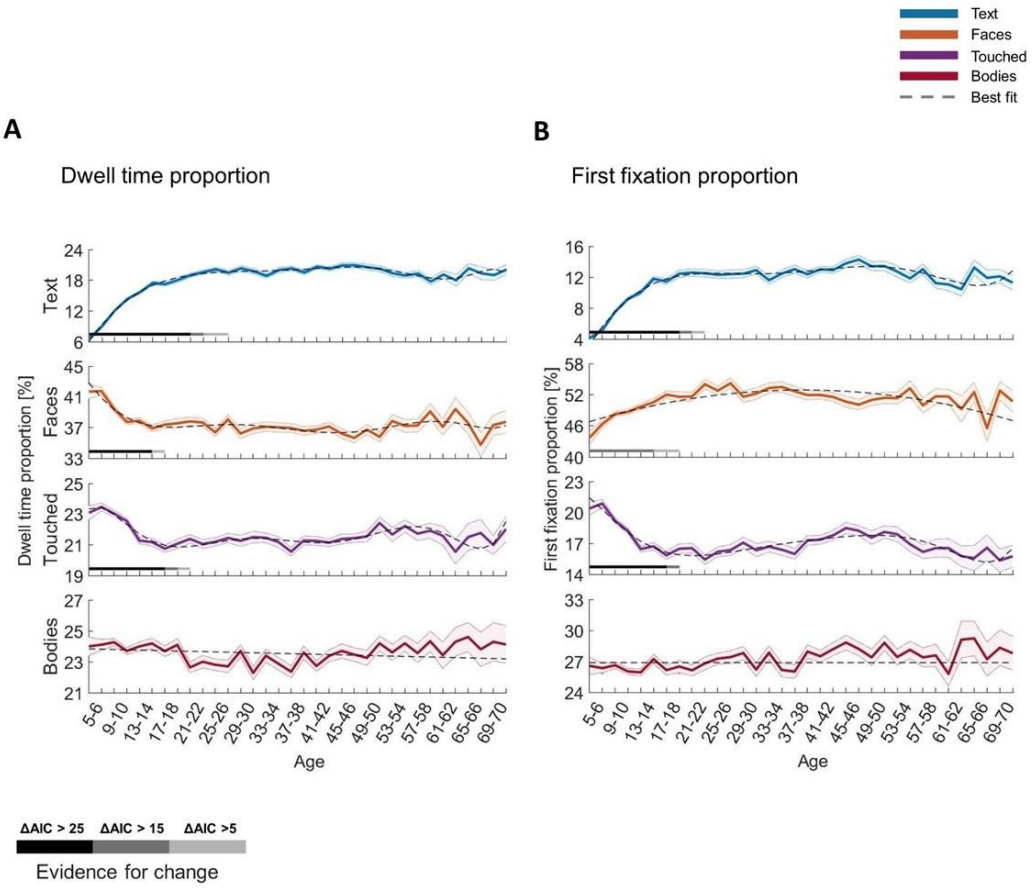

Figure 5. **Protracted development of semantic salience**. Line plots display the mean proportion of (A) dwell time and (B) first fixations directed towards four semantic features across age bins. Semantic categories are indicated by colours, as shown in the inset. Error shades indicate +/- 1 standard error of the mean (SEM). Dotted lines show the corresponding best fitting polynomials. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. Linka et al., 2024).

Finally, we examined the development of inter-individual similarity in gaze patterns. First, we computed individual distributions of dwell time across all depicted objects and then calculated the average pairwise correlation between observers, separately for each age bin. The pairwise observer similarities first somewhat decreased between ages 5-6 and 13-14, before showing a strong linear increase until age 23-24. Second, we calculated the Shannon entropy of fixation maps for each given image and then averaged them for each age group, providing us with a complementary measure for inter-observer variance. In line with the findings above, this group-level entropy showed a strong increase until age 11-12, before decreasing until age 25-26, and a second more gradual decrease from age 47-48 onwards (Figure 6B). Additionally, we computed the *intra*individual entropy of individual fixation maps and averaged it for each age group, an indicator for individual visual dispersion. In line with our findings on visual exploration, intraindividual entropy increased until the age of 37-38, before steadily declining (Figure 6C). Thus, visual exploration strongly increased during the first two decades of life. Interestingly, this increase in individual exploration was accompanied by a small initial decrease but subsequent increase in inter-observer similarity during adolescence.

This indicates that from age 13-14 onwards, gaze distributions became both more dispersed at the individual level and at the same time more canonical across observers (cf. Figure 6D).

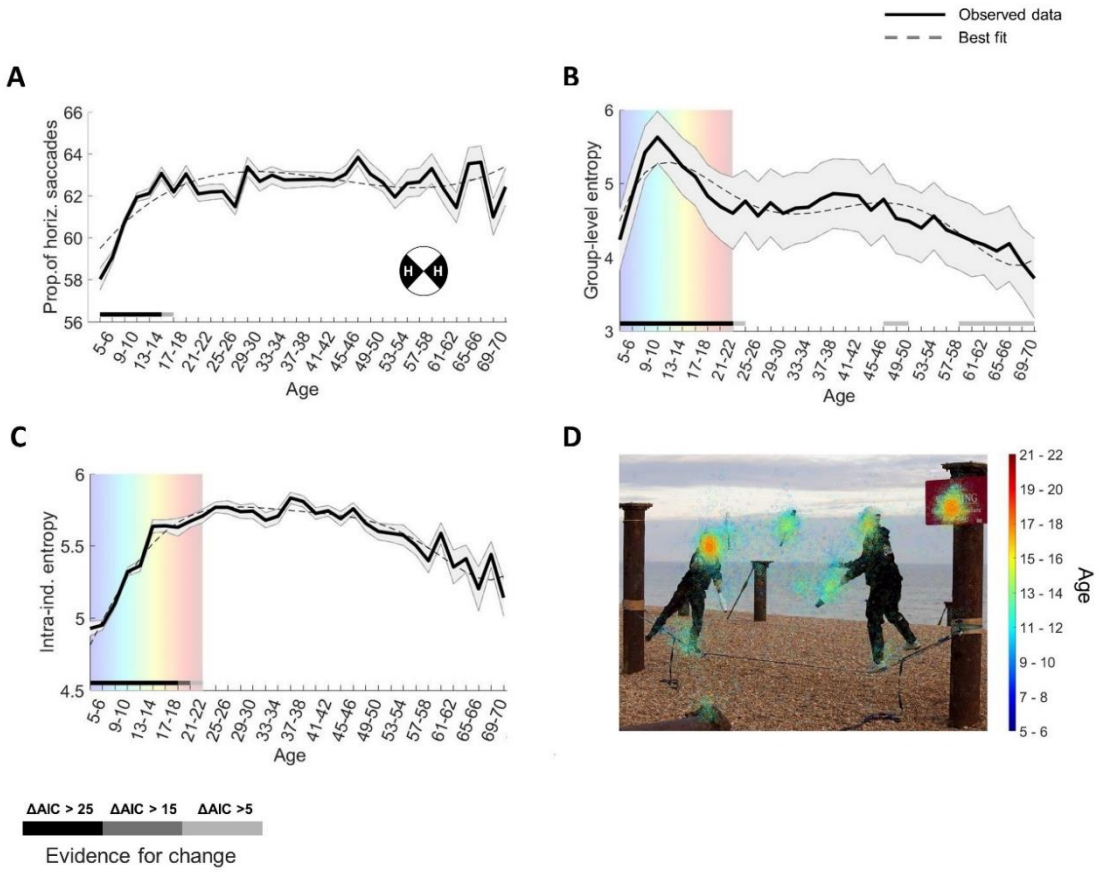

Figure 6. **Protracted development of the horizontal bias and inter-observer similarity**. (A) shows the average proportion of horizontal fixations across age bins. (B) shows the average entropy of group-level fixation maps across age bins. (C) shows the average entropy of observer-specific fixation maps across age bins. (D) shows fixation data from observers aged 5-22 overlaid as colored circles on one example image. Colors correspond to age groups as indicated by the colorbar (the same color coding is used in section B and C to indicate the age range). Note the lower dispersion of hotter colors, indicating less dispersion on the group level for adult participants. All error shades indicate +/- 1 standard error of the mean (SEM) and the dotted lines the best fitting polynomial. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. Linka et al., 2024).

Taken together, study 4 showed that gaze behaviour towards semantic object categories, as well as visual exploration and saccadic visuo-spatial biases develop across the first two decades of life. At the same time, while visual exploration increased on an individual level, gaze behaviour started to become more canonical across observers from the early teenage years on.

6. Discussion

Research on gaze behaviour towards natural scenes has largely focused on the influence of image features on an observer-general level. The aim of this thesis is to shift the focus towards studying potential gaze tendencies of the observer. The findings presented here suggest that free viewing behaviour towards high-level image features depends on distinct and systematic characteristics of the observer. Here we confirmed the relevance of two such

characteristics: face processing ability and age. Below, I will examine these insights in more detail and discuss their implications for potential underlying cognitive and neural mechanisms as well as for the development of the visual system.

6.1. A continuous spectrum of semantic salience

Differences in gaze behaviour have been linked to various clinical conditions and abilities. For example, people with ASD (Wang et al., 2015) or DP (Bobak et al., 2017), tend to focus less on social stimuli such as faces and eyes. In study 1, we show that SRs – individuals with exceptional face identity processing abilities – exhibit a preference for looking at faces and a reduced tendency to look at text or objects that have been touched. Study 3 indicates that preschool children pay more attention to objects that have been touched and to faces, but significantly less to text. Notably, across nearly all semantic dimensions, there is some degree of overlap between the specific groups being tested and the control groups. In adults, we replicated stable individual differences across these semantic dimensions (study 2; see also de Haas et al., 2019). These findings suggest an interesting characteristic regarding the distribution of gaze tendencies. It seems that visual preferences for objects across several semantic properties are distributed across a wide spectrum within the population. The maximum-to-minimum ratios of these preferences range from 1.37 for Motion to 2.47 for Text for a sample of 100 participants (see study 2). This spectrum, particularly in the context of faces, is characterized by specific groups of individuals who cluster at opposite ends. Importantly, most of these semantic gaze tendencies do not exist independently, but rather in a push-pull relationship (cf. de Haas et al., 2019). That is, if I spend a lot of time focusing on faces, it is likely that I will spend less time looking at text.

6.2. Functional implications of fixation biases towards natural scenes

Recently, de Haas et al. (2019) have reported a correlation between the tendency to first fixate on faces and face recognition performance. Others have shown reduced gaze towards faces in individuals with DP (Bobak et al., 2017) and decreased face recognition abilities in those with ASD (Dawson et al., 2005). Along with the visual preference for faces in SRs demonstrated in study 1, these results suggest a link between the ability to recognize faces and enhanced gaze towards them. It is important to note that we did not explicitly test face recognition performance in our control group. Consequently, we had limited statistical power to examine the link between visual salience and face cognition skills, as we could only test a categorical rather than a graded relationship. Nevertheless, it is tempting to speculate about a potential causal interplay between them. Do face recognition performance and face preferences operate and evolve independently, or does one precede and reinforce the other? For instance, early preferences for faces and eyes may lead to corresponding advantages in face recognition performance. At the same time, early advantages in face identity processing might result in enhanced visual attention to faces. Interestingly, gaze behaviour towards faces and face recognition abilities are *both* highly heritable (Constantino et al., 2017; Wilmer et al., 2011). However, whereas heritability estimates for gaze behaviour towards social scenes are substantially higher for infants (Constantino et al., 2017) compared to older children (9–14 years; Kennedy et al., 2017), those for face identity processing performance seem to increase with age (Zhu et al., 2010). Future genetic research could investigate the extent to which these genetic factors overlap and conduct longitudinal studies to examine how they interact with each other and with experience over time.

In study 1, we further showed that SRs fixate closer to a point just below the eyes, which has been identified as optimal for face identification following a Foveated Bayesian Ideal Observer Model (Peterson & Eckstein, 2012). This model relies on assumptions such as perfect symmetric foveation. SRs may adhere more closely to these assumptions than controls, exhibiting more symmetric visual field geometries. This symmetry could result in fewer compensatory eye movements when integrating facial information. Future research could investigate this further by using psychophysical tests to measure acuity and crowding across the visual field in SRs compared to controls.

As noted earlier, evidence suggests that the influence of genetic factors on gaze behaviour towards social content appears to decrease with age (Constantino et al., 2017; Kennedy et al., 2017). One possible explanation is that as individuals age, genetic predispositions may increasingly be overtaken by learned gaze pattern. Prior research has shown that basic oculomotor behaviour (Helo et al., 2014) and the influence of low-level image features (Acik et al., 2010) undergo changes over time. May this also be true for high-level features? To test this, in study 2 we first validated a short test that could capture most of the tested dimensions with acceptable re-test consistency. In study 3, we then applied this test and found that gaze behaviour towards objects across different semantic dimensions varies among age groups. Specifically, preschoolers look significantly less towards text compared to adults, suggesting that literacy mediates attention towards text. Furthermore, children demonstrate a visual preference for objects that have been touched and, albeit to a lesser extent, faces. In study 4, we then demonstrated that these changes in semantic biases develop over decades. These findings lead to two main conclusions: First, a link between literacy and text salience may indicate a broader functional relevance of gaze biases that extend beyond those towards faces as discussed earlier; and second, the protracted development of these biases suggest that gaze towards semantic information is remarkably plastic. What factors may drive these developments? Given that the emergence of text salience occurs throughout much of an individual's educational trajectory, it is possible that literacy and reading experience are key factors in driving text salience. This process could be self-reinforcing: As literacy improves, text becomes a more frequent component of our visual diet, enhancing our salience for text and in turn further strengthening literacy skills. Neither in study 3 nor in study 4 did we explicitly investigate the mediating effect of literacy. Future research may therefore examine the development of gaze biases towards text longitudinally while also assessing literacy development.

Furthermore, studies 3 and 4 indicate that gaze towards touched objects is more pronounced in preschoolers and takes about 20 years to reach an adult-like state. One explanation for this may lie in the development of understanding (inter-)actions among people and objects. Representations of typically occurring actions (e.g. a person throwing a darts arrow) may develop with experiences. With increasing experience, adults may no longer need to fixate directly on such regions, but extract semantic information from the periphery (Munneke et al., 2013). Children on the other hand may have experienced a given (inter-)action (or its variations) less frequently and thus require more extensive foveation of touched objects and hands in order to obtain sufficient information for action understanding. This may be enhanced by the strong curiosity in children to learn about the functions and causal properties of new objects (Chouinard, 2007; Gopnik, 1998). A hypothesis that could be drawn from this is that children might exhibit more (re-)fixations between objects, agents and their interactions in order to better integrate their spatial and semantic relationships. Interestingly, preliminary findings from our lab suggest that refixations indeed show a protracted

decrease until the third decade of life. Future work may probe whether this development is a result of (re-)fixations aimed at (inter-)actions or extends to other objects, potentially due to a protracted enhancement of visual working memory (Luna et al., 2004).

Studies 3 and 4 additionally indicate that children spend more time focusing on faces, and that this preference for faces gradually decreases until the early twenties. Surprisingly, we found an opposite developmental trend for face preference for first fixations towards a scene. Children showed a *lower* proportion of first fixations (study 3), and it takes approximately 20 years to reach a level similar to that of adults (study 4). At the same time, evidence suggests that face identity processing matures over a similarly extended period (Germine et al., 2011). Together, we speculate that these findings may point to two distinct mechanisms influencing the developmental paths of face preference: one that is primarily bottom-up and captures attention immediately after the stimulus is presented. This mechanism may show a stronger, perhaps more finely tuned, bias towards faces in adults, possibly reflecting their enhanced speed to saccade towards faces (Borovska & de Haas, 2023; Crouzet et al., 2010), as well as the protracted development of face recognition capabilities (Germine et al., 2011). A second mechanism, inferred from dwell time distributions, might be influenced more by top-down processes and reflect adults' increased competition between text and other semantic categories (de Haas et al., 2019). This mechanism could also perform a compensatory function; younger individuals, with less mature face identity processing abilities, might dwell longer on faces to accumulate enough visual information for face processing. Future studies could test these mechanisms by assessing the speed and accuracy of early face-directed saccades in children and adults. Additionally, longitudinal research might investigate the link between the development of (initial) face fixation tendencies and face recognition skills.

Beyond developmental differences in gaze towards object semantics, study 4 revealed a protracted increase of the horizontal bias and a parallel decrease of the center bias. Interestingly, the developmental trajectory of the horizontal bias remained when excluding images with text elements. This indicates that the development of the horizontal bias is not merely an artifact of text fixations. An additional control analysis, which tested whether the correlation between text salience and horizontal bias persists when controlling for age, suggests that the horizontal bias and text salience develop independently of each other. The protracted changes in horizontal bias are consistent with recent studies, which demonstrate that children and adolescents have a reduced sensitivity advantage for the horizontal meridian when compared to adults (Carrasco et al., 2022, 2023). Future longitudinal research could clarify the link between the development of this horizontal sensitivity advantage and the saccadic horizontal bias reported here. Moreover, we could show that visual exploration develops in a protracted manner: Throughout the first twenty years, visual exploration increases in a linear fashion before it slowly decreases until age 60. Surprisingly, at the same time we found that from childhood until young adulthood gaze behaviour becomes more canonical. Thus, while individuals appear to focus on more aspects of their visual environment as they grow older, they become more similar in doing so. What might be a reason for this? With ongoing age, we may develop a shared understanding of objects and their informative features. In infancy, there may be an initial visual preference for a limited number of objects (or categories). These preferences may be highly individual as they are influenced by a person's learning history, genetic makeup, or personal environment. As we move to out-of-home care, our visual interests may shift dramatically, yet in a collective manner, potentially resulting in less idiosyncratic gaze patterns.

6.3. A potential neural basis for gaze biases towards object semantics

Up to this point, the discussion has mostly focused on exploring potential cognitive mechanisms that might underlie reported semantic gaze biases. What may be a potential neural basis?

It is well known that regions in the Ventral Temporal Vortex (VTC) are selective for object categories such as faces (Kanwisher & Yovel, 2006), bodies (Downing et al., 2001), hands (Bracci et al., 2010), and words (McCandliss et al., 2003). These category selective patches are located close to each other and seem to compete for cortical space (Arcaro et al., 2017; Dehaene et al., 2010; Grill-Spector & Weiner, 2014; McCandliss et al., 2003). Interestingly, Arcaro et al. (2017) demonstrated that face-deprived macaques show a decrease in gaze behaviour towards faces and an increase in gaze towards hands, which is accompanied by the development of cortical regions for hands, at the expense of those for faces. This indicates a link between gaze biases towards higher-level object features, corresponding visual experiences and the competitive dynamics of high-level cortical areas. Notably, in our first study, we observed that SRs show a visual preference for faces, but spent less time looking at touched objects and text. Additionally, gaze biases towards faces are negatively correlated with those towards text and touched objects/hands, as reported by de Haas et al. (2019). Future work may investigate whether semantic fixation biases in SRs and neurotypical individuals are rooted in the functional organization of the ventral stream, reflecting a similar competition for cortical territory. Given that variations in face selectivity within the Fusiform Face Area have been linked to individual differences in face recognition performance (Furl et al., 2011), future research may test a potential mediating role of face recognition ability.

The study by Arcaro et al., (2017) points to an interesting interplay between cortical and perceptual plasticity and visual experiences in primates. For humans, a drastic change in their visual diet occurs with the onset of reading (see above). A recent longitudinal study by Nordt et al., (2021) demonstrates that from the beginning of reading until the ages of 15-17, areas of the VTC that previously preferred limbs start to show a preference for text (and faces). This suggests a prolonged process of cortical recycling, which may be related to changes in visual diet (Kubota et al., 2023; Nordt et al., 2021). Interestingly, these findings converge with the protracted increase in fixation bias towards text elements and the parallel decrease in gaze to touched objects/hands observed in studies 3 and 4. This suggests a link between the development of (competing) semantic gaze biases, changes in visual diet, and cortical recycling. Future research may test this relationship in a longitudinal design assessing whether and how these factors interact over time following the start of school. This could involve employing experience sampling methods such as mobile eye tracking to monitor changes in visual diet and semantic salience, combined with neuroimaging and literacy assessments. Recent longitudinal data demonstrate that the cortical development for face and text selectivity go hand in hand with corresponding recognition performance (Nordt et al., 2023). Therefore, it would be interesting to also examine whether object recognition abilities evolve in relation to changes in visual diet and semantic gaze tendencies.

6.4. Overall limitations

While specific study limitations are addressed throughout this discussion, some broader methodological shortcomings will be outlined here. It is important to mention that all findings and their interpretations presented here

derived from analysing eye movements towards static natural scenes (refer to Figure 1 for example stimuli). Although natural scenes tend to be more ecologically valid than highly controlled stimuli, they (1) typically constitute a relatively small portion of our visual field, (2) evoke only a limited sense of immersion and (3) lack motion and auditory elements, both of which can influence gaze behaviour (Coutrot & Guyader, 2014; Dorr et al., 2010; Mital et al., 2011). Furthermore, we presented images for either 2 or 3 seconds each. Yet, fixation dynamics change as a function of viewing time (Pannasch et al., 2011; Linka & de Haas, 2023), perhaps due to a change in experimental pace (Linka & de Haas, 2023). The reported variations in gaze behaviour may thus alter with changing viewing duration. Additionally, all present findings are based on free viewing data. As mentioned earlier, previous research has demonstrated that (top-down) goal-oriented tasks affect eye movements (e.g. Castelhano, 2008; Hayhoe & Ballard, 2005). Exploring the effects of age and enhanced face recognition abilities on goal-driven tasks like visual search is an interesting avenue for future research. Finally, in studies 3 and 4, we examined developmental effects through a cross-sectional approach which risks blending purely age-related differences with generational effects. Therefore, the longitudinal studies mentioned earlier are essential for accurately monitoring and understanding individual developmental trajectories over time.

6.5. Conclusion

This thesis begins with us imagining a friend who saw a stop sign directly ahead of us, which we failed to notice. The outcomes of the four studies presented here provide evidence that gaze behaviour towards natural scenes may be driven not only by bottom-up, top-down influences and generic visuo-spatial biases, but also by traits of the observer. Specifically, we found that gaze behaviour towards high-level object attributes is highly individual across the population and atypical within certain cohorts, covering a spectrum of semantic gaze biases. Thus, one reason for my overlooking of a stop sign might trace back to a particular characteristic of mine – perhaps I am at the lower end of a text salience spectrum. But why? We speculate that face recognition ability, literacy, and action understanding may be related to fixation tendencies towards higher-level object attributes, which unfold in a protracted manner throughout the lifespan. These developmental shifts might coincide with changes in our visual diet, ultimately resulting in a more canonical gaze pattern. Therefore, age differences between me and my friend, as well as an enhanced face recognition ability may lead me to focus more on faces at the expense of text. Recent neurodevelopmental studies further suggest an extended development of high-level cortical areas, accompanied by cortical recycling that aligns with our behavioural findings. A pressing open question, thus, is how these neural developments connect to changes in visual diet and semantic salience reported here, and how these are mediated by functional factors like face recognition and literacy.

7. References

Acik, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental Changes in Natural Viewing Behavior: Bottom-Up and Top-Down Differences between Children, Young Adults and Older Adults. *Frontiers in Psychology*, *1*, 207. https://doi.org/10.3389/fpsyg.2010.00207

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, *15*(5), 9. https://doi.org/10.1167/15.5.9

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, *20*(10), 1404–1412. https://doi.org/10.1038/nn.4635

Bahill, A. T., & Troost, B. T. (1979). Types of saccadic eye movements. *Neurology*, *29*(8), 1150–1152.

Bestelmeyer, P. E. G., Tatler, B. W., Phillips, L. H., Fraser, G., Benson, P. J., & St.Clair, D. (2006). Global visual scanning abnormalities in schizophrenia and bipolar disorder. *Schizophrenia Research*, *87*(1), 212–222. https://doi.org/https://doi.org/10.1016/j.schres.2006.06.015

Betz, T., Kietzmann, T. C., Wilming, N., & Koenig, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *Journal of Vision*, *10*(3), 15.

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*(23), 2577–2587.

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-Movement Strategies in Developmental Prosopagnosia and "Super" Face Recognition. *Quarterly Journal of Experimental Psychology*, *70*(2), 201–217. https://doi.org/10.1080/17470218.2016.1161059

Borovska, P., & de Haas, B. (2023). Faces in scenes attract rapid saccades. *Journal of Vision*, *23*(8), 11. https://doi.org/10.1167/jov.23.8.11

Bracci, S., Ietswaart, M., Peelen, M. V, & Cavina-Pratesi, C. (2010). Dissociable Neural Responses to Hands and Non-Hand Body Parts in Human Left Extrastriate Visual Cortex. *Journal of Neurophysiology*, *103*(6), 3389–3397. https://doi.org/10.1152/jn.00215.2010

Broda, M. D., & de Haas, B. (2022). Individual differences in looking at persons in scenes. *Journal of Vision*, *22*(12), 9. https://doi.org/10.1167/jov.22.12.9

Buswell, G. T. (1935). *How people look at pictures: a study of the psychology and perception in art.*

Carrasco, M., Myers, C., & Roberts, M. (2023). Visual field asymmetries vary between adolescents and adults. *BioRxiv*, 2023.03.04.531124. https://doi.org/10.1101/2023.03.04.531124

Carrasco, M., Roberts, M., Myers, C., & Shukla, L. (2022). Visual field asymmetries vary between children and adults. *Current Biology*, *32*(11), R509–R510. https://doi.org/10.1016/j.cub.2022.04.052

Carrasco, M., Talgar, C., & Cameron, E. (2001). Characterizing visual performance fields: Effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spatial Vision*, *15*, 61–75. https://doi.org/10.1163/15685680152692015

Castelhano, M. S. (2008). Eye movements during reading, visual search, and scene perception: An overview. In K. Rayner (Ed.), *Cognitive and cultural influences on eye movements.* (pp. 3–33). Tianjin People's Publishing House.

Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *62*(1), 1.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, *9*(12), 10.

Chouinard, M. (2007). Children's questions: A mechanism for cognitive development [Monograph]. *Monographs of the Society for Research in Child Development*, *72*, 1, vii–ix. https://doi.org/10.1111/j.1540-5834.2007.00412.x

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., Gillespie, S., Klaiman, C., Klin, A., & Jones, W. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*, *547*(7663), 340–344. https://doi.org/10.1038/nature22999

Coren, S., & Girgus, J. (2020). *Seeing is Deceiving: The Psychology of Visual Illusions*. https://doi.org/10.4324/9781003050681

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 5. https://doi.org/10.1167/14.8.5

Crouzet, S., Kirchner, H., & Thorpe, S. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*, 16.1-17. https://doi.org/10.1167/10.4.16

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, *292*(4), 497–523. https://doi.org/https://doi.org/10.1002/cne.902920402

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the Nature of Face Processing Impairment in Autism: Insights From Behavioral and Electrophysiological Studies. *Developmental Neuropsychology*, *27*, 403–424. https://doi.org/10.1207/s15326942dn2703_6

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317–329.

De Haas, B., Iakovidis, A., Schwarzkopf, D., & Gegenfurtner, K. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences*, 116, 11687. https://doi.org/10.1101/444257

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, *16*(3), e2004103-. https://doi.org/10.1371/journal.pbio.2004103

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, *330*(6009), 1359–1364. https://doi.org/10.1126/science.1194140

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28. https://doi.org/10.1167/10.10.28

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of Vision*, *3*(10), 1. https://doi.org/10.1167/3.10.1

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5), 14.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 2.

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research*, *50*(8), 779–795.

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, *48*(17), 1777–1790. https://doi.org/https://doi.org/10.1016/j.visres.2008.05.018

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6.

Frank, M., Vul, E., & Saxe, R. (2011). Measuring the Development of Social Attention Using Free-Viewing. *Infancy*, *17*. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform Gyrus Face Selectivity Relates to Individual Differences in Facial Recognition Ability. *Journal of Cognitive Neuroscience*, *23*(7), 1723–1740. https://doi.org/10.1162/jocn.2010.21545

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, *118*(2), 201–210. https://doi.org/https://doi.org/10.1016/j.cognition.2010.11.002

Gopnik, A. (1998). Explanation as Orgasm*. *Minds and Machines*, *8*(1), 101–118. https://doi.org/10.1023/A:1008290415597

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Harel, J., Koch, C., & Perona, P. (2006). Graph-Based Visual Saliency. In *Adv. Neural Inform. Process. Syst* (Vol. 19).

Harvey, B. M., & Dumoulin, S. O. (2011). The Relationship between Cortical Magnification Factor and Population Receptive Field Size in Human Visual Cortex: Constancies in Cortical Architecture. *The Journal of Neuroscience*, *31*(38), 13604. https://doi.org/10.1523/JNEUROSCI.2572-11.2011

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Hayhoe, M., Mennie, N., Sullivan, B., & Gorgos, K. (2005). The role of internal models and prediction in catching balls. *Proceedings of the American Association for Artificial Intelligence*, 1–5.

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, *103*, 83–91. https://doi.org/10.1016/j.visres.2014.08.006

Henderson, J. M. (2008). Eye movements and scene memory. *Visual Memory*, 87–121.

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*(10), 743–747. https://doi.org/10.1038/s41562-017-0208-0

Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1390.

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210.

Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*(2), 204–210. https://doi.org/10.3758/s13423-011-0205-7

Himmelberg, M. M., Winawer, J., & Carrasco, M. (2023). Polar angle asymmetries in visual perception and neural architecture. *Trends in Neurosciences*, *46*(6), 445–458. https://doi.org/https://doi.org/10.1016/j.tins.2023.03.006

Horton, J. C., & Hoyt, W. F. (1991). The Representation of the Visual Field in Human Striate Cortex: A Revision of the Classic Holmes Map. *Archives of Ophthalmology*, *109*(6), 816–824. https://doi.org/10.1001/archopht.1991.01080060080030

Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *The American Journal of Psychology*, *9*(4), 575–586.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1), 1–19. https://doi.org/https://doi.org/10.1016/0010-0277(91)90045-6

Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, *50*, 169–179. https://doi.org/https://doi.org/10.1016/j.neubiorev.2014.10.009

Jones, W., Carr, K., & Klin, A. (2008). Absence of Preferential Looking to the Eyes of Approaching Adults Predicts Level of Social Disability in 2-Year-Old Toddlers With Autism Spectrum Disorder. *Archives of General Psychiatry*, *65*, 946–954. https://doi.org/10.1001/archpsyc.65.8.946

Kanwisher, N., & Yovel, G. (2006). The Fusiform Face Area: A Cortical Region Specialized for the Perception of Faces. *Philosophical Transactions: Biological Sciences*, *361*(1476), 2109–2128. http://www.jstor.org/stable/20209804

Kennedy, D. P., D'Onofrio, B. M., Quinn, P. D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic Influence on Eye Movements to Complex Scenes at Short Timescales. *Current Biology*, *27*(22), 3554-3560.e3. https://doi.org/https://doi.org/10.1016/j.cub.2017.10.007

Klin, A., & Jones, W. (2008). Altered face scanning and impaired recognition of biological motion in a 15-month-old infant with autism: REPORT. *Developmental Science*, *11*, 40–46. https://doi.org/10.1111/j.1467-7687.2007.00608.x

Kowler, E. (2011). Eye movements: The past 25years. *Vision Research*, *51*(13), 1457–1483. https://doi.org/https://doi.org/10.1016/j.visres.2010.12.014

Krishna, O., Helo, A., Rämä, P., & Aizawa, K. (2018). Gaze distribution analysis and saliency prediction across age groups. *PLOS ONE*, *13*(2), e0193149-. https://doi.org/10.1371/journal.pone.0193149

Kubota, E., Grill-Spector, K., & Nordt, M. (2023). Rethinking cortical recycling in ventral temporal cortex. *Trends in Cognitive Sciences*. https://doi.org/https://doi.org/10.1016/j.tics.2023.09.006

Kümmerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding low-and high-level contributions to fixation prediction. *Proceedings of the IEEE International Conference on Computer Vision*, 4789–4798.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, *3*(12), 1340–1345.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Linka, M., & de Haas, B. (2023, March 30). Detection, Inspection, Return: A functional classification of fixations in complex scenes. https://doi.org/10.31234/osf.io/bqfdy

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of Cognitive Processes From Late Childhood to Adulthood. *Child Development*, *75*(5), 1357–1372. https://doi.org/https://doi.org/10.1111/j.1467-8624.2004.00745.x

Mackay, M., Cerf, M., & Koch, C. (2012). Evidence for two distinct mechanisms directing gaze in natural scenes. *Journal of Vision*, *12*(4), 9. https://doi.org/10.1167/12.4.9

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299. https://doi.org/https://doi.org/10.1016/S1364-6613(03)00134-7

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation*, *3*(1), 5–24. https://doi.org/10.1007/s12559-010-9074-z

Munneke, J., Brentari, V., & Peelen, M. (2013). The influence of scene context on object recognition is independent of attentional focus. *Frontiers in Psychology*, *4*, 552. https://doi.org/10.3389/fpsyg.2013.00552

Niebur, E., & Koch, C. (1995). Control of selective visual attention: Modeling the" where" pathway. *Advances in Neural Information Processing Systems*, *8*.

Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2021). Cortical recycling in high-level visual cortex during childhood development. *Nature Human Behaviour*, *5*(12), 1686–1697. https://doi.org/10.1038/s41562-021-01141-5

Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2023). Longitudinal development of category representations in ventral temporal cortex predicts word and face recognition. *Nature Communications*, *14*(1), 8010. https://doi.org/10.1038/s41467-023-43146-w

Oliva, A. (2005). CHAPTER 41 - Gist of the Scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 251–256). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012375731-9/50045-8

Orschansky, J. (1899). Eine Methode die Augenbewegungen direct zu untersuchen. *Zentralblatt Für Physiologie*, *12*, 785.

Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics*, *73*(4), 1120–1132. https://doi.org/10.3758/s13414-011-0090-1

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, *109*(48), E3314–E3323. https://doi.org/10.1073/pnas.1214269109

Peterson, M. F., & Eckstein, M. P. (2013). Individual Differences in Eye Movements During Face Identification Reflect Observer-Specific Optimal Points of Fixation. *Psychological Science*, *24*(7), 1216–1225. https://doi.org/10.1177/0956797612471684

Peterson, M. F., Zaun, I., Hoke, H., Jiahui, G., Duchaine, B., & Kanwisher, N. (2019). Eye movements and retinotopic tuning in developmental prosopagnosia. *Journal of Vision*, *19*(9), 7. https://doi.org/10.1167/19.9.7

Peterson, M., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of Vision*, *16*, 12. https://doi.org/10.1167/16.7.12

Ramon, M. (2021). Super-Recognizers – a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia*, *158*, 107809. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2021.107809

Riby, D. M., Hancock, P. J. B., Jones, N., & Hanley, M. (2013). Spontaneous and cued gaze-following in autism and Williams syndrome. *Journal of Neurodevelopmental Disorders*, *5*(1), 13. https://doi.org/10.1186/1866-1955-5-13

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2019). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, *19*(3), 1. https://doi.org/10.1167/19.3.1

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5), 9.

Sliney, D. H. (2016). What is light? The visible spectrum and beyond. *Eye*, *30*(2), 222–229. https://doi.org/10.1038/eye.2015.252

Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural scenes: Objects dominate features. *Vision Research*, *107*, 36–48. https://doi.org/https://doi.org/10.1016/j.visres.2014.11.006

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

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2). https://doi.org/10.16910/jemr.2.2.5

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6–7), 1029–1054.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. https://doi.org/https://doi.org/10.1016/j.actpsy.2010.02.006

Tseng, P.-H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, *9*(7), 4.

Van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E. J., & Visser, I. (2016). Do infants have the horizontal bias? *Infant Behavior and Development*, *44*, 38–48. https://doi.org/https://doi.org/10.1016/j.infbeh.2016.05.005

Võ, M. L.-H., Boettcher, S. E. P., & Draschkow, D. (2019). Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205–210. https://doi.org/https://doi.org/10.1016/j.copsyc.2019.03.009

Wang, H.-C., & Pomplun, M. (2012). The attraction of visual attention to texts in real-world scenes. *Journal of Vision*, *12*. https://doi.org/10.1167/12.6.26

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., & Zhao, Q. (2015). Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron*, *88*(3), 604–616. https://doi.org/https://doi.org/10.1016/j.neuron.2015.09.042

Wilmer, J., Germine, L., Chabris, Chatterjee, G., Nakayama, Williams, M., Loken, & Duchaine, B. (2011). Wilmer, J.B., Germine, L., Chabris, C.F., Chatterjee, G., Nakayama, K., Williams, M.A., Loken, E. & Duchaine, B. (2010). Human face recognition ability is highly heritable. Proceedings of the National Academy of Sciences, 107, 5238-5241. doi:10.1073/pnas. *Proceedings of the National Academy of Sciences.*

Xu, J., Jiang, M., Wang, S., Kankanhalli, M., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision*, *14*. https://doi.org/10.1167/14.1.28

Yarbus, A. L. (2013). *Eye movements and vision*. Springer.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the Specific Cognitive Ability of Face Perception. *Current Biology*, *20*(2), 137–142. https://doi.org/https://doi.org/10.1016/j.cub.2009.11.067

II. Publications

Study 1

Linka, M., Broda, M. D., Alsheimer, T., de Haas, B., & Ramon, M. (2022). Characteristic fixation biases in Super-Recognizers. *Journal of Vision*, 22(8), 17.

Characteristic fixation biases in Super-Recognizers

Marcel Linka Experimental Psychology, Justus Liebig University Giessen, Giessen, Germany

Maximilian Davide Broda Experimental Psychology, Justus Liebig University Giessen, Giessen, Germany

Tamara Alsheimer Experimental Psychology, Justus Liebig University Giessen, Giessen, Germany; Applied Face Cognition Lab, University of Lausanne, Institute of Psychology, Lausanne, Switzerland

Benjamin de Haas* Experimental Psychology, Justus Liebig University Giessen, Giessen, Germany

Meike Ramon* Applied Face Cognition Lab, University of Lausanne, Institute of Psychology, Lausanne, Switzerland

Neurotypical observers show large and reliable individual differences in gaze behavior along several semantic object dimensions. Individual gaze behavior toward faces has been linked to face identity processing, including that of neurotypical observers. Here, we investigated potential gaze biases in Super-Recognizers (SRs), individuals with exceptional face identity processing skills. Ten SRs, identified with a novel conservative diagnostic framework, and 43 controls freely viewed 700 complex scenes depicting more than 5000 objects. First, we tested whether SRs and controls differ in fixation biases along four semantic dimensions: faces, text, objects being touched, and bodies. Second, we tested potential group differences in fixation biases toward eyes and mouths. Finally, we tested whether SRs fixate closer to the theoretical optimal fixation point for face identification. SRs showed a stronger gaze bias toward faces and away from text and touched objects, starting from the first fixation onward. Further, SRs spent a significantly smaller proportion of first fixations and dwell time toward faces on mouths but did not differ in dwell time or first fixations devoted to eyes. Face fixation of SRs also fell significantly closer to the theoretical optimal fixation point for identification, just below the eyes. Our findings suggest that reliable superiority for face identity processing is accompanied by early fixation biases toward faces and preferred saccadic landing positions close to the theoretical optimum for face identification. We discuss future directions to investigate the functional basis of individual fixation behavior and face identity processing ability.

Introduction

Gaze behavior is traditionally studied via investigations emphasizing commonalities among observers. More recent studies, on the other hand, have shifted toward understanding *individual differences*, and they have revealed substantial reliable idiosyncratic patterns across observers. For example, neurotypical observers' oculomotor behavior is characterized by stable interindividual differences in general oculomotor parameters (i.e., saccades, anti-saccades, smooth pursuit) (Bargary, Bosten, Goodbourn, Lawrance-Owen, Hogg, & Mollon, 2017), as well as visual saliency, or fixation biases for distinct semantic categories, such as faces or text (de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2019; Kanan, Bseiso, Ray, Hsiao, & Cottrell, 2015; Linka & de Haas, 2020).

For faces specifically, neurotypical observers also exhibit reliable idiosyncratic fixation biases (i.e., fixation preferences for *specific parts* of the face) (Arizpe, Walsh, Yovel, & Baker, 2017; Mehoudar, Arizpe, Baker, & Yovel, 2014; Stacchi, Liu-Shuang, Ramon, & Caldara, 2019). Interestingly, a Bayesian foveated ideal observer model predicts a location just below the eyes as optimal for sampling identity information from the face (Peterson & Eckstein, 2012), but such face-specific sampling biases nevertheless are not generally predictive of observers' behavioral proficiency. Although most observers fixate close to this optimal location (Peterson & Eckstein, 2012), deviations from this appear to

Citation: Linka, M., Broda, M. D., Alsheimer, T., de Haas, B., & Ramon, M. (2022). Characteristic fixation biases in Super-Recognizers. *Journal of Vision*, *22*(8):17, 1–15, https://doi.org/10.1167/jov.22.8.17.

https://doi.org/10.1167/jov.22.8.17 Received February 15, 2022; published July 28, 2022 ISSN 1534-7362

reflect *individually* optimal behavior (Peterson & Eckstein, 2013), either reflecting retinotopic tuning biases (de Haas, Schwarzkopf, Alvarez, Lawson, Henriksson, Kriegeskorte, & Rees, 2016; de Haas, Sereno, & Schwarzkopf, 2021; Poltoratski, Kay, Finzi, & Grill-Spector, 2021) or prioritized processing of the facial information that is most diagnostic for a given observer (Stacchi et al., 2019).

Investigations of gaze behavior in special populations, on the other hand, are often characterized by more heterogeneous findings. For example, individuals with autism have been reported to exhibit an abnormally low tendency to fixate faces (Amestoy, Guillaud, Bouvard, & Cazalets, 2015; Constantino et al., 2017), or specific facial features (Tanaka & Sung, 2016). However, developmental changes have also been reported (Fedor, Lynn, Foran, DiCicco-Bloom, Luna, & O'Hearn, 2018), and, more generally, "a large inconsistency in findings" (Tang, Falkmer, Horlin, Tan, Vaz, & Falkmer, 2015, in response to Weigelt, Koldewyn, & Kanwisher, 2012) may be observed. At the same time, investigating the traits of extreme groups can offer significant power advantages over random samples of comparable sizes (de Haas, 2018), and studies of rare neuropsychological cohorts continue to advance our understanding of brain functioning (e.g., Avidan & Behrmann, 2021; Geskin & Behrmann, 2018; Ramon, 2018; Rossion, 2014). A finding of particular interest for the present study is that deviations from the theoretically optimal fixation location on faces do not appear to reflect individually optimal behavior in observers suffering from developmental prosopagnosia (Peterson, Zaun, Hoke, Jiahui, Duchaine, & Kanwisher, 2019). Specifically, a tendency to fixate lower on the face may be detrimental to face perception in these participants.

The present study investigated visual saliency in a unique observer cohort: Super-Recognizers (SRs), individuals with superior skills for processing facial identity information (Ramon, 2021; Ramon et al., 2019; Russell, Bobak, & White, 2009). To date, only one study has compared the gaze behavior of SRs to that of controls and reported reduced visual attention toward bodies and an increased dwell time on inner facial features, especially the nose (Bobak, Parris, Gregory, Bennetts, & Bate, 2016). At the same time, this study reported no significant difference in the overall salience of faces, suggesting that SRs may differ in the *way* they look at faces rather than their tendency to do so overall. We surmised that at present visual saliency in SRs had not been sufficiently addressed for two methodological reasons at least.

First, in the study by Bobak et al. (2016), observers' gaze patterns were determined based on free viewing of only 20 images of social scenes. Recent findings suggest that a minimum of 40 to 100 images are required to establish reliable estimates of individual observers' fixation biases (Linka & de Haas, 2020). Therefore, the study by Bobak et al. (2016) may have lacked the sensitivity necessary to identify stable idiosyncratic fixation biases. Furthermore, previous results have shown a correlation between face recognition performance and the specific tendency of the *first* fixation (the immediate saccade after image onset) to land on faces (de Haas et al., 2019), a measure that was not investigated in the study by Bobak and is particularly dependent on an adequate number of trials (Linka & de Haas, 2020).

Second, the individuals reported by Bobak et al. (2016) were identified as SRs based on their performance of a well-established measure of face *recognition* (i.e., the ability to locate experimentally learned among novel identities). That is, establishing their SR status did not include formal and isolated assessment of potentially superior face *perception* (i.e., simultaneous matching of face images portraying the same vs. discrimination from other identities). This may have reduced the sensitivity to find differences compared with controls for the following reason: Superiority in face recognition could reflect (domain-specific or -general) mnemonic capacities that need not necessarily coincide with superior perceptual skills. Therefore, it remains to be determined whether SRs that present with superiority in face perception, or face perception *and* recognition, exhibit atypical biases in visual salience toward faces.

In this study, we investigated natural information sampling in a group of SRs recently identified using a novel and conservative diagnostic framework (Ramon, 2021). Contrary to previous studies that identified SRs based on a performance on a single test of face (re)cognition (e.g., Bobak et al., 2016; Phillips et al., 2018), the cases reported by Ramon (2021) excelled across a number of highly challenging tests of face identity processing (assessing perception and recognition). Recent findings suggest that SRs identified in this manner utilize the same bandwidth of spatial frequency information for identity matching as typical observers but do so more consistently (Nador, Zoia, Pachai, & Ramon, 2021). Moreover, they encode novel facial identities in a viewpoint-invariant manner that facilitates later (surreptitiously solicited) recognition in a more consistent manner (Nador, Vomland, Thielgen, & Ramon, 2022), and regardless of face memorability (Nador, Alsheimer, Gay, & Ramon, 2021).

The present study sought to answer three main questions. First, we aimed to determine whether SRs show enhanced visual preference for faces compared to other observers. To this end, we compared the gaze behavior of SRs and controls exhibited during free-viewing of 700 complex scenes (with detailed metadata for visual object categories) (Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014) using a paradigm that provides highly reliable estimates of individual saliency

biases for various semantic dimensions including faces (de Haas et al., 2019). Second, we specifically investigated potential fixation differences for first fixations after image onset (i.e., landing position of the first saccade), which are thought to be dominated by bottom–up processing and more difficult to suppress than immediate saccades directed toward other stimulus categories (Crouzet, Kirchner, & Thorpe, 2010; de Haas et al., 2019). Third, and finally, using enhanced annotations for facial features in the stimuli (Broda & de Haas, 2022), we investigated observers' face fixations in greater detail to compare the *way* in which SRs and controls fixate faces. Specifically, we compared groups regarding their distance between (first) face fixations and the face region just below the eye, which is optimal for face identity processing according to a foveated Bayesian ideal observer model (cf. Peterson et al., 2019).

To preview our findings, SRs showed a significantly stronger tendency to fixate faces and significantly reduced fixation tendencies for text and objects being touched. These biases were already present for first fixations. Furthermore, SRs showed a tendency to fixate faces differently than controls, as they devoted a significantly smaller proportion of their face fixations toward mouths and fixated closer to the theoretically optimal landing point just below the eyes.

Methods

Observers

In total, 53 healthy observers with normal or corrected-to-normal vision took part in this study. All observers provided written informed consent, and all research procedures were approved by the respective local ethics committees (approval nos. 473 and 486, University of Fribourg; approval no. 2018-0051, University of Giessen) and adhered to the tenets of the Declaration of Helsinki.

SRs ($n = 10$; $M_{age} = 36.10$ years; $SD = 7.28$; one left-handed; five females) (Table 1) were recruited from a larger, recently assembled cohort of SR individuals (Ramon, 2021) and received no reimbursement for participation. SR status was ascertained according to the criteria and cut-offs defined recently by Ramon (2021). In brief, the superior face identity processing abilities of SRs were determined via achievement of high scores in at least two of three demanding tests probing processing of facial identity: the 40-item long form of the Yearbook Test (YBT) (Bruck, Cavanagh, & Ceci, 1991; Fysh, Stacchi, & Ramon, 2020; Stacchi et al., 2020) and the Facial Image Card Sorting Test (FICST) (Fysh et al., 2020; Jenkins, White, Van Montfort & Burton, 2011; Stacchi et al., 2020) as measures of face perception (i.e., matching), and the 102-item version of the Cambridge Face Memory Test (CFMT+) (Russell, Duchaine, & Nakayama, 2009) as a measure of face recognition (i.e., learning and memory).

The control sample ($n = 43$; $M_{age} = 23.37$ years; $SD = 4.19$; three left-handed; 33 females) was comprised of observers who were recruited at Justus-Liebig University Giessen (Germany) and were compensated with course credit or received payment (7€/hr) for participation. Control observers were recruited as pairs of friends and completed several other questionnaires and tasks that were unrelated to the present study (for which the pairing was also irrelevant).

	Demographics			Face perception			Face fixation bias	
Code	Gender	Age (y)	Handedness	FICST score (piles + errors – 2) (Max: 0)	YBT long raw score (Max: 35)	Face recognition CFMT+ raw score (Max: 102)	Dwell time (%)	First fixations (%)
UC1	M	42	R	0	20	92	30	37
FW1	M	32	R	1	22	97	30	44
NC1	F	41	R	7	17	92	42	35
PT1	F	33	R	0	17	78	50	52
MB2	F	45	L	0	18	96	32	42
AM1	F	31	R	0	17	99	36	46
VZ1	M	24	R	6	21	90	32	36
MB1	M	34	R	0	20	97	38	43
GP1	M	47	R	7	15	99	32	40
CB1	F	32	R	2	20	87	31	42

Table 1. Demographics and performance of recruited Super-Recognizers. The table provides information on demographics (gender, age, and handedness) and scores for three tests of face cognition, as well as fixation biases toward faces. SRs were recruited based on the diagnostic framework proposed by Ramon (2021).

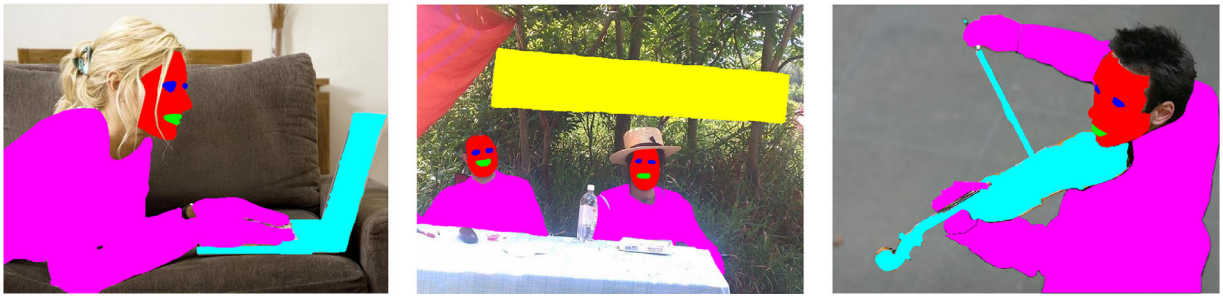

Figure 1. Example stimuli with pixel masks. Example stimuli with overlaid pixel masks for objects of the semantic categories: *Faces* (red), *Eyes* (blue), *Mouths* (green), *Bodies* (violet), *Touched* (cyan), and *Text* (yellow). All images were presented without pixel masks.

Apparatus

The free-viewing task was created and implemented with Psychtoolbox 3.0.12 (Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard, 2007; Pelli, 1997) in MATLAB R2019a (MathWorks, Natick, MA). Controls saw the stimuli presented on an ultra-high-definition monitor (3840 × 2160 resolution; LG Corporation, Seoul, South Korea) and SRs on a SyncMaster 2233RZ 3D liquid-crystal display monitor (1680 × 1050 resolution; Samsung Electronics, Suwon-si, South Korea). Eye tracking data from the left eye were measured with a tower-mounted EyeLink 1000 (SR Research, Kanata, Ontario, Canada) at a frequency of 1 kHz.

Stimuli and procedure

The free-viewing task included 700 images of everyday scenes, each showing multiple potentially dominant objects (Xu et al., 2014). Most of the scenes contained objects of multiple semantic dimensions. The images were displayed at a resolution of 1200 × 900 pixels and 31.5 × 23.6 degrees of visual angle (DVA), with a viewing distance of 62 cm (SR sample), and at a resolution of 2400 × 1800 pixels and 34.3 × 25.7 DVA, with a viewing distance of 55 cm (control sample). Semantic metadata for the full stimulus set included binary pixel masks for 5551 objects and corresponding labels for 12 semantic dimensions (OSIE dataset) (Xu et al., 2014), as well as the additional 6365 masks and labels for 10 semantic labels that came with the $OSIE_{Person}$ dataset (Broda & de Haas, 2022) (see Figure 1 for example pixel masks). To reduce overlap between labels, the *Smell* label was removed from all objects with the label *Text*, the labels *Operable* and *Gazed* were removed from all objects with the *Touched* label, and the *Watchable* label was removed from all objects with the label *Text* (cf. de Haas et al., 2019; Linka & de Haas, 2020).

After completing a nine-point calibration and validation procedure, participants were instructed to freely view seven blocks of 100 images each. Each image was presented for 2 seconds before a self-paced fixation cross appeared. The appearance of the next image could then be initiated by the participant by pressing the spacebar. The order of image presentation was identical across groups and participants.

Analysis

Data processing

Statistical analyses were performed with MATLAB R2019a. To ensure that only object-specific fixations were included, all onset fixations and fixations that could not be assigned to an object label were disregarded. Following de Haas et al. (2019) and Linka & de Haas (2020), fixations that fell within ~0.5 DVA from a labeled object were assigned the respective attribute label. When a fixation fell within ~0.5 DVA between two or more labeled objects, we assigned all respective object labels. Fixations were defined using the standard settings of the EyeLink parser, and fixations with a duration below 100 ms were removed (following the manufacturer's recommendations). Data and code allowing the reproduction of all presented results and figures can be downloaded from the Open Science Framework (https://osf.io/bdafk/).

Bootstrapping

To test group differences between SRs and the control group, we applied a nonparametric bootstrapping approach. This method maximizes statistical power and controls the risk of type 1 errors for unequal group sizes. In each iteration, we pooled individual fixation data across all 53 observers and drew a random subsample of $n = 10$ from the pooled sample. After this, we calculated the statistic of interest from the given fixation data. This procedure was repeated 10,000 times, resulting in a sampling distribution consisting of 10,000 bootstrap estimates. Under the null hypothesis (no systematic difference between SRs and controls), the SR sample

characteristic is not expected to be extreme relative to this bootstrapped null distribution. Therefore, the *p* value for the observed statistic of interest for the SR group was calculated as the proportion of random draws from the pooled groups that lay below or above that value.

Visual saliency across four semantic dimensions

We investigated group differences in visual saliency along four semantic dimensions: *Faces* to test the hypothesis of altered face saliency in SRs; *Text* and *Touched*, because they have previously shown a strong anti-correlation with individual face saliency (de Haas et al., 2019); and *Bodies*, for which Bobak et al. (2016) reported reduced fixation behavior in SRs. Note that most scenes included objects of multiple semantic categories; hence, fixations toward specific dimensions can lead to fewer fixations toward objects of competing semantic categories. Nevertheless, some dimensions of semantic salience have been found to correlate positively with each other for these images (de Haas et al., 2019). We first computed individual cumulative fixation times across fixations toward all labeled objects. The proportion of time spent on a given semantic attribute (cumulative dwell time ratio) was then determined for each subject. Additionally, we computed the proportion of *first fixations* toward objects of a given semantic dimension after image onset for each individual (first fixation ratio). To test group differences, we then pooled participants across groups and bootstrapped a null distribution of mean fixation ratios, against which we compared the observed value for SRs (see above). This was repeated for each dimension, and the corresponding *p* values were adjusted for the four semantic attributes using the Holm–Bonferroni correction (Holm, 1979).

Differences in fixation behavior toward inner face regions

Tendency to fixate eyes and mouths

To test whether SRs and control participants differed in their tendency to fixate Eyes and Mouths (Bobak et al., 2016), we first excluded all faces with an eye-to-mouth distance below 2.5 DVA. This was done to avoid inconclusive assignments of fixations to both eyes and mouths (due to the margin of tolerance), as well as false positives due to calibration inaccuracy. After that we calculated the cumulative fixation time on faces for each participant to then determine the individual proportion of this face fixation time spent on Eyes and Mouths. Further, we computed the individual proportions of fixations toward Eyes and Mouths among the first fixations landing on a given face (note that here we analyzed first fixations toward each observed face in a scene). Again, group differences were tested for statistical significance by bootstrapping the group mean for random samples drawn from all participants and comparing the observed mean for SRs against this null distribution (see above). Again, *p* values for Eye and Mouth fixations were adjusted using the Holm–Bonferroni correction (Holm, 1979).

Deviation from optimal fixation location for face identity processing

To investigate whether SRs tend to fixate closer to the ideal fixation location for identification just below the eyes, as predicted by a foveated Baysian ideal observer model (cf. Peterson & Eckstein, 2012), we first again removed all fixations with an eye to mouth distance of below 2.5 DVA. We then calculated the approximate optimal fixation point for each fixated face by calculating the point that lay at a height of 70% along the direct line connecting the centroids of mouth and eyes (Peterson & Eckstein, 2012, figure 5B). Finally, we projected each face fixation of each observer onto this connecting line and calculated its distance from the optimal location along this line normalized as percentages from the distance from the respective mouth centroid to the ideal fixation point. We then compared the resulting mean distances for each observer across groups, again bootstrapping the group mean for random samples drawn from all participants, and we compared the observed mean for SRs against this null distribution. We applied the Holm–Bonferroni correction (Holm, 1979) to adjust *p* values for first fixations and all fixations.

Results

Salience differences across object categories

First, we tested whether SRs would show higher cumulative dwell time and first fixation ratios for *Faces* than controls and reduced dwell time and first fixation ratios toward *Text*, *Touched* objects, and *Bodies*. Figure 2 shows individual gaze tendencies for each SR (red lines), superimposed on raincloud plots (Allen, Poggiali, Whitaker, Marshall, van Langen, & Kievit, 2019) of the distribution for controls. Figure 3 shows the bootstrapped null distributions of random sample means drawn from all participants (pooled across groups), with the observed mean of the SRs superimposed as a red line.

Bootstrap analyses confirmed that SRs spent more time fixating *Faces* than did the controls ($p_{FWE} < 0.01$). Further, SRs showed lower cumulative dwell-time ratios for the dimension *Text* ($p_{FWE} < 0.05$) and *Touched*

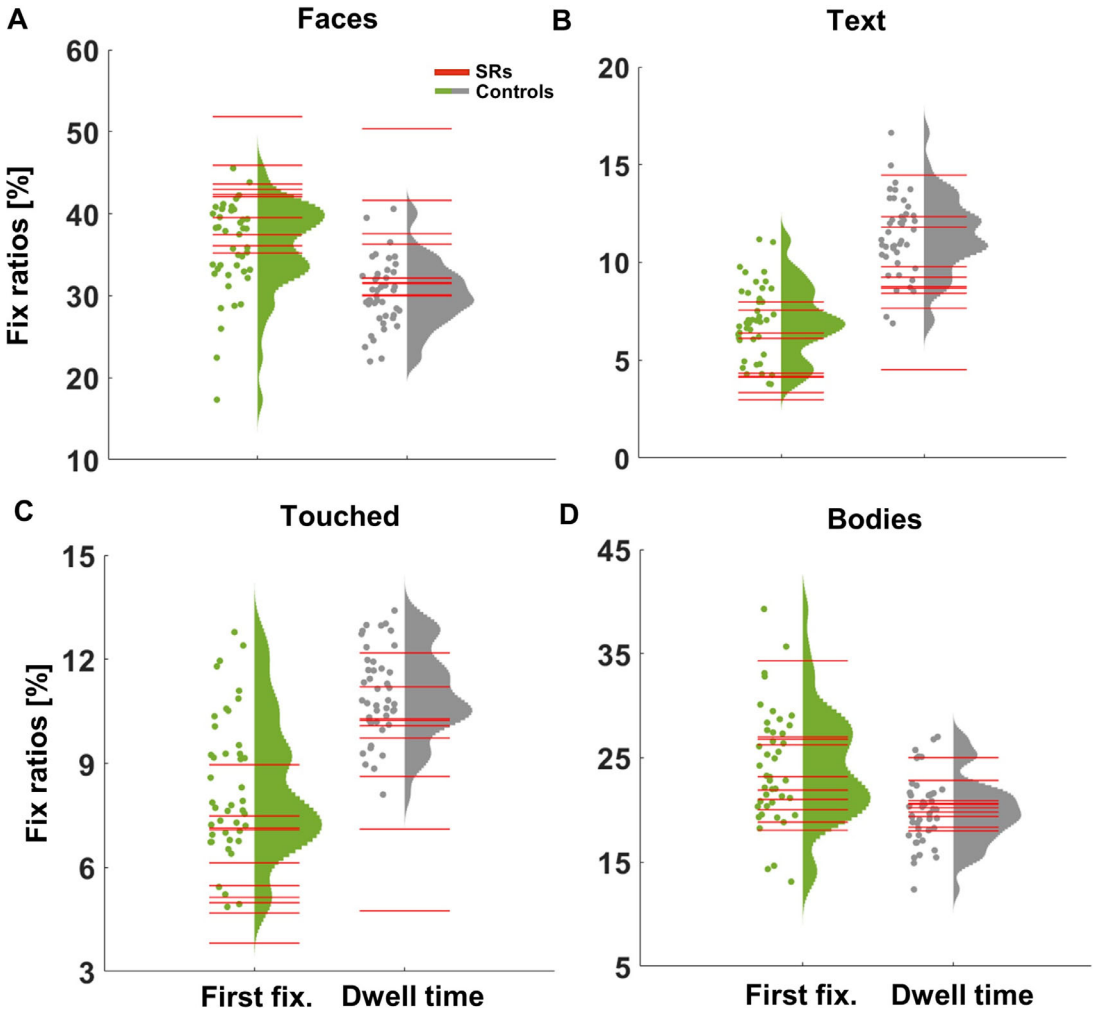

Figure 2. Individual gaze tendencies toward four semantic dimensions. (A–D) Right-hand leaves show the distributions of the control group for percent first fixations (green) and percent cumulative dwell time (gray) along each semantic dimension. Dots depicted in the left-hand raincloud plots indicate the corresponding individual data for each control subject. Superimposed red lines refer to the fixation ratios for each SR.

($p_{FWE} < 0.01$) compared with the control group. Further, SRs and controls did not differ in dwell time proportion toward *Bodies* ($p_{FWE} = 0.27$). Similar results were found for analyses testing group differences in the proportion of first fixations; SRs spent a significantly larger proportion of first fixations on *Faces* than did the control group ($p_{FWE} < 0.01$). Further, SRs directed a smaller proportion of first fixations toward *Text* ($p_{FWE} < 0.01$) and *Touched* objects ($p_{FWE} < 0.01$) compared with the control group. SRs did not differ from controls in the proportion of first fixations toward *Bodies* ($p_{FWE} = 0.42$).

Correlations between face cognition performance and visual saliency

Previous research has shown a correlation between the tendency to direct first fixations toward faces and CFMT performance in controls (de Haas et al., 2019). To test whether variance in face saliency is also associated with enhanced face cognition performance within the group of SRs, we computed two Pearson's correlations: (1) between observers' face cognition performance composite scores (across CFMT+, YBT long, and FICST) (cf. Nador et al., 2022) and percent dwell time toward faces, and (2) between their composite score and percent first fixations toward faces. The results show no significant correlation between face cognition performance and percent dwell time (if anything a negative trend; $r = -0.63$, $p_{FWE} = 0.10$) or percent first fixation ($r = -0.07$, $p_{FWE} = 0.85$). Note, that the limited sample size of SRs ($n = 10$) and the reduced variance in this group implied an extremely limited sensitivity of our design for this exploratory analysis.

Differences in the tendency to fixate eyes and mouths

Given that SRs exhibited a stronger tendency to fixate faces compared with controls, we next probed whether SRs and control participants differed in *how*

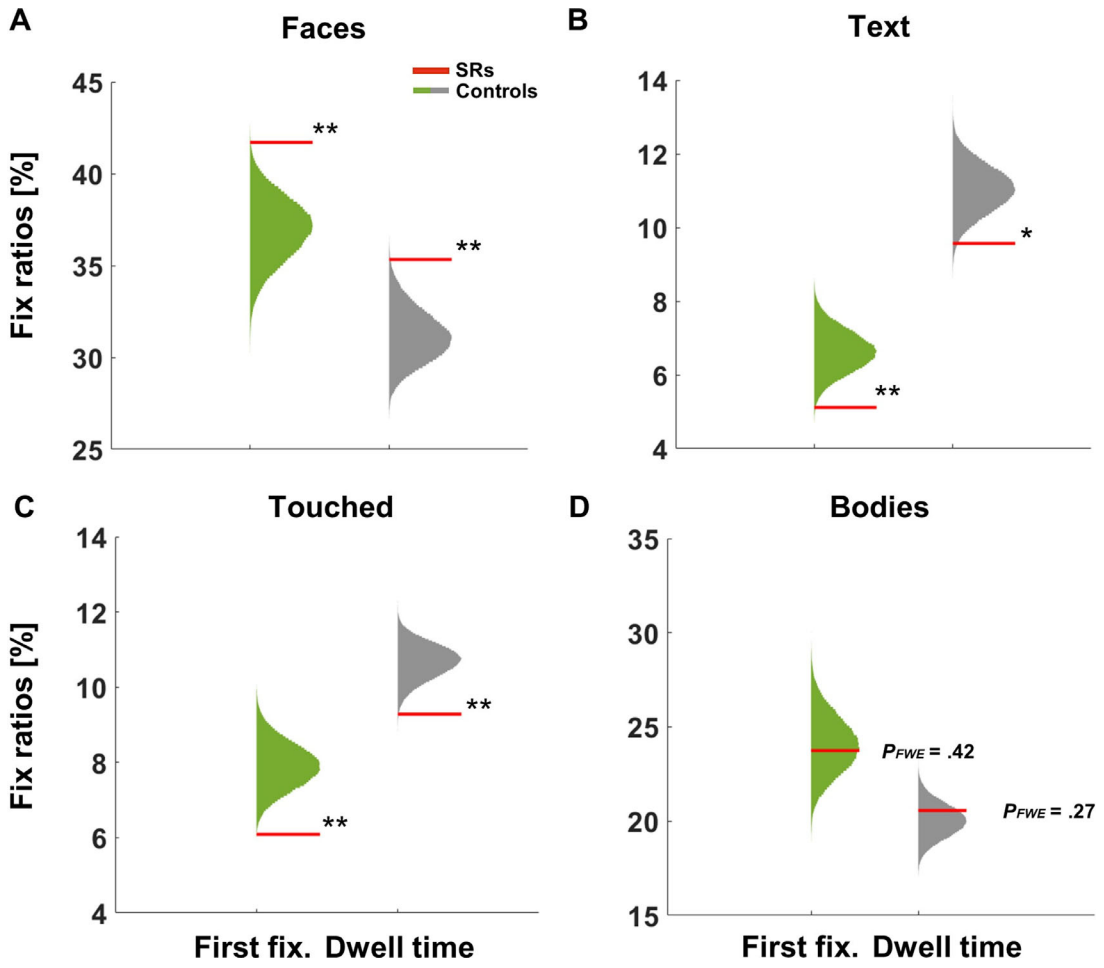

Figure 3. Group difference in gaze tendency across four semantic dimensions. (A–D) Bootstrapped null distributions of 10,000 random sample means drawn from all participants (pooled across groups) for percent first fixations (green) and percent of cumulative dwell time (gray). The superimposed red lines refer to the observed mean percent first fixations and mean percent cumulative dwell time of SRs for each given semantic dimension. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Holm–Bonferroni corrected; see Methods).

they fixated faces. We first tested group differences in the proportion of cumulative fixation time spent on faces that was directed toward Eyes and Mouths. Figure 4 shows bootstrapped mean distributions and the respective observed mean values from SRs for mean cumulative dwell-time ratios and first fixation ratios toward Eyes and Mouths. SRs spent significantly less time fixating Mouths when looking at a face than did controlse ($p_{FWE} < 0.05$). This was also true when considering the proportion of first fixations toward faces ($p_{FWE} < 0.01$). SRs and controls did not differ in percent dwell time ($p_{FWE} = 0.30$) and percent first fixation ($p_{FWE} = 0.33$) spent on Eyes.

Differences in face fixations toward the theoretical optimal fixation point

To test whether SRs compared with controls fixate closer to the optimal fixation point for identification, as predicted by an ideal Bayesian foveated observer model (Peterson & Eckstein, 2012), we computed the mean distance between (first) face fixations and the respective estimated optimal landing point for each observer. Compared with controls, SRs fixated significantly closer to the estimated optimal fixation point ($p_{FWE} < 0.01$). This also held for first fixations toward a given face ($p_{FWE} < 0.01$). Figure 5 shows the distribution of first fixations and all fixations toward faces for SRs and control subjects, as well as the average distance from the ideal fixation point for SRs and controls.

Discussion

Previous studies revealed that neurotypical observers exhibit reliable idiosyncratic gaze biases toward specific semantic stimulus categories and that their tendency to immediately fixate a face within a scene is correlated with face recognition ability (de Haas et al.,

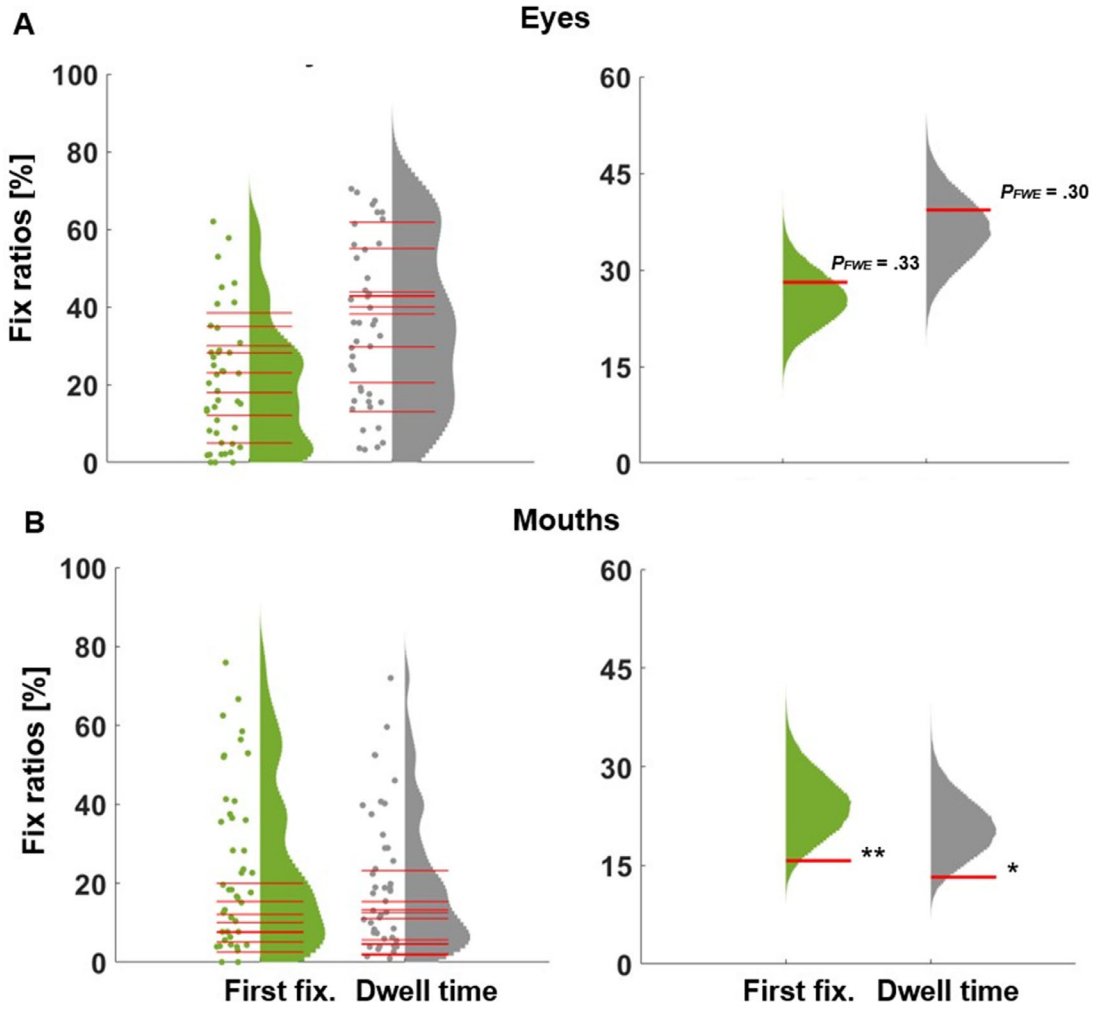

Figure 4. Fixations toward eyes and mouths. (A, B) Left-hand sides show the distributions of the control group for percent first fixations (green) and percent cumulative dwell time (gray) for fixations toward (A) eyes and (B) mouths relative to the amount of time spent looking at faces. Dots depicted in the left-hand raincloud plots indicate the corresponding individual data for each control subject and red lines indicate the fixation ratios for each SR. Right-hand sides show bootstrapped null distributions of 10,000 random sample means drawn from all participants (pooled across groups) for percent first fixations (green) and percent of cumulative dwell time (gray), respectively, for fixations toward (A) eyes and (B) mouths. The superimposed red lines refer to the corresponding observed mean of SRs. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ (Holm–Bonferroni corrected; see Methods).

2019). Extending this work, here we examined visual saliency in Super-Recognizers (SRs), individuals with exceptional face identity processing skills (Ramon, 2021; Ramon et al., 2019; Russell et al., 2009). To this end we registered oculomotor behavior of 10 SRs and 43 controls. Observers freely viewed 700 unique naturalistic scenes, depicting over 5000 objects, which were individually annotated by manually created pixel masks (Xu et al., 2014).

First, we tested whether SRs show enhanced dwell time proportions toward faces relative to objects of other semantic categories. Our findings show that SRs spend a significantly larger proportion of their dwell time fixating faces and a significantly smaller proportion of time fixating touched objects and text compared with controls. SRs and controls did not differ in proportion of dwell time spent on bodies.

We sought to further scrutinize these observations through a targeted analysis focusing on the proportion of first fixations executed after image onset. This revealed that, compared with controls, SRs showed a significantly stronger tendency to fixate faces immediately after image onset and directed a significantly lower proportion of initial fixations toward touched objects and text. In line with the findings for overall dwell time, again there was no group difference in fixations toward bodies.

Moreover, we did not find a significant correlation between the SRs' face cognition performance and their tendency to (first) fixate faces; if anything, the opposite

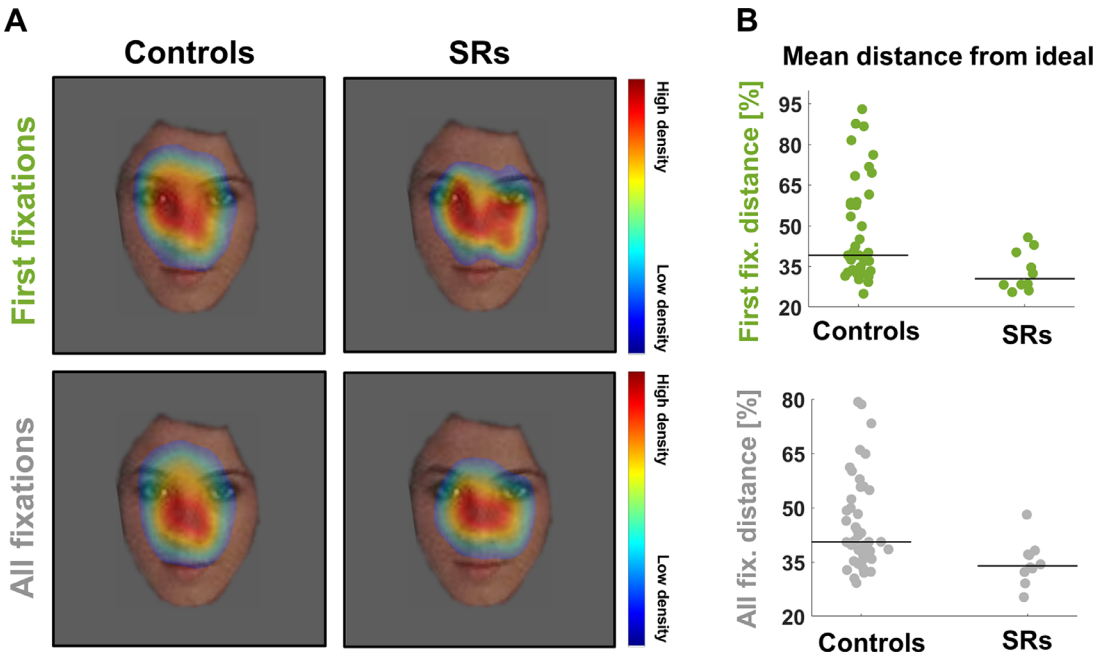

Figure 5. Results for the stimulus category *Faces*. (A) Heatmaps showing the distribution of first fixations (top row) and all fixations (bottom row) on all observed faces with an eye-to-mouth distance > 2.5 DVA superimposed over an example image (selected from one stimulus used in the free-viewing task). The heatmap was horizontally compressed to match the example face. Heatmaps on the left-hand side show fixation data from controls, maps on the right-hand side show fixations from SRs. All face fixation coordinates were normalized to the horizontal (*X*) and vertical (*Y*) extent of the respective observed face. Warmer colors indicate higher density of fixations. (B) Average of relative distances from ideal fixation point for first fixations (top row) and all fixations (bottom row) toward faces with an eye to mouth distance >2.5 DVA for each observer. Distances from the ideal fixation point just below the eyes (Peterson & Eckstein, 2012) were calculated for each face and fixation, normalized as percentages from the distance from the respective mouth centroid to the ideal fixation point and then averaged for each observer. The resulting mean distances for controls are shown on the left-hand side, those for SRs on the right-hand side. See Methods for further details.

trend for dwell time was observed. Importantly, however, these results must be interpreted with caution. We acquired data on face processing abilities only from SRs, and the small size of this group ($n = 10$) together with reduced variance within this group implies a lack of statistical power and a high probability of type 2 errors. Thus, although our data provide strong evidence for heightened face salience in SRs compared with controls, it cannot provide reliable evidence regarding the question of associations between face salience and recognition performance *within* this group. Together, our results indicate that SRs show characteristic fixation biases that differ from those shown by neurotypical controls.

Atypical saliency in relation to previously reported fixation behavior in SRs

Our findings differ from and extend the results from the only previous study on natural scene viewing in SRs in several important ways. Most strikingly, we found a significantly enhanced gaze tendency toward faces in SRs than control subjects, whereas no such effect was reported in the previous work (Bobak et al., 2016). These seemingly contradictory findings may be explained by procedural differences between studies.

First, the present study involved a more sensitive measurement (700 vs. 20 images in Bobak et al., 2016); recent findings suggest that a minimum of 40 to 100 images are necessary to reliably capture salience biases toward complex scenes (Linka & de Haas, 2020). Additionally, the SR individuals tested here were identified using a more stringent diagnostic framework (Ramon, 2021), which involves the use of a combination of the most challenging tests of face perception *and* recognition (vs. a single recognition test, as in Bobak et al., 2016).

Further, we analyzed overall dwell time, as well as first fixations. We report increased proportions of first fixations toward faces in SRs versus controls. These findings match recent evidence reported by de Haas et al. (2019) showing that individual differences in the proportion of first fixations toward faces correlate with face recognition performance in typical observers,

pointing to potential differences in bottom–up salience processing.

Finally, Bobak et al. (2016) found that SRs spent a significantly lower proportion of dwell time fixating bodies. We investigated further types of objects known to be highly salient and found no group difference for bodies but a significant tendency for SRs to fixate text and touched objects less. This is in line with the anti-correlation between individual saliency for faces, on the one hand, and text and objects being touched, on the other, which has been described for controls (de Haas et al., 2019). The general nature of this push–pull mechanism may point to SRs being at the extreme end of a general and continuous space of individual saliency, rather than showing a gaze behavior that is qualitatively different from controls.

SRs and the continuous spectrum of face salience

Together with previous work, the present findings draw a heterogenous picture of visual salience biases in the population. Across neurotypical observers, there seems to be large variability in the tendency to attend faces (de Haas et al., 2019). Certain cohorts, such as patients with autism spectrum disorder (ASD) (Wang, Jiang, Duchesne, Laugeson, Kennedy, Adolphs, & Zhao, 2015) or developmental prosopagnosia (DP) (Bobak et al., 2016), show reduced visual salience toward social stimuli such as faces, already in infancy (Constantino et al., 2017), as well as large deficits in face identity processing (Bowles et al., 2009; Griffin, Bauer, & Scherf, 2021). Notably, although most patients with ASD or DP show atypical gaze behavior toward faces, the extent can vary idiosyncratically across individuals (Pantelis & Kennedy, 2017). Here, we show that the superior face identity processing skills of SRs (Ramon, 2021; Ramon et al., 2019; Russell et al., 2009) can go along with enhanced salience for faces. However, we also observe large individual differences in gaze biases among SRs. Moreover, our preliminary findings show no significant association between fixations tendency toward faces and face processing ability within the group of SRs. This suggests that enhanced face salience is a common but most likely neither necessary nor sufficient ingredient for superior face identity processing skills. Taken together, on a population level, individual face salience seems to fall on a wide spectrum, with specific groups showing a tendency to cluster at either end rather than differing qualitatively from controls. This resembles the large individual differences reported for face identity processing skills (e.g., Bobak, Jones, Hilker, Mestry, Bate, & Hancock, 2022; Fysh & Bindemann, 2018; Fysh & Ramon, 2022; Fysh et al., 2020; Stacchi et al., 2020; Stantić et al., 2021; Stantić, Ichijo, Catmur, & Bird, 2022; Wilmer, 2017).

Functional implications of atypical saliency in SRs

Given the higher face saliency in SRs versus controls reported here, it is tempting to speculate about a potential causal relationship between individual face salience and (superior) face identity processing. Do inherent visual face biases affect face identity processing skills, or vice versa? As mentioned above, the contingency between the two is not perfect, as not all SRs show enhanced face salience and not all individuals with high face salience are SRs. Nevertheless, increased (early) face salience may promote social interaction via mutual reinforcement leading to advanced processing of facial information. This idea is partly consistent with data showing an association between extraversion and individual differences in face identity processing (Li, Tian, Fang, Xu, Li, & Liu, 2010). It also matches theories on learning style that explain difficulties in face processing in autism (Qian & Lipkin, 2011). Future research may investigate whether early tendencies to attend faces develop into corresponding levels of face identity processing abilities.

Recent twin studies in infants and children have shown strong heritability for individual face fixation biases (Constantino et al., 2017; Kennedy, D'Onofrio, Quinn, Bölte, Lichtenstein, & Falck-Ytter, 2017). On the other hand, further twin studies have shown that face identity processing abilities are heritable, as well (Wilmer et al., 2010; Zhu et al., 2010). Some studies provide evidence that face identity processing performance, as well as its genetic component, increases from preschool until adolescence (Germine, Duchaine, & Nakayama, 2011; Zhu et al., 2010). Together with the evidence for heritable gaze biases in infants, this suggests that individual salience biases may be innate and contribute to the later development of face identity processing abilities.

Retinotopic feature-location biases in SR face fixation behavior

To probe the fixation behavior of SRs *within* the semantic category of faces, we compared the proportion of dwell time and first fixations toward eyes and mouths between SRs and controls. We found that when fixating a face, compared with controls SRs fixated mouths significantly less and did not differ in the tendency to fixate eyes. The same pattern of results emerged when only the first fixation toward faces was considered. Moreover, we found that SRs fixated faces closer to

a region just below the eyes, which, according to a Bayesian ideal foveated observer model, is optimal for face identification (Peterson & Eckstein, 2012). This converges with the report by Bobak et al. (2016) of an increased fixation tendency toward noses in SRs versus typical observers (cf. Bennetts, Mole, & Bate, 2017). Interestingly, individual variance around this theoretical optimum has previously been shown to be individually optimal (Peterson & Eckstein, 2013), except for some observers suffering from developmental prosopagnosia (Peterson et al., 2019).

It is tempting to speculate that SRs may have face templates or visual field layouts aligning their individually optimal (and preferred) fixation locations more closely with the theoretical optimum than is the case for neurotypical controls. Further studies using forced fixation paradigms (Peterson & Eckstein, 2013) and more accurate, face-specific Bayesian model predictions of optimal fixation locations are required to test this. Furthermore, the stereotypical gaze behavior of SRs seems aligned with visual field tuning for the recognition of isolated facial features (de Haas & Schwarzkopf, 2018; de Haas et al., 2016) as well as their neural processing (de Haas et al., 2016; de Haas et al., 2021). Future studies should also test whether such feature-location effects are more pronounced in SRs than controls.

It is currently unclear why the individually optimal fixation point of some observers is shifted from the theoretical (and most common empirical) optimal fixation location just below the eyes. One possible explanation is that individual observers may deviate from model assumptions such as symmetric foveation. This could lead to overall detrimental effects on face identity processing performance which could partially be compensated for by altered fixation behavior. Although this remains speculative at present, it is noteworthy that our current results suggest that the fixation behavior of SRs is more tightly clustered around the theoretical optimum predicted by a foveated ideal observer than that of controls. This behavior could reflect increased spatial coverage potentially due to larger population receptive fields enabling more efficient spatial integration of information in face selective areas (Witthoft et al., 2016), or entorhinal cortex (Avidan & Behrmann, 2021).

Limitations and further considerations

One limitation of the present study is that we have no information about face identity processing ability in the control sample. This limits our power to test associations between visual salience and face cognition abilities, because we can only do so in a categorical (rather than graded) manner. Furthermore, to date there is no reliable estimate regarding the prevalence of SRs in the general population. However, note that in a recent study, Nador et al. (2022) reported that five of the final sample of 114 police professionals, whose data were subject to analysis (and six of the original 146 who completed the lab tests) met the SR criteria used here and elsewhere (Ramon, 2021; Nador et al., 2021). Therefore, although we cannot exclude the possibility that our control sample included individuals meeting SR criteria, we would expect their number to be very low.

Although here we show a significant difference in visual salience between SRs and controls, there is considerable overlap between both groups (see Figure 2). This overlap could potentially be driven by control observers located at the upper end of the face identity processing spectrum. Future research may additionally test face identity processing abilities and use regression approaches to exploit individual variance across the whole spectrum.

One could reasonably argue that the present findings could have been impacted by an implicit factor related to the experimental design: SRs who volunteered to participate in the present study were aware of their superior face identity processing skills. Therefore, they may have focused toward faces more often and/or longer, purely because they expected the experiment to address questions related to face cognition.

However, compared with controls, SRs showed stronger differences in their tendency to first fixate faces versus non-face objects than in their allocation of overall dwell time. Previous studies have shown very low latencies for first saccades after image onset directed toward faces. Specifically, observers have a hard time suppressing such face-directed fixations (Crouzet et al., 2010). Thus, it seems unlikely that these effects are significantly impacted by top–down expectancy effects.

Here, we investigated individual differences in fixation behavior during free viewing of complex static scenes containing people and other objects. Such individual differences have been shown to be highly stable over time (de Haas et al., 2019). Furthermore, individual differences in fixation behavior *within* faces has been shown to be extremely stable across lab-based identification tasks and real world free-viewing (Peterson, Lin, Zaun,& Kanwisher, 2016). Nevertheless, at the group level, face fixation behavior varies with task (Buchan, Paré, & Munhall, 2007) and between static and dynamic stimuli (Võ, Smith, Mital, & Henderson, 2012). Future studies will have to determine whether and how individual and group differences vary with task and stimulus modality.

Conclusions

In sum, we find that SRs compared with controls exhibit significantly enhanced gaze biases toward

faces and significantly diminished tendencies to fixate touched objects and text when freely viewing complex scenes. These differences are evident from the first fixation onward, suggesting a substantial bottom–up component in salience processing. However, the distributions of the fixation tendencies of SRs and control observers are overlapping, and the push–pull relationship between different fixation tendencies is seen in both groups. This suggests a continuous space of individual salience in which SRs cluster toward the high face salience end (rather than inhabit a qualitatively different space). Finally, within the category of faces, SRs show a more stereotypical fixation pattern, exhibiting a stronger tendency than controls to preferentially fixate near the theoretical optimum just below the eyes. Further research is needed to probe the developmental and causal relationships between individual gaze behavior and face identity processing skills.

Keywords: super-recognizers, face identity processing, perception and recognition, visual salience, eye movements

Acknowledgments

Supported by an ERC Starting Grant (852885 INDIVISUAL to SuML, MB, and BdH); by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 222641018–SFB/TRR 135 TP C9 (BdH); by "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art (BdH); and by a grant from the Swiss National Science Foundation PRIMA (Promoting Women in Academia; PR00P1_179872 to MR).

Commercial relationships: none.
Corresponding author: Meike Ramon.
Email: meike.ramon@gmail.com.
Address: Applied Face Cognition Lab, Institute of Psychology, University of Lausanne, Lausanne, Switzerland.

*BDH and MR contributed equally to this work.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T., van Langen, J., & Kievit, R. (2019). Raincloud plots: A multi-platform tool for robust data visualization [version 2; peer review: 2 approved]. *Wellcome Open Research, 4*, 16, https://doi.org/10.12688/wellcomeopenres.15191.2.

Amestoy, A., Guillaud, E., Bouvard, M. P., & Cazalets, J.-R. (2015). Developmental changes in face visual scanning in autism spectrum disorder as assessed by data-based analysis. *Frontiers in Psychology, 6*, 989, https://doi.org/10.3389/fpsyg.2015.00989.

Arizpe, J., Walsh, V., Yovel, G., & Baker, C. I. (2017). The categories, frequencies, and stability of idiosyncratic eye-movement patterns to faces. *Vision Research, 141*, 191–203, https://doi.org/10.1016/j.visres.2016.10.013.

Avidan, G., & Behrmann, M. (2021). Spatial integration in normal face processing and its breakdown in congenital prosopagnosia. *Annual Review of Vision Science, 7*, 301–321, https://doi.org/10.1146/annurev-vision-113020-012740.

Bargary, G., Bosten, J.M., Goodbourn, P.T., Lawrance-Owen, A.J., Hogg, R.E., & Mollon, D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research, 141*, 157–169, https://doi.org/10.1016/j.visres.2017.03.001.

Bennetts, R. J., Mole, J., & Bate, S. (2017). Super-recognition in development: a case study of an adolescent with extraordinary face recognition skills. *Cognitive Neuropsychology, 34*, 357–376, https://doi.org/10.1080/02643294.2017.1402755.

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2016). Eye-movement strategies in developmental prosopagnosia and "super" face recognition. *Quarterly Journal of Experimental Psychology, 70*, 201–217, https://doi.org/10.1080/17470218.2016.1161059.

Bobak, A.K., Jones, A.L., Hilker, Z., Mestry, N., Bate, S., & Hancock, P.J.B. (2022). Commonly used face cognition tests yield low reliability and inconsistent performance: Implications for test design, analysis, and interpretation of individual differences data. Manuscript in preparation.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., & Schmalzl, L., …Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology, 26*, 423–455, https://doi.org/10.1080/02643290903343149.

Broda, M. D., & de Haas, B. (2022). Individual differences in looking at persons in scenes. Retrieved from https://psyarxiv.com/xyz9u/.

Bruck, M., Cavanagh, P., & Ceci, S. (1991). Fortysomething: Recognizing faces at one's 25th reunion. *Memory & Cognition, 19*, 221–228, https://doi.org/10.3758/BF03211146.

Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic

face processing. *Social Neuroscience, 2*, 1–13, https://doi.org/10.1080/17470910601043644.

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 is atypical in autism. *Nature, 547*, 340–344, https://doi.org/10.1038/nature22999.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision, 10*(4):16, 1–17, https://doi.org/10.1167/10.4.16.

de Haas, B. (2018). How to enhance the power to detect brain–behavior correlations with limited resources. *Frontiers in Human Neuroscience, 12*, 421, https://doi.org/10.3389/fnhum.2018.00421.

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences, USA, 116*, 11687, https://doi.org/10.1073/pnas.1820553116.

de Haas, B., & Schwarzkopf, D. S. (2018). Feature–location effects in the Thatcher illusion. *Journal of Vision, 18*(*4*):16, 1–12, https://doi.org/10.1167/18.4.16.

de Haas, B., Schwarzkopf, D. S., Alvarez, I., Lawson, R. P., Henriksson, L., Kriegeskorte, N., . . . Rees, G. (2016). Perception and processing of faces in the human brain is tuned to typical feature locations. *Journal of Neuroscience, 36*, 9289–9302, https://doi.org/10.1523/JNEUROSCI.4131-14.2016.

de Haas, B., Sereno, M. I., & Schwarzkopf, D. S. (2021). Inferior occipital gyrus is organized along common gradients of spatial and face-part selectivity. *The Journal of Neuroscience, 41*, 5511, https://doi.org/10.1523/JNEUROSCI.2415-20.2021.

Fedor, J., Lynn, A., Foran, W., DiCicco-Bloom, J., Luna, B., & O'Hearn, K. (2018). Patterns of fixation during face recognition: Differences in autism across age. *Autism, 22*, 866–880, https://doi.org/10.1177/1362361317714989.

Fysh, M. C., & Bindemann, M. (2018). Human–Computer interaction in face matching. *Cognitive Science, 42*, 1714–1732, https://doi.org/10.1111/cogs.12633.

Fysh, M. C., & Ramon, M. (2022). Accurate but inefficient: Standard face identity matching tests fail to identify prosopagnosia. *Neuropsychologia, 165*, 1–10, https://doi.org/10.1016/j.neuropsychologia.2021.108119.

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: Implications for theory, research and personnel selection. *Royal Society Open Science, 7*, 200233, https://doi.org/10.1098/rsos.200233.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition, 118*, 201–210, https://doi.org/10.1016/j.cognition.2010.11.002.

Geskin, J., & Behrmann, M. (2018). Congenital prosopagnosia without object agnosia? A literature review. *Cognitive Neuropsychology, 35*, 4–54, https://doi.org/10.1080/02643294.2017.1392295.

Griffin, J. W., Bauer, R., & Scherf, K. S. (2021). A quantitative meta-analysis of face recognition deficits in autism: 40 years of research. *Psychological Bulletin, 147*, 268–292, https://doi.org/10.1037/bul0000310.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*, 313–323, https://doi.org/10.1016/j.cognition.2011.08.001.

Kanan, C., Bseiso, D.N., Ray, N.A., Hsiao, J.H., & Cottrell, G.W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research, 108*, 67–76, https://doi.org/10.1016/j.visres.2015.01.013.

Kennedy, D.P., D'Onofrio, B.M., Quinn, P.D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic influence on eye movements to complex scenes at short timescales. *Current Biology, 27*, 3554–3560.e3, https://doi.org/10.1016/j.cub.2017.10.007.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception, 36*, 1–16.

Li, J., Tian, M., Fang, H., Xu, M., Li, H., & Liu, J. (2010). Extraversion predicts individual differences in face recognition. *Communicative & Integrative Biology, 3*, 295–298, https://doi.org/10.4161/cib.3.4.12093.

Linka, M., & de Haas, B. (2020). OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *Journal of Vision, 20*(9):13, 1–9, https://doi.org/10.1167/jov.20.9.13.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision, 14*(7):6, 1–11, https://doi.org/10.1167/14.7.6.

Nador, J. D., Alsheimer, T. A., Gay, A., & Ramon, M. (2021). Image or identity? Only super-

recognizers' (memor)ability is consistently viewpoint-invariant. *Swiss Psychology Open, 1*, 2, http://doi.org/10.5334/spo.28.

Nador, J., Vomland, M., Thielgen, M., & Ramon, M. (2022). Face recognition in police officers: Who fits the bill? *Forensic Science International: Reports, 5*, 100267, https://doi.org/10.1016/j.fsir.2022.100267.

Nador, J. D., Zoia, M., Pachai, M. V., & Ramon, M. (2021). Psychophysical profiles in super-recognizers. *Scientific Reports, 11*, 13184, https://doi.org/10.1038/s41598-021-92549-6.

Pantelis, P.C., & Kennedy, D.P. (2017). Deconstructing atypical eye gaze perception in autism spectrum disorder. *Scientific Reports, 7*, 14990, https://doi.org/10.1038/s41598-017-14919-3.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, USA, 109*, E3314–E3323, https://doi.org/10.1073/pnas.1214269109.

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science, 24*, 1216–1225, https://doi.org/10.1177/0956797612471684.

Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of Vision, 16*(7):12, 1–18, https://doi.org/10.1167/16.7.12.

Peterson, M. F., Zaun, I., Hoke, H., Jiahui, G., Duchaine, B., & Kanwisher, N. (2019). Eye movements and retinotopic tuning in developmental prosopagnosia. *Journal of Vision, 19*(9):7, 1–19, https://doi.org/10.1167/19.9.7.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., & Jackson, K., …O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, super recognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences, USA, 115*, 6171, https://doi.org/10.1073/pnas.1721355115.

Poltoratski, S., Kay, K., Finzi, D., & Grill-Spector, K. (2021). Holistic face recognition is an emergent phenomenon of spatial processing in face-selective regions. *Nature Communications, 12*, 4745, https://doi.org/10.1038/s41467-021-24806-1.

Qian, N., & Lipkin, R. M. (2011). A learning-style theory for understanding autistic behaviors. *Frontiers in Human Neuroscience, 5*, 77, https://doi.org/10.3389/fnhum.2011.00077.

Ramon, M. (2018). The power of how-lessons learned from neuropsychology and face processing. *Cognitive Neuropsychology, 35*, 83–86, https://doi.org/10.1080/02643294.2017.1414777.

Ramon, M. (2021). Super-Recognizers - a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia, 158*, 107809, https://doi.org/10.1016/j.neuropsychologia.2021.107809.

Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology, 110*, 461–479, https://doi.org/10.1111/bjop.12368.

Rossion, B. (2014). Understanding individual face discrimination by means of fast periodic visual stimulation. *Experimental Brain Research, 232*, 1599–1621, https://doi.org/10.1007/s00221-014-3934-9.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review, 16*, 252–257, https://doi.org/10.3758/PBR.16.2.252 .

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive Research: Principles and Implications, 5*, 8, https://doi.org/10.1186/s41235-019-0205-0.

Stacchi, L., Liu-Shuang, J., Ramon, M., & Caldara, R. (2019). Reliability of individual differences in neural face identity discrimination. *NeuroImage, 189*, 468–475, https://doi.org/10.1016/j.neuroimage.2019.01.023.

Stantić, M., Brewer, R., Duchaine, B., Banissy, M.J., Bate, S., & Susilo, T., …Bird, G. (2021). The Oxford Face Matching Test: A non-biased test of the full range of individual differences in face perception. *Behavior Research Methods*. *54*, 158–173, https://doi.org/10.3758/s13428-021-01609-2.

Stantić, M., Ichijo, E., Catmur, C., & Bird, G. (2022). Face memory and face perception in autism. *Autism, 26*, 276–280, https://doi.org/10.1177/13623613211027685.

Tanaka, J. W., & Sung, A. (2016). The "eye avoidance" hypothesis of autism face processing. *Journal of Autism and Developmental Disorders, 46*, 1538–1552, https://doi.org/10.1007/s10803-013-1976-7.

Tang, J., Falkmer, M., Horlin, C., Tan, T., Vaz, S., & Falkmer, T. (2015). Face recognition and visual search strategies in autism spectrum disorders: amending and extending a recent review by Weigelt et al. *PLoS One, 10*, e0134439, https://doi.org/10.1371/journal.pone.0134439.

Võ, M. L., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing

moving faces. *Journal of Vision, 12*(13):, 1–14, https://doi.org/10.1167/12.13.3.

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., … Zhao, Q. (2015). Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron, 88*, 604–616, https://doi.org/10.1016/j.neuron.2015.09.042.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews, 36*, 1060–1084, https://doi.org/10.1016/j.neubiorev.2011.12.008.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., & Loken, E., …Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences, USA, 107*, 5238, https://doi.org/10.1073/pnas.0913053107.

Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science, 26*, 225–230, https://doi.org/10.1177/0963721417710693.

Witthoft, N., Poltoratski, S., Nguyen, M., Golarai, G., Liberman, A., & LaRocque, K. F., …Grill-Spector, K. (2016). Reduced spatial integration in the ventral visual cortex underlies face recognition deficits in developmental prosopagnosia [preprint]. bioRxiv, https://doi.org/10.1101/051102.

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision, 14*(1):28, 1–20, https://doi.org/10.1167/14.1.28.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., & Zhen, Z., …Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology, 20*, 137–142, https://doi.org/10.1016/j.cub.2009.11.067.

Study 2

Linka, M., & de Haas, B. (2020). OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *Journal of Vision*, 20(9), 13.

OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience

Marcel Linka
Experimental Psychology, Justus Liebig Universität, Giessen, Germany

Benjamin de Haas
Experimental Psychology, Justus Liebig Universität, Giessen, Germany

Recent findings revealed consistent individual differences in fixation tendencies among observers free-viewing complex scenes. The present study aimed at (1) replicating these differences, and (2) testing whether they can be estimated using a shorter test. In total, 103 participants completed two eye-tracking sessions. The first session was a direct replication of the original study, but the second session used a smaller subset of images, optimized to capture individual differences efficiently. The first session replicated the large and consistent individual differences along five semantic dimensions observed in the original study. The second session showed that these differences can be estimated using about 40 to 100 images (depending on the tested dimension). Additional analyses revealed that only the first 2 seconds of viewing duration seem to be informative regarding these differences. Taken together, our findings suggest that reliable individual differences in semantic salience can be estimated with a test totaling less than 2 minutes of viewing duration.

Introduction

In order to make sense of our visual environment, we constantly move our eyes. This allows us to observe fixed and moving objects with high foveal resolution and to inspect regions of interest while ignoring many others (Gegenfurtner, 2016). To explain this selection process, models of attentional guidance have tried to predict gaze behavior based on, for example, task relevance (Fecteau & Munoz, 2006); low-level image features, such as local contrast for color, intensity, and orientation (Harel, Koch, & Perona, 2006; Itti & Koch, 2000); and high-level, semantic features, such as faces or text (Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014). What these models all have in common is that they try to predict a typical observer and treat individual differences as noise.

More recent findings, however, show stable, trait-like differences in eye movements; for example, oculomotor measures such as mean saccade amplitude, smooth pursuit duration, or mean fixation duration vary reliably among people (Bargary, Bosten, Goodbourn, Lawrance-Owen, Hogg, & Mollon, 2017; Castelhano & Henderson, 2008; Henderson & Luke, 2014). Moreover, recent twin studies have provided evidence for a strong genetic component in individual gaze (Constantino et al., 2017; Kennedy, D'Onofrio, Quinn, Bölte, Lichtenstein, & Falck-Ytter, 2017).

Most relevant to the current study, observers freely viewing hundreds of complex scenes showed large individual differences in the number of fixated objects and in fixation tendencies toward objects from six semantic categories (de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2019). These differences were consistent across images and time and extended to first fixations after image onset, suggesting a bottom–up component.

Such semantic salience biases may be useful in the study of neurobiological mechanisms of attentional gaze control (de Haas et al., 2019) and form a crucial baseline for evaluating the diagnostic potential of gaze behavior for neurodevelomental and clinical conditions. In fact, research investigating eye movements in autism spectrum disorder (ASD) has demonstrated reduced social visual engagement in infants and adults with ASD compared to healthy controls (Constantino et al., 2017; Jones & Klin, 2013; Wang, Jiang, Duchesne, Laugeson, Kennedy, Adolphs, & Zhao, 2015). Others have found evidence for abnormal eye movements in patients with major depression (Armstrong & Olatunji, 2012). Further, patients with schizophrenia show a reduced number and spatial dispersion of fixations (Benson, Beedie, Shephard, Giegling, Rujescu, & St. Clair, 2012). However, the free viewing paradigm in de Haas et al. (2019) used the full stimulus set of the Object and Semantic Images and Eye-tracking (OSIE) dataset, comprised of 700 images (Xu et al., 2014). For practical

Citation: Linka, M., & de Haas, B. (2020). OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *Journal of Vision*, *20*(9):13, 1–9, https://doi.org/10.1167/jov.20.9.13.

https://doi.org/10.1167/jov.20.9.13 Received May 7, 2020; published September 18, 2020 ISSN 1534-7362

purposes, it would be desirable to estimate individual gaze biases with a more economical test.

The present study had two main objectives. The first was to replicate the findings of stable individual gaze differences along semantic dimensions found in de Haas et al. (2019). Therefore, we tested whether the proportion of cumulative fixation times for objects of the categories *Text* and *Faces*, objects with implied *Motion*, objects with a characteristic *Taste*, or *Touched* objects vary reliably between observers. Further, we tested whether the proportion of first fixations varies reliably for objects of the categories *Text*, *Faces*, and *Touched*. We chose and preregistered these object dimensions, because they yielded large and consistent individual gaze differences ($r > 0.6$) in the original study by de Haas et al. (2019). Note, we combined neutral and emotional *Faces*, given the high covariance in fixation tendencies toward both. Moreover, we tried to replicate the findings of stable individual differences in visual exploration, as indicated by the number of objects an observer fixated. Second, we tested whether a smaller stimulus set can reliably estimate these differences. For this purpose, we selected subsets of OSIE images ($OSIE_{40}$, $OSIE_{100}$, and $OSIE_{200}$), which a search algorithm predicted would yield high congruence with individual differences observed for the full set.

Two eye-tracking sessions were completed by 103 participants on separate days; one day they free-viewed the full OSIE set (700 images), and another day they free-viewed the smaller subsets. We then computed the correlation among individual fixation tendencies between the full and smaller sets to determine the minimum set size required for reliable estimates. Additionally, we repeated these analyses for gaze data truncated to the first 1 and 2 seconds of each trial (from a total of 3 seconds). Results showed that the most prominent individual fixation biases could be estimated reliably from just 40 images shown for 2 seconds each.

Methods

Subjects

A total of 103 healthy participants with normal or corrected-to-normal vision were recruited at Leibniz Institute of Psychology Information (ZPID) using their PsychLab offline service ($M_{age} = 25.17$; $SD = 5.50$; seven left-handed; 72 females). The sample size was based on an a priori power analysis (de Haas, 2019). All participants took part in two sessions, with the second appointment following the first after an average of 16 days. The study was approved by the local ethics committee, and all participants gave written informed consent before participating.

Apparatus

Stimuli were displayed on a BenQ XL2430T monitor (BenQ Corporation, Taipei, Taiwan) using Psychtoolbox 3.0.12 (Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard, 2007; Pelli, 1997) in MATLAB R2019a (MathWorks, Natick, MA). Participants viewed the stimuli at a resolution of 1920 × 1080 pixels and at 29.0 × 22.2 degrees visual angle. Eye gaze from the left eye was measured using a desktop-mounted EyeLink 1000 Plus eye tracker (SR Research, Ottawa, Canada) at a frequency of 1 kHz.

Stimuli and procedure

Participants completed free-viewing tasks that included a set of 700 images (day 1) and 200 images (day 2) with natural everyday scenes, each of which contained multiple objects. Semantic metadata for the full stimulus set consisted of binary pixel masks for 5551 objects and corresponding labels for 12 semantic dimensions as provided in the OSIE dataset (Xu et al., 2014). The provided labels were modified in order to minimize overlap between them, as in de Haas et al. (2019). Specifically, the *Smell* label was removed from all objects with the *Text* label, the *Operable* and *Gazed* labels were removed from all objects with the *Touched* label, and the *Watchable* label was removed from all objects with the *Text* label.

The design and procedure for both testing days were adopted from de Haas et al. (2019). After calibration, participants freely viewed seven blocks of 100 images each while sitting at a distance of ∼64 cm from the computer screen. Each trial started with the presentation of a central fixation disk followed by a 3-second presentation of the image that was initiated by a button press by the participant. All images appeared in the same order across participants. During the second appointment, this procedure was repeated with only two blocks (200 images).

The order of images presented during the first two blocks of the first day was determined a priori, based on the dataset by de Haas et al. (2019). An iterative search algorithm aimed at selecting images in an order that maximized the predictive validity for the full set at each step (i.e., for each number of images between 10 and 200, incrementing by one image). The order of images on the second day was pseudo-randomized relative to the first two blocks of the first day to minimize possible effects of order expectation. The order of images was randomized separately for images 1 to 40, 41 to

100, and 101 to 200 of the first day, but this division had no further relevance for the analyses performed here.

Analysis

All statistical analyses were performed in MATLAB R2019a (MathWorks). Data processing closely followed the procedures in de Haas et al. (2019) and was preregistered (de Haas, 2019). To limit our dataset to object-directed fixations after stimulus onset, we excluded central onset fixations (onset time < 100 ms), fixations with duration below 100 ms (following standard settings from the eye tracker manufacturer), and fixations that could not be assigned a label. Fixations that fell on or within a distance of ~0.5 degrees visual angle from a labeled object were assigned the corresponding label. To assess individual fixation tendencies, for each observer we computed the proportion of first fixations (after image onset) and of cumulative dwell time falling onto objects of a given label.

Consistency and retest correlations

We used the data from the first testing day to replicate the study by de Haas et al. (2019), testing the split-half consistency of individual fixation proportions for each semantic dimension. This was repeated across 1000 random splits of images. We calculated two-sided p values for the median correlations across these splits, which were Bonferroni adjusted for the number of tested dimensions: *Faces*, *Text*, *Touched*, *Taste*, and *Motion* for cumulative dwell times and *Faces*, *Text*, and *Touched* for the proportion of first fixations, as preregistered (de Haas et al., 2019). Finally, we calculated the minimum/maximum ratios of individual fixations between observers and for each dimension. This was done by dividing the minimum percent cumulative dwell time and percent first fixation for a given dimension by the corresponding maximum across all observers.

To test the validity of estimates from smaller stimulus sets, we calculated fixation proportions as above, but for 190 subsets (from 10 to 200 images in steps of one). Next, we determined the correlation of individual differences for a given dimension and subset from day 2 with those observed for the full set on day 1. Note that only the preregistered set sizes of 40, 100, and 200 images contained identical images, because of the pseudo-randomized order between days (see above). We also probed the consistency of individual differences in the number of objects fixated, an indicator of the individual tendency for visual exploration. This part of the analyses was not preregistered but replicated de Haas et al. (2019).

Further, we explored the effects of shorter trial durations (also not preregistered). First, we probed the consistency of individual differences when truncating viewing data from day 1 to the first n milliseconds of each trial. For this analysis, we computed the proportion of cumulative dwell times for each semantic dimension considered and individual dwell times for different steps of truncation, from 100 ms to 3 seconds in steps of 100 ms. We then correlated these proportions with those observed for the full trial duration. Second, we computed split-half correlations between cumulative dwell time proportions across odd- and even-numbered images for each step of truncation. Third, we repeated all consistency analyses for smaller stimulus sets with gaze data trimmed to the first 1 and 2 seconds of each trial (data truncated for both day 1 and day 2).

To determine whether a subset of images (presented on day 2) yielded a valid estimate of individual differences along a given dimension, we tested its correlation with individual differences along the same dimension observed for the full set (on day 1). For this analysis, we applied a preregistered correlation threshold of $r = 0.7$. This criterion was very conservative, given the noise ceiling of full sets shown on separate days. Test–retest correlations for the full set ranged from $r = 0.68$ (*Motion*) to $r = 0.84$ (*Text*) in de Haas et al. (2019). We therefore additionally report the proportion of explainable variance explained by a given subset, relative to this estimate of the noise ceiling.

Prevalence of image features across subsets

Moreover, to explore possible differences in the prevalence of image features that may lead to different consistency estimates across subsets, we additionally determined the relative and absolute number of occurrences for each semantic image feature for each subset.

Data availability and preregistration

All anonymized data and code to reproduce the presented results, as well as the stimulus sets, are available at https://osf.io/ekvj4/. A preregistered study protocol is available at http://dx.doi.org/10.23668/psycharchives.2454.

Results

First, we replicated the large and consistent individual differences found by de Haas et al. (2019). Proportions of cumulative fixation time varied strongly among observers, with maximum/minimum ratios

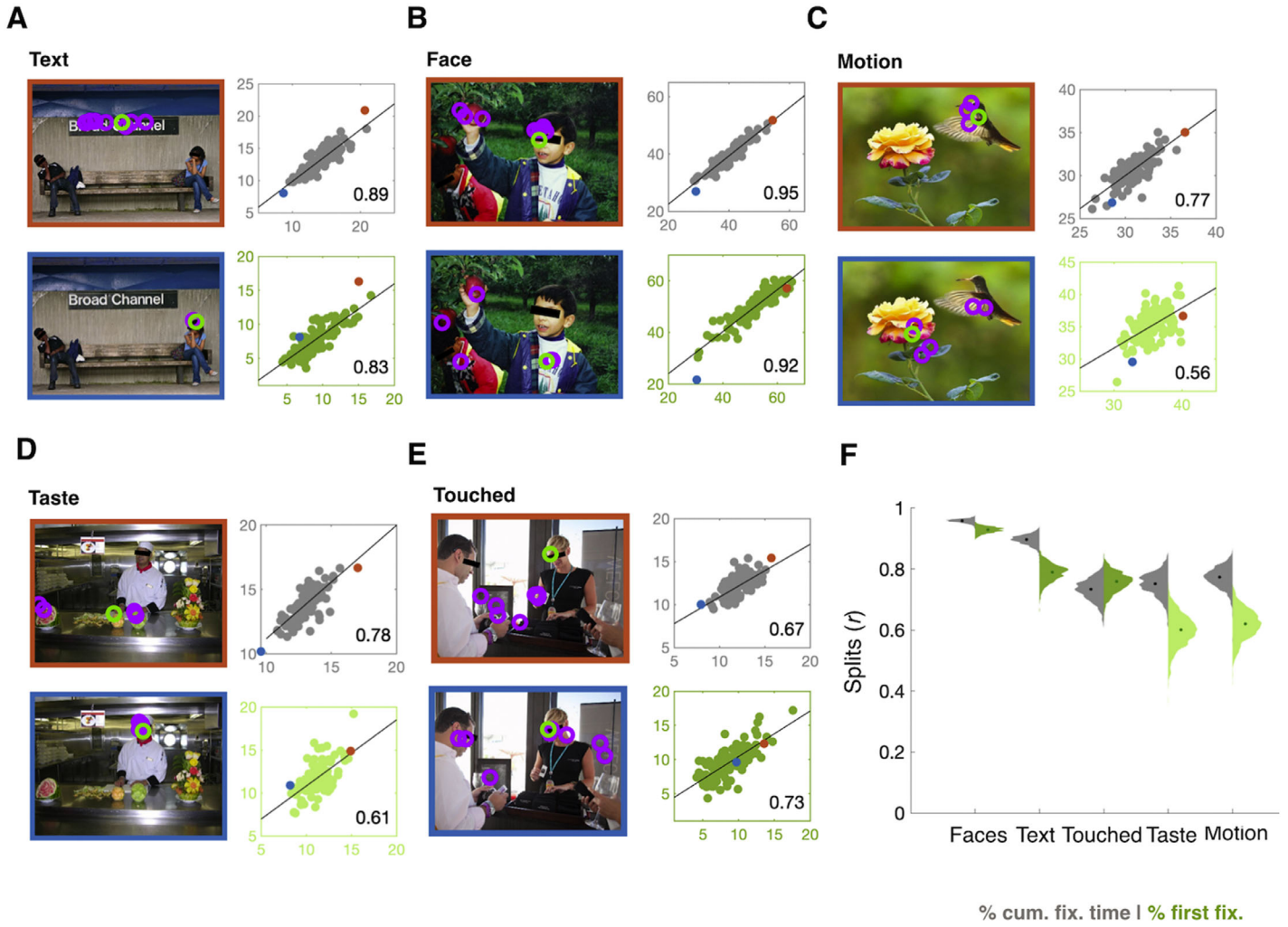

Figure 1. Consistency of fixation ratios. (A–E) Scatterplots show the split-half correlation between odd- versus even-numbered images for percent cumulative fixation time (gray) for the dimensions *Text*, *Faces*, *Motion*, *Taste*, and *Touched* and percent first fixations (green) for the dimensions *Text*, *Faces*, and *Touched*. These dimensions were preregistered and included based on the large and reliable individual differences along them ($r > 0.6$) reported by de Haas et al. (2019). Black inset numbers give the respective Pearson's correlation coefficient. For each dimension, one example is given showing the fixations from a participant with a strong tendency to fixate objects of the given dimension (orange frames) and one participant with a weaker tendency (blue frame). The green circles represent the first fixation after image onset, and purple circles represent any subsequent fixations. The orange and blue data points in the scatterplots correspond to the respective example participants and image frame color. Black bars in the images were added for display purposes only and were not seen during the free-viewing task. (F) Distribution of bootstrapped split-half correlations for each of the five included the semantic dimensions *Faces*, *Text*, *Touched*, *Taste*, and *Motion* (*x*-axis). The gray left-hand leaves show the distribution of split-half correlations for cumulative fixation time between image sets for 1000 random half-splits of images. The green right-hand leaves show the corresponding distributions for first fixations. Dots inside the distributions refer to the median correlation for each histogram and semantic dimension. Desaturated green scatterplots (A–D) and right-hand leaves (F) for the dimensions *Motion* and *Taste* are shown for completeness and were not preregistered (de Haas et al., 2019).

ranging from 1.37 (*Motion*) to 2.47 (*Text*). These differences were highly reliable across image splits for each semantic dimension (from $r = 0.74$, $p < 0.001$ for *Touched* to $r = 0.96$, $p < 0.001$ for *Faces*). Similar results were observed for the proportion of first fixations, with maximum/minimum ratios ranging from 2.40 for *Faces* to 3.47 for *Touched* and consistency ranging from $r = 0.74$, $p < 0.001$ for *Touched* to $r = 0.93$, $p < 0.001$ for *Faces* (Figure 1). Similarly, the number of distinct objects fixated consistently varied across observers, with a maximum/minimum ratio of 1.95 ($r = 0.97$, $p < 0.001$). Table 1 provides an overview of the results and gives the ranges of fixation proportions across observers.

	Cumulative fixation time				First fixations			
	r	*P*	Range (%)	Max/min	*r*	*P*	Range (%)	Max/min
Faces	0.96	<0.001	28–53	1.90	0.93	<0.001	26–63	2.40
Text	0.90	<0.001	8–21	2.50	0.79	<0.001	4–16	3.54
Touched	0.74	<0.001	9–16	1.73	0.76	<0.001	5–17	3.47
Taste	0.75	<0.001	10–17	1.71	0.61	<0.001	8–17	2.09
Motion	0.78	<0.001	26–36	1.37	0.62	<0.001	28–40	1.42

Table 1. Median consistency and corresponding range of individual differences in fixation toward objects of five semantic dimensions. The left-hand side of the table shows results referring to individual differences in the proportion of cumulative dwell time toward objects of the five semantic attributes included in the study; the right-hand side displays results based on the proportion of first fixations landing on an object with a given attribute after image onset. Pearson correlations (*r*) indicate the median split-half correlation of individual differences measured over 1000 random image splits, and the corresponding *p* values are Bonferroni adjusted for the five semantic attributes. The range across observers is presented in percent cumulative fixation time and percent first fixation and the maximum/minimum ratio (max/min) indicates the maximum relative difference in individual gaze difference between observers.

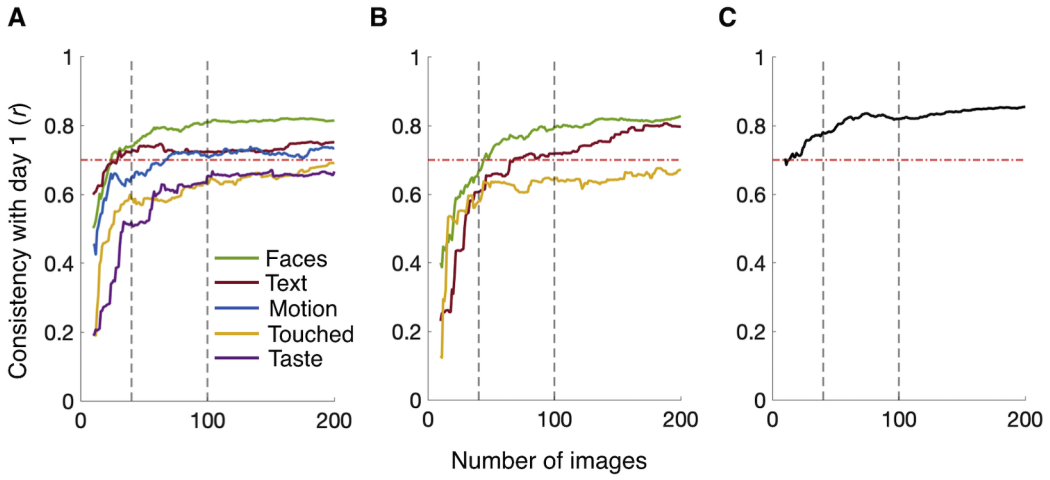

Figure 2. Consistency of fixation ratios based on smaller stimulus sets. Line plots depicting Pearson consistency correlations of individual differences in percent cumulative fixation time (A), percent first fixations (B), and visual exploration (number of distinct objects fixated) (C) between the full image set (day 1) and a given subset (10 to 200 images from day 2; *x*-axis). Dashed gray lines mark stimulus sets of 40 and 100 images. Dashed red lines indicate the consistency threshold of $r = 0.7$. Colors in (A) and (B) indicate semantic dimensions as shown in the inset.

Estimating individual differences with a smaller stimulus set

Next, we probed whether these differences can be tested with a smaller set of images. Figure 2 shows the consistency (Pearson's *r*) of estimates for smaller stimulus sets from day 2 compared with those for the full dataset from day 1 as a function of set size, both for cumulative dwell time (Figure 2A) and first fixations (Figure 2B). Figure 2C shows the same analysis for visual exploration. For cumulative dwell times, a stimulus set of just 40 images yielded reliable estimates ($r > 0.7$, $p < 0.001$ with the full set) for individual differences along the dimensions *Text* and *Faces*. We found that 100 and 200 images reliably captured *Text*, *Faces*, and *Motion*, with the dimensions *Taste* and *Touched* being just below this threshold for the set of 200 images ($r = 0.67$, $p < 0.001$ for *Taste* and $r = 0.69$, $p < 0.001$ for *Touched*). For first fixations, a set of 100 images captured reliable inter-individual differences ($r > 0.7$, $p < 0.001$ with the full set) for *Text* and *Faces*. *Touched* came close to this threshold ($r = 0.65$, $p < 0.001$). Individual differences in visual exploration could reliably be estimated with a set size of 40 images ($r = 0.78$, $p < 0.001$).

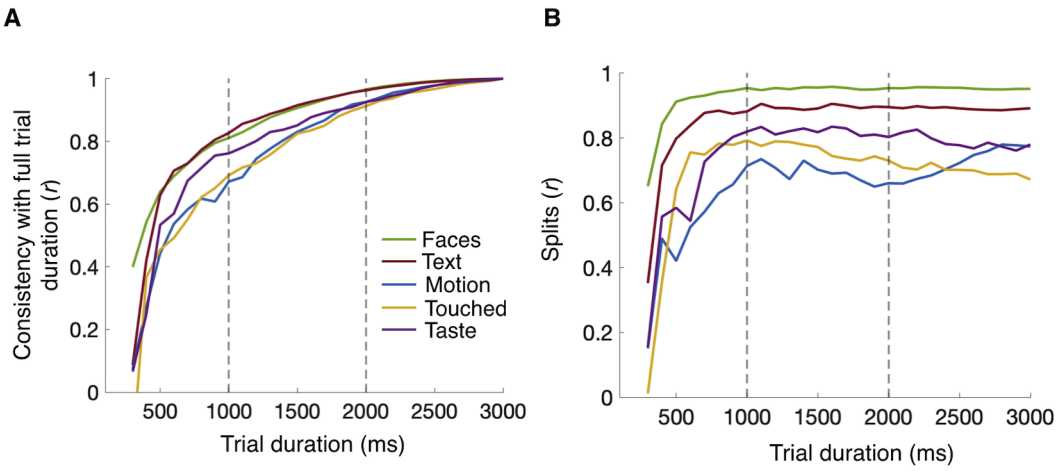

Figure 3. Consistency of estimates for truncated trial durations. (A) Line plots show the correlation of percent cumulative fixation time observed for the full trial duration with the corresponding values for different steps of truncation (starting from 1000 ms to 3000 ms in steps of 100 ms as shown on the *x*-axis; all data from day 1). (B) Line plot on the right-hand side presents split-half correlations across odd and even images for different steps of truncation. Dashed gray lines mark free viewing durations of 1000 ms and 2000 ms. Semantic dimensions are indicated by color as shown in the insets.

Estimating individual differences with shorter trial durations

We further explored the effects of (virtually) shortened trial durations by trimming the data from each trial. Figure 3A considers estimated individual differences in cumulative dwell time from day 1 and shows the consistency correlation between these estimates based on the whole trial duration of 3 seconds versus the same estimates for trimmed durations. Figure 3B shows split-half reliabilities of estimates between odd and even images as a function of trimmed trial times. Across all tested dimensions, estimates based on the first 2 seconds of trials are highly consistent with those for the full trial duration (all $r > 0.91$, $p < 0.001$) and yield split-half reliabilities approaching or exceeding those seen for 3 seconds (ranging from $r = 0.66$, $p < 0.001$ for *Motion* to $r = 0.95$, $p < 0.001$ for *Faces*).

Combining small stimulus sets with shorter trial durations

Next, we considered the effects of trimmed trial durations on smaller stimulus sets. Figure 3C and Figures 4A and 4B show the number of images necessary to estimate reliable individual differences ($r \geq 0.7$) in cumulative dwell time as a function of (trimmed) trial duration. *Faces* and *Text* reached this criterion with fewer than 40 images for all trial durations. *Motion* could reliably be estimated with fewer than 100 images when using at least the first second of trial data. Interestingly, *Touched* and *Taste* crossed the threshold with fewer than 100 images for the first second of trial duration (Figure 4A) but required more than 100 (*Touched*) or even 200 (*Taste*) images for trial durations of 2 seconds and 3 seconds. Figure 4C shows the consistency of individual visual exploration (numbers of objects fixated) with day 1 as a function of trial duration and considered set size from day 2. At 40 images, the data are above the consistency threshold for both 1-second and 2-second trial durations. Supplementary Table S1 provides an overview of results and compares consistency correlations to the proportion of explainable variance (as indicated by test–retest reliabilities for the full set reported by de Haas et al., 2019).

Prevalence of image features

In a final step we determined the prevalence of image features for the image sets $OSIE_{40}$, $OSIE_{100}$, $OSIE_{200}$, and $OSIE_{700}$. Figure 5 shows the (log10) absolute frequency (Figure 5A) and the relative frequency (Figure 5B) of these features across stimulus sets. Results show that the relative distribution of features was stable across image sets (Figure 5B). The proportion of *Faces* was most similarly distributed across image sets and ranged from 47.31 ($OSIE_{40}$) to 48.74 ($OSIE_{700}$). The proportion of *Text* was least stable and ranged from 18.88 ($OSIE_{700}$) to 22.1 ($OSIE_{40}$). However, the absolute number of all feature occurrences (unsurprisingly) was heavily reduced for the shorter sets (Figure 5A).

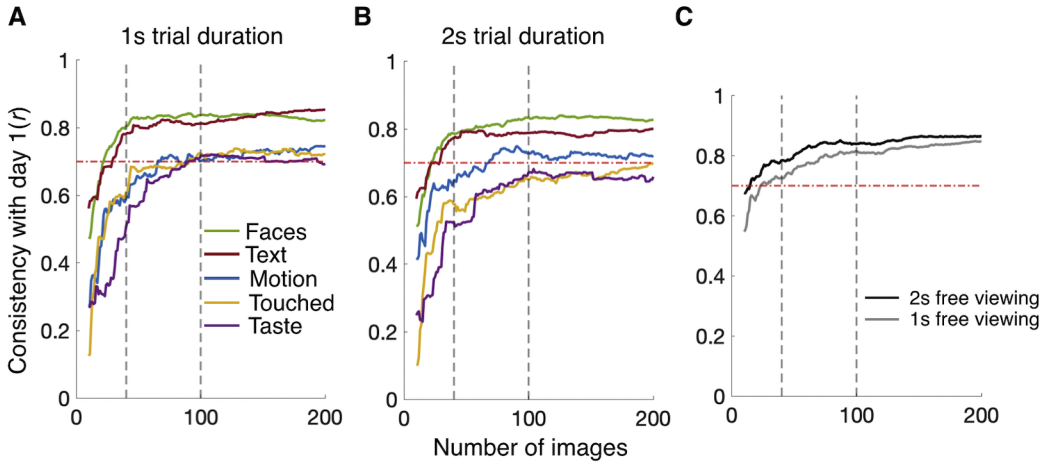

Figure 4. Consistency of estimates based on smaller stimulus sets for shorter trial durations. Line plots show the Pearson consistency of estimates for the full stimulus set (day 1) with smaller subsets from day 2. (A, B) Consistency correlations for estimated proportions of cumulative viewing times for trial durations trimmed to 1 second (A) and 2 seconds (B); semantic dimensions are indicated by color as shown in the inset. (C) Consistency correlations for the individual tendency for visual exploration (number of distinct objects fixated) for 1 second and 2 seconds as indicated by the inset. Dashed gray lines mark stimulus sets of 40 and 100 images. Dashed red lines indicate the desired consistency threshold of $r = 0.70$.

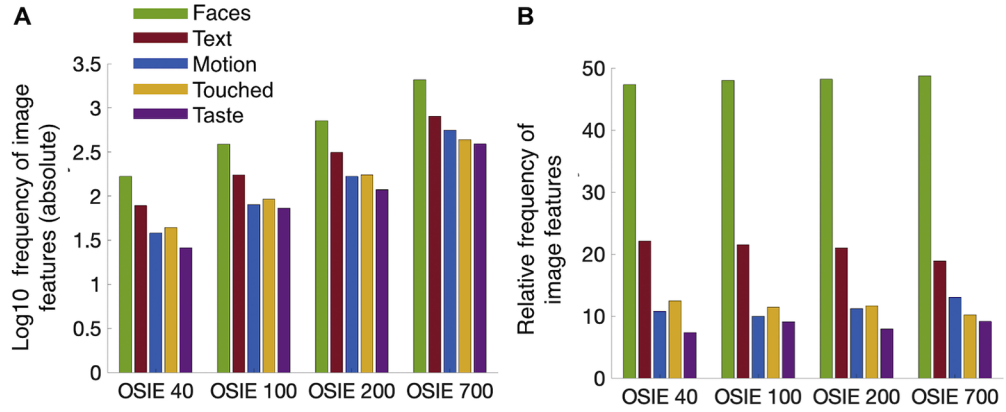

Figure 5. Prevalence of image features along all semantic dimensions and across image sets. (A) Bar plot describing the absolute frequency (log10 scale) of image features for a given semantic dimension and across image sets. (B) Bar plot displaying the relative frequency in percent for image features of a given dimension and image set. Semantic dimensions are indicated by color as shown in the inset.

Discussion

The results of the present study allow two conclusions. First, they corroborate the finding of reliable individual differences in gaze behavior reported by de Haas et al. (2019). Individual differences in visual exploration and systematic fixation biases toward a set of semantic features were replicated for both dwell time and first fixations. Second, the present results show how to test these differences efficiently.

Based on our results, we propose the $OSIE_{40}$ subset with a presentation time of 2 seconds per image as an attractive option for estimating individual gaze behavior efficiently. When using the re–test reliabilities for the full image set reported by de Haas et al. (2019) as an estimate of the noise ceiling, this test captures more than 80% of the explainable variance in visual exploration and dwell time for three out of five dimensions (*Face*, *Text*, and *Motion*) and 69% of explainable dwell-time variance for *Touched*. The remaining fifth dimension of *Taste* was not well captured by any of the short tests, which perhaps is unsurprising given its relatively low internal consistency, as reported by de Haas et al. (2019) and replicated here.

Researchers interested in individual differences in first fixations can achieve acceptable results using the $OSIE_{100}$ but can expect considerable gains from using the full stimulus set. All subsets of stimuli can be retrieved from Supplementary Table S2 and downloaded with the online data and code at https://osf.io/ekvj4/.

Exploratory analyses showed that the relative prevalence of image features was similarly distributed across dimensions for the different image subsets (Figure 5B). Nevertheless, it is possible that the lower absolute number of features in the image subsets caused floor effects for consistency estimates of less salient dimensions (i.e., *Touched* and *Taste*) (de Haas et al., 2019; Xu et al., 2014), which also appear less frequently in the images. The ability to estimate individual differences in fixation tendencies along a given dimension depends on sufficient data. If the number of relevant fixations in the dataset is too low, it will lack the necessary granularity to capture smaller but relevant inter-individual differences. It seems likely that the relatively low saliency for *Taste* and *Touched* (de Haas et al., 2019; Xu et al., 2014) in the smaller image sets is compounded by a low absolute number of occurrences of these features, yielding such floor effects.

Individual differences in gaze behavior are gaining interest in vision and computer science (Li & Chen, 2018; Yu & Clark, 2017), as well as clinical psychology (Armstrong & Olatunji, 2012). Especially in developmental studies and clinical settings, free-viewing experiments using natural images are ideal because of their minimum task requirements. Fixation measurements also come with minimal technical requirements and could soon be bedside compatible. At the same time, applied settings and clinical research often come with severe time limits and prohibit the use of extensive testing batteries. We hope our current results contribute to closing this gap and will provide a useful standard for researchers aiming to explore individual fixation tendencies in a time-efficient manner.

The explorative analyses of truncating trial durations suggested that individual differences can reliably be estimated based on trials of just 1 or 2 seconds rather than 3 seconds. Notably, we observed slightly higher validity correlations for 1 second versus 2 and 3 seconds for the dimensions *Touched* and *Taste* (Figures 3A, 3B). Although these consistency differences between truncated datasets were small, they were constant across subsets. A possible reason for this is that individual fixations tendencies toward *Touched* objects or objects with *Taste* are only stable insofar as they are guided by bottom–up processes and have a higher degree of intra-subject variability as soon as later, top–down processes come into play. For example, earlier fixations toward objects with *Taste* may be driven by relatively stable individual levels of saliency for low-level features, such as the high color contrast of these objects. Later, top–down-directed fixations may in turn more heavily depend on the variable states of the observer, such as hunger or satiety.

Further, these analyses were based on trimming the viewing data rather than the actual viewing duration during the experiment. We cannot exclude the possibility that shortening trial times during testing would have different and maybe unexpected consequences due to the altered pace of the experiment. Pilot data from our lab suggest that experiments with a 2-second viewing duration yield individual differences with excellent internal consistency; however, this comes with an intercept shift across the given semantic dimensions. Future research could aim for an in-depth comparison of 2-second and 3-second trial durations in the same sample.

In the present study fixation differences were assessed with a free-viewing task. Past research has demonstrated that low-level properties of gaze behavior and the low-level features of fixated regions vary between tasks and can also interact with individual differences (Kardan, Berman, Yourganov, Schmidt, & Henderson, 2015; Kardan, Henderson, Yourganov, & Bergmann, 2016). However, these studies did not consider individual differences in semantic saliency. Future research should test whether saliency along these dimensions is modulated by task and in particular whether such modulations are limited to intercept shifts on the group level or interact with individual differences.

In sum, we have replicated consistent individual differences in gaze behavior and documented that they can be estimated using less than 2 minutes of free-viewing data.

Keywords: eye movements, individual differences, free-viewing, saliency

Acknowledgments

The authors thank Xu et al. (2014) for making their stimuli and annotations freely available. Further, we thank Stefanie Mueller and the PsychLab team at ZPID for their help with data collection.

Supported by European Research Council Starting Grant 852885 INDIVISUAL; BdH was further supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 222641018–SFB/TRR 135 TP A8.

Commercial relationships: none.
Corresponding author: Marcel Linka.
Email: marcellinka54@gmail.com.

Address: Experimental Psychology, Justus Liebig Universität, Giessen, Germany.

References

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review, 32*, 704–723.

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research, 141*, 157–169.

Benson, P. J., Beedie, S. A., Shephard, E., Giegling, I., Rujescu, D., & St. Clair, D. (2012). Simple viewing tests can detect eye movement abnormalities that distinguish schizophrenia cases from controls with exceptional accuracy. *Biological Psychiatry, 72*, 716–724.

Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology, 62*, 1–14.

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 is atypical in autism. *Nature, 547*, 340–344.

de Haas, B. (2019). *Can a small stimulus set reliably estimate individual differences in semantic salience?* Trier, Germany: Leibniz Institut für Psychologische Information und Dokumentation (ZPID).

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences, USA, 116*, 11687.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences, 10*, 382–390.

Gegenfurtner, K. R. (2016). The interaction between vision and eye movements. *Perception, 45*, 1333–1357.

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Proceedings of the 19th International Conference on Neural Information Processing Systems* (pp. 545–552) Cambridge, MA: MIT Press.

Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1390–1400.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature, 504*, 427–431.

Kardan, O., Berman, M. G., Yourganov, G., Schmidt, J., & Henderson, J. M. (2015). Classifying mental states from eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 1502–1514.

Kardan, O., Henderson, J.M., Yourganov, G., & Berman, M.G. (2016). Observers' cognitive states modulate how visual inputs relate to gaze control. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1429–1442.

Kennedy, D.P., D'Onofrio, B.M., Quinn, P.D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic influence on eye movements to complex scenes at short timescales. *Current Biology, 27*, 3554–3560.e3.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception, 36*, 1–16.

Li, A., & Chen, Z. (2018). Personalized visual saliency: Individuality affects image perception. *IEEE Access, 6*, 16099–16109.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., . . . Zhao, Q. (2015). Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron, 88*, 604–616.

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision, 14*(1):28, 1–20, https://doi.org/10.1167/14.1.28.

Yu, B., & Clark, J. J. (2017). Personalization of saliency estimation. Retrieved from https://arxiv.org/pdf/1711.08000.pdf.

Supplementary materials

Table S1

Consistency correlations for small stimulus sets and truncated trial durations. Rows show consistency correlations for each semantic dimension and visual exploration for the $OSIE_{40}$, $_{100}$ and $_{200}$ sets. Columns show Pearson consistency correlations (r) between the estimated proportion of individual dwell time (or first fixations rightmost columns) on the respective objects for the full set seen on day 1 and the respective subset on day 2. Dwell time estimates from day 2 were either based on the full trial duration (3s) or trials truncated to 1 or 2s, as indicated by the column headings. % expl. indicates which proportion of explainable variance a given consistency correlation corresponds to. This was estimated based on the test-retest reliabilities for the full stimulus set (700 images) reported by de Haas et al. (2019), which served as an independent estimate of the noise-ceiling.

	Cumulative fixation time						First fixations	
	Trial dur. (1s)		**Trial dur. (2s)**		Trial dur. (3s)			
	r	% exp.	***r***	**% exp.**	*r*	% exp.	*r*	% exp.
$OSIE_{40}$								
Faces	.81	103	**.79**	**98**	.74	86	.67	61
Text	.78	86	**.77**	**84**	.72	73	.60	46
Motion	.59	75	**.65**	**91**	.65	91	.51	43
Touched	.61	76	**.58**	**69**	.59	71	.58	55
Taste	.51	43	**.53**	**46**	.51	43	.38	32
Vis. exploration	.73	83	**.79**	**98**	.78	95	-	-
$OSIE_{100}$								
Faces	.84	110	.83	108	.81	103	**.80**	**86**
Text	.81	93	.79	88	.72	73	**.72**	**65**
Motion	.70	106	.73	115	.71	109	**.54**	**48**
Touched	.72	106	.65	86	.63	81	**.65**	**69**
Taste	.72	85	.67	74	.65	69	**.57**	**72**
Vis. exploration	.82	105	.84	110	.82	105	-	-
$OSIE_{200}$								
Faces	.82	105	.83	108	.82	105	.83	93

Text	.85	102	.80	91	.75	80	.81	79
Motion	.75	122	.72	112	.73	115	.65	69
Touched	.72	106	.70	100	.69	97	.67	74
Taste	.69	78	.66	72	.67	74	.63	88
Vis. exploration	.85	113	.86	116	.85	113	-	-

Table S2

Images included in the three tested stimulus subsets $OSIE_{40}$, $OSIE_{100}$ and $OSIE_{200}$

Image number		
$OSIE_{40}$	$OSIE_{100}$	$OSIE_{200}$
1608	1608	1608
1226	1226	1226
1233	1233	1233
1335	1335	1335
1472	1472	1472
1391	1391	1391
1479	1479	1479
1277	1277	1277
1590	1590	1590
1118	1118	1118
1296	1296	1296
1094	1094	1094
1178	1178	1178
1460	1460	1460
1101	1101	1101
1620	1620	1620
1576	1576	1576
1287	1287	1287

1265	1265	1265
1566	1566	1566
1573	1573	1573
1420	1420	1420
1188	1188	1188
1227	1227	1227
1433	1433	1433
1112	1112	1112
1614	1614	1614
1001	1001	1001
1606	1606	1606
1643	1643	1643
1567	1567	1567
1522	1522	1522
1526	1526	1526
1480	1480	1480
1543	1543	1543
1544	1544	1544
1491	1491	1491
1006	1006	1006
1132	1132	1132
1481	1481	1481
	1247	1247
	1586	1586
	1187	1187
	1550	1550
	1441	1441
	1395	1395

1468	1468
1254	1254
1182	1182
1579	1579
1246	1246
1009	1009
1286	1286
1527	1527
1035	1035
1138	1138
1148	1148
1425	1425
1096	1096
1694	1694
1637	1637
1268	1268
1663	1663
1456	1456
1162	1162
1301	1301
1349	1349
1621	1621
1628	1628
1011	1011
1416	1416
1022	1022
1484	1484
1128	1128

1474	1474
1229	1229
1580	1580
1023	1023
1353	1353
1565	1565
1308	1308
1103	1103
1169	1169
1015	1015
1578	1578
1539	1539
1452	1452
1236	1236
1018	1018
1531	1531
1583	1583
1327	1327
1209	1209
1197	1197
1248	1248
1044	1044
1385	1385
1602	1602
1465	1465
1352	1352
	1208
	1570

1386

1431

1325

1204

1547

1382

1445

1106

1568

1309

1406

1670

1288

1478

1366

1077

1683

1363

1303

1401

1657

1358

1508

1430

1540

1377

1372

1210

1111

1113

1562

1455

1059

1279

1130

1109

1512

1646

1313

1356

1273

1461

1556

1002

1675

1075

1680

1381

1088

1376

1393

1219

1462

1206

1292

1257

1284

1241

1412

1440

1156

1196

1518

1215

1624

1592

1351

1302

1281

1150

1609

1650

1483

1528

1551

1546

1146

1244

1216

1272

1371

1079

1147

1258

1548

1131

1183

1238

1700

1192

1555

1201

1501

1593

1502

1119

1340

1610

Study 3

Linka, M., Sensoy, Ö., Karimpur, H., Schwarzer, G., & de Haas, B. (2023). Free viewing biases for complex scenes in preschoolers and adults. *Scientific Reports*, 13(1), 11803.

scientific reports

Check for updates

OPEN

Free viewing biases for complex scenes in preschoolers and adults

Marcel Linka[1]✉, Özlem Sensoy[2], Harun Karimpur[1], Gudrun Schwarzer[2] & Benjamin de Haas[1]

Adult gaze behaviour towards naturalistic scenes is highly biased towards semantic object classes. Little is known about the ontological development of these biases, nor about group-level differences in gaze behaviour between adults and preschoolers. Here, we let preschoolers ($n = 34$, age 5 years) and adults ($n = 42$, age 18–59 years) freely view 40 complex scenes containing objects with different semantic attributes to compare their fixation behaviour. Results show that preschool children allocate a significantly smaller proportion of dwell time and first fixations on *Text* and instead fixate *Faces*, *Touched* objects, *Hands* and *Bodies* more. A predictive model of object fixations controlling for a range of potential confounds suggests that most of these differences can be explained by drastically reduced text salience in pre-schoolers and that this effect is independent of low-level salience. These findings are in line with a developmental attentional antagonism between text and body parts (touched objects and hands in particular), which resonates with recent findings regarding 'cortical recycling'. We discuss this and other potential mechanisms driving salience differences between children and adults.

To process visual information at high resolution, we constantly have to move our eyes. The periphery of the visual field suffers from crowding and lack of acuity, implying the need to prioritise certain areas in a scene over others[1,2]. How does the oculomotor system decide these priorities? Does it learn them over time? And if so, how do they change across the lifespan?

Adult gaze behavior. The past 25 years have seen extensive attempts to predict adult gaze behaviour based on the content of a visual scene. Two types of scene features have emerged as particularly relevant, namely low-level features such as the local contrast in orientation, intensity and color[3,4] and high-level objects such as faces and text[5]. Objects[6] and their semantic features[5] have been found to be substantially more predictive of gaze behaviour than classic low-level salience. For instance, objects[6,7] and in particular those that are faces, text, touched objects and objects with implied motion[5,8] hold large weights in predicting adult gaze behaviour, outweighing low-and mid-level features in complex scenes[5,9]. These attentional biases towards certain semantic features are shared between individuals, but their degree differs substantially and reliably[10,11]. Moreover, some gaze biases seem to be particularly pronounced among observers with specific perceptual abilities or challenges. For instance, patients with autism spectrum disorder (ASD) spend less time fixating faces and social features compared to controls[12,13]. Super-recognizers—people with exceptional abilities for processing facial identity information—spend more time fixating faces and less time focusing on touched objects and text[14].

The development of gaze behavior. While visual preferences for semantic features have been studied extensively for adults, much less is known about their development. A notable exception are faces. Infants show a visual preference for face-like dot patterns over inverted patterns very early on[15,16]. In fact, recent evidence points to increased behavioural responses to such face-like stimuli by human foetuses even during the third trimester of pregnancy[17] (but note that this conclusion has been questioned on methodological grounds[18,19]). These findings contribute to a long-standing debate on the question whether human attentional biases towards faces are driven by innate mechanisms[16,20]. In the course of infancy, this attentional bias develops into a more differentiated preference for faces over mere face-like stimuli[21] and non-face objects in complex scenes[22–24]. Aside from faces, infants show gaze biases towards other socially relevant stimuli. For example, Frank et al. let children from 3- to 30-months of age freely view videos of complex social scenes depicting children performing everyday actions (mostly playing with toys). They found that infants spent a larger proportion of their dwell time fixating hands when a depicted scene turned socially complex, that is, when agents in the videos used their hands for more complex actions[25]. This effect was positively correlated with age, suggesting that throughout infancy and early childhood, gaze allocation becomes increasingly tuned to socially relevant visual information.

[1]Department of Experimental Psychology, Justus Liebig University Giessen, 35394 Giessen, Germany. [2]Department of Developmental Psychology, Justus Liebig University Giessen, 35394 Giessen, Germany. ✉email: marcellinka54@gmail.com

This bias for social information might go hand in hand with an early visual sensitivity for biological motion in (new-born) infants[26]. More recent work shows that a visual preference for implied motion in static stimuli might emerge already at around 5 months of age[27,28]. As with adults[10,29], the degree of attentional biases towards social information varies reliably in childhood and is highly heritable[12,30]. Visual attention in infants is however not only dependent on the object's semantics or the individuals' gaze biases. Work using head camera recordings of infants and their parents playing in naturalistic scenes has shown that prior parental attention towards objects results in extended visual selection for a given object[31]. Interestingly, infant attention and even more so, joint attention towards objects during word learning predicted later vocabulary size in infants[32]. Fewer studies have investigated the development of attentional biases beyond three years of age.

Açık et al. have looked at free-viewing behaviour towards complex scenes across three age groups (7–9-year-old children, 19–27-year-old adults, and > 72-year-olds) and showed that the effect of low-level saliency on gaze drops with age[33]. Further, Helo et al. compared free-viewing behaviour of adults with those of children between 2 and 10 years and found that saccadic amplitudes increased and fixation duration decreased as a function of age[34]. Krishna et al. found that the center bias in image viewing decreases with age[35]. These studies suggest that gaze behaviour towards complex scenes changes from preschool/school age to adulthood, shifting from fixating more local, low-level features like luminance contrasts and orientation, to more global, explorative viewing behavior. However, more recent work predicting gaze behaviour in children (age 6–14) and adults showed that a face-based model outperformed (low-level) saliency models in predicting gaze towards complex scenes even in the youngest age group. The face-based model however was more predictive for gaze behaviour of adults than children[34]. This suggests that gaze behavior is more attracted to faces than low-level features and that this face preference increases with age.

Taken together, we know of distinct semantic information that guides viewing behaviour in adults and infants. Gaze behaviour in older children seems to differ significantly from adults in terms of enhanced low-level salience, center-bias, fixation durations and lower saccadic amplitudes[33,34]. However, just as for adults, the gaze behaviour of children is better predicted by high-level semantic information compared to low-level features[5,36]. Although we know of qualitative differences in the effect of semantic information on fixation behavior in adults and infants, it is unclear whether and how gaze behaviour towards semantic information differs between preschool children and adults.

Besides *Faces* and *Touched* objects, *Bodies* and *Hands* adults also show high (but individually variable) salience for *Text*[5,10]. To date, it is unclear how these biases for *Text* develop, whether they exist before reading acquisition and how they interact with other semantic gaze biases. Here, we probe this question by comparing free viewing biases for *Text*, *Faces*, *Touched* objects, as well as *Bodies* and *Hands* between preschool children and adults free-viewing complex scenes.

Understanding whether and how gaze behaviour in children is drawn towards semantic content can have important implications. Atypical gaze behaviour towards faces has been found in infants and children with ASD[37,38]. Insights on semantic gaze biases in healthy children may help establishing a baseline to test the diagnostic potential of gaze behaviour for neurodevelopmental disorders. Moreover, given the increasing exposure of children to screen-based education, understanding their attentional biases may help in the design of efficient learning material. Investigating visual attention in preschoolers is particularly relevant, because this age represents a time of remarkable psychological growth; For example, selective attention develops significantly during later childhood, especially between 4 and 7 years of age[39]. Further, during the preschool years, children have an innate 'theory drive', that is they ask many questions and actively seek causal explanations in order to interpret and understand things in their environment[40,41]. This and a highly dynamic neural development (also in the visual cortices[42,43]) make preschoolers a particularly relevant age group for studying visual attention towards complex scenes as insights may inform theories on neural and cognitive functioning.

The present study. In the present study we investigated gaze behaviour towards 40 complex scenes in 5-year-old preschoolers and adults. In particular, we tested whether and to what extent children compared to adults differ in their proportions of (1) cumulative dwell time and (2) first fixations towards objects of multiple semantic categories depicted in these scenes, namely *Faces*, *Text* and *Touched* objects. First fixations have been described as driven by automatic or bottom-up processes[5,44–46]. Differences in first fixation allocation may thus reflect observer traits which are less malleable than those reflected in total dwell time.

Here, proportion of cumulative dwell time refers to the summed duration of all fixations falling on a given feature, divided by the summed duration of fixations falling on any object. First fixation proportions refer to the proportion of first fixations landing on any object, which landed on objects of a given category. First fixations refer to the first fixation initiated at least 100 ms after image onset (i.e. following the first saccade after image onset). Note that these proportions do not necessarily add up to 1. Any object-directed fixation can be labelled with none or multiple of the features analysed here. For example, a plate of pasta may bear none of the features analysed here, while text can be touched, too.

We chose these dimensions because of their importance for predicting adult gaze behaviour. Moreover, we have previously found that individual gaze tendencies along these dimensions can be tested reliably with this small stimulus set[5,10,11]. The recent publication of additional pixel masks and meta data for this stimulus set[47] further allows the quantification of fixation biases towards *Bodies* and *Hands* depicted in the scenes. Adult gaze behavior is also predicted by objects with implied *Motion*. However, the moving objects in our stimulus set were almost exclusively person-related features (*Bodies*, *Hands* or *Faces*), rendering it ill-suited to estimate an independent contribution of this dimension. We did not consider *Motion* for our main analyses, but include a corresponding analysis in the Supplementary Material.

Results

Simple descriptive and inferential statistics. First, we investigated whether children vs. adults show differences in gaze behaviour towards semantic information on a simple descriptive and inferential level. We analysed group differences in the proportion of cumulative dwell time and first fixations on *Text, Faces* and *Touched* objects. Figure 1 shows example heatmaps of cumulative dwell time for both groups as well as overlaid pixel masks. Figure 2 shows the distributions of cumulative dwell time proportions for children and adults respectively for each of the three semantic dimensions. See Table 1 for the mean pixel coverage of each semantic category.

Dwell time differences towards semantic categories. Results showed that the proportion of cumulative dwell time children spent on *Text* was reduced fivefold compared to adults, $t\text{obs} = -14.59$, $p < 0.001$. Further, children spent a significantly larger proportion of their dwell time on *Faces*, $t\text{obs} = 2.02$, $p = 0.02$. and *Touched* objects, $t\text{obs} = 4.28$, $p < 0.001$ than adults.

First fixations towards semantic categories. In a further descriptive analysis, we tested whether children and adults differ in their proportion of *first* fixations towards *Text*, *Faces* and *Touched* objects, that is the proportion of immediate saccades after image onset landing on objects from these categories. Figure 3 shows the distributions of first fixation proportions for children and adults. Findings show that children directed a significantly smaller proportion of first fixations towards *Text,* $t\text{obs} = -5.76$, $p < 0.001$ and *Faces*, $t\text{obs} = -3.02$, $p = 0.002$ and a significantly larger proportion of first fixations towards *Touched* objects, $t\text{obs} = 4.28$, $p < 0.001$.

Gaze behaviour towards hands and bodies. Given enhanced fixation tendencies towards *Faces* and *Touched* objects in children and recent findings on fixation tendencies in adults[47], we additionally tested whether similar effects would emerge for *Hands* and *Bodies* (with the latter excluding hands). Figure 4 shows the distributions of cumulative dwell time (a, b) and first fixation (c, d) proportions spent on *Hands* and *Bodies* for children and adults. Children did not significantly differ from adults in the proportion of dwell time spent on *Bodies* (although there was a trend for children to fixate more towards *Bodies*, $t\text{obs} = 1.5$, $p = 0.07$), but spent a significantly larger proportion of dwell time on *Hands* ($t\text{obs} = 2.78$, $p = 0.01$). Moreover, children directed a larger proportion of first saccades towards *Hands* and *Bodies* compared to adults (*Hands*: $t\text{obs} = 2.83$, $p = 0.007$, *Bodies*: $t\text{obs} = 2.58$, $p = 0.007$).

Note, for these analyses, we focussed on group differences in object-directed dwell time and first fixations. Although children and adults both spent the vast majority of their dwell time and first fixations on objects (rather than the background), we found that children did so significantly less (dwell time on objects for children

Figure 1. Pixel masks and heat maps showing group-wise fixations for two example images. Example stimuli on the left-hand side depict overlaid pixel masks for objects of the semantic dimensions: *Text* (green), *Faces* (red), *Touched* (violet), *Hands* (blue) and *Bodies* (cyan). Heat maps in the middle show corresponding duration weighted fixation data from children, maps on the right-hand side show fixations from adults. The transparent white scatters show all first fixations towards the image by all subjects of the respective group. Please note the increased gaze towards touched objects/hands and moving objects in the heatmaps for children, as well as the lower levels of text fixations compared to those of adults. Heatmaps and pixel mask overlays were created using MATLAB R2019B (MathWorks).

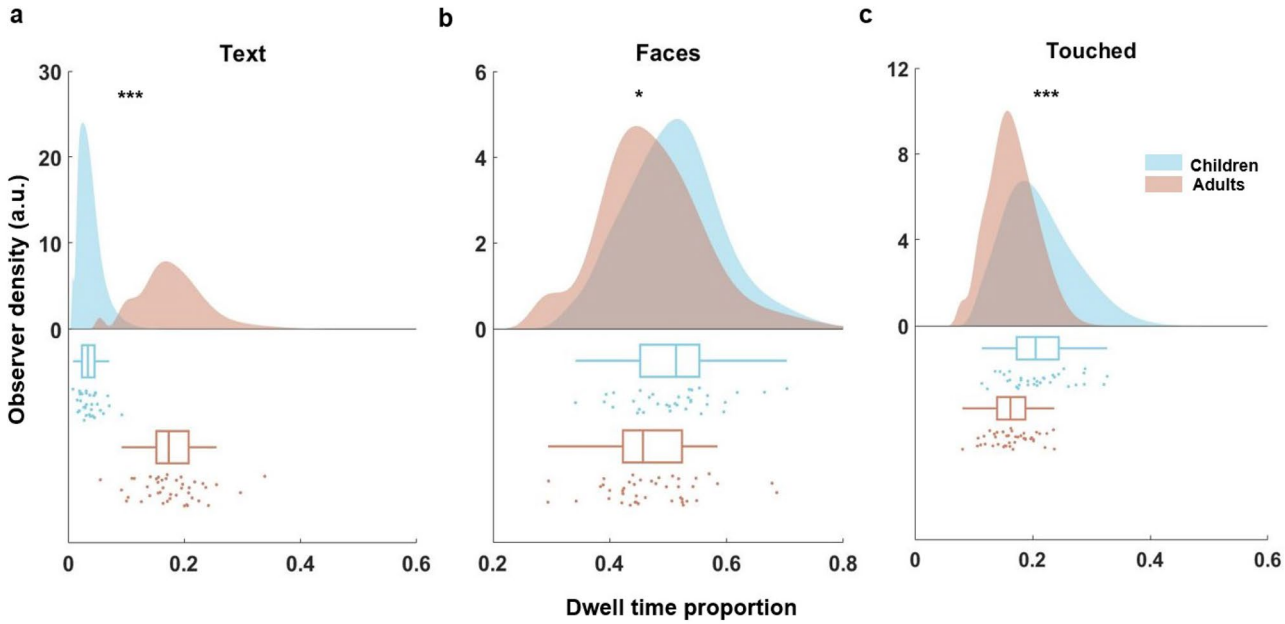

Figure 2. Differences in gaze behaviour along three semantic dimensions. Density plots showing the probability distribution of cumulative dwell time proportion towards objects of the dimensions *Face* (**a**), *Text* (**b**), and *Touched* (**c**) respectively for children (blue) and adults (red). Data points below the distributions indicate the individual dwell time proportion. Box plots depict the summary statistics for each dimension and group. For a respective box plot, the vertical line indicates the mean cumulative dwell time proportion. The left side of the box indicates the 25th percentile and the right side the 75th percentile. The whiskers represent the minimum and maximum values. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ (Holm-Bonferroni corrected; see Methods).

Pixel masks	Text	Faces	Touched	Bodies	Hands	Total
Coverage (mean %)	2.83	4.94	5.21	14.51	1.53	27.36
SD (across images)	2.3	5.09	5.45	13.8	1.8	17.11

Table 1. Mean % coverage of pixel masks per included semantic category Mean % of coverage represents the mean % of pixels covered in an image by objects of a given semantic dimension across all images depicting objects of a given dimension. SD gives the standard deviation of relative pixel coverage per semantic dimension across images depicting objects of a given dimension. Total includes the mean % and standard deviation (SD) of pixel masks of all 5 analysed dimensions across images.

and adults: $M = 88\%$ and 90%, respectively; $tobs = -1.98$, $p = 0.02$; first fixations $M = 86\%$ and 92%, $tobs = -5.47$, $p < 0.001$). Nevertheless, including background fixations in our analyses did not change the pattern of results. Results of the corresponding control analyses are reported in the supplementary materials.

(G)LMM analyses. Adults may spend less time fixating *Faces*, *Touched* objects and *Hands* because of their higher propensity to fixate *Text*, or an inherently lower tendency to fixate these other dimensions. To juxtapose these hypotheses and control for explicit lower-level salience features[48,49], we fitted a Linear Mixed Effect Model (LMM) using fitlme() from the Statistics and Machine Learning Toolbox in MATLAB R2019b. This model predicted the cumulative dwell time towards an object given its high-level semantic features as well as size, image eccentricity, and graph-based visual salience (GBVS)[3], for which previous work has shown age-dependent differences[33,35,48].

Moreover, we fitted a binomial Generalized Linear Mixed Model (GLMM) using the function fitglme() from the Statistics and Machine Learning Toolbox in MATLAB R2019b to estimate the probability that a given object was fixated immediately after image onset (first fixation), again, as a function of high-level semantic features, size, image eccentricity, and graph-based visual salience (GBVS)[3]. To show the influence of the group difference in Text fixations on gaze differences towards any other semantic category, we have additionally fitted trimmed (G)LMMs, excluding Text (see supplementary materials for results).

Model coefficients, standard errors, and corresponding *t*-values and *p*-values for results are reported in Tables 2 and 3 respectively. Figure 5 shows the estimates for each included interaction for the (a) linear and (b) binomial model.

The LMM's intercept holds no specific significance in this context, as it indicates the estimated average value (log-transformed and standardized) of object-directed dwell time when all predictors, including semantic

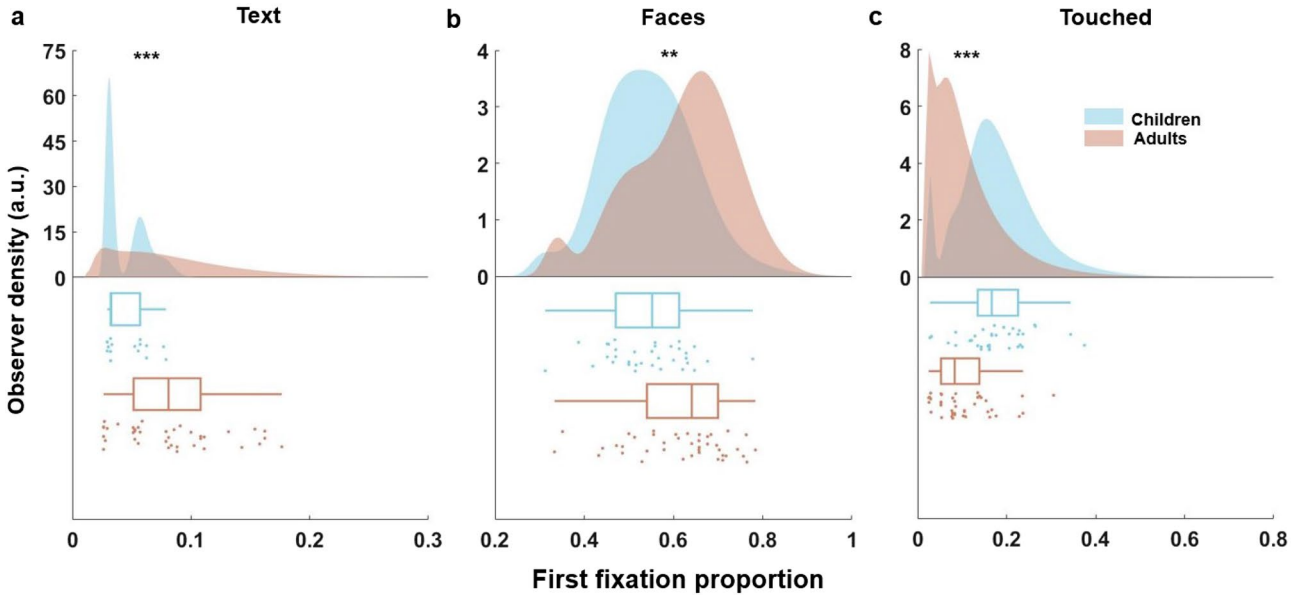

Figure 3. Differences between children and adults in first fixation proportion for three semantic dimensions. Density plots show the probability distribution of first fixation proportion towards *Faces* (**a**), *Text* (**b**) and *Touched* objects (**c**) respectively for children (blue) and adults (red). Box plots below show an overview of the summary statistics for each group and semantic dimension; the vertical line within a respective box represents the mean value. The left side of a box indicates the 25th percentile and the right side the 75th percentile. The whiskers represent the minimum and maximum values. Data points represent the individual corresponding first fixation proportions. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ (Holm-Bonferroni corrected; see Methods).

variables are equal to zero. When referring to z-standardized continuous predictors, the value of zero corresponds to the average value of the predictor. The significant effect for *Age Group*, ($b = 0.21$, $t = 4.17$, $p < 0.001$), indicates that children spent more dwell time towards a selected object than adults. Note, this does not contradict the earlier finding that children spent less time fixating objects overall compared to adults (see simple descriptive and inferential statistics), because here we look at the average dwell time for a given object (children tend to look at fewer objects overall and therefore spent more time engaging with them). We also found a significant interaction between *Age Group* and *Text*, $b = -0.84$, $t = -10.16$, $p < 0.001$, indicating longer dwell time towards *Text* for adults compared to children. Further, we found a small but significant interaction between *Age Group* and *Faces*, $b = -0.12$, $t = -2.05$, $p = 0.04$, also indicating somewhat higher dwell time on faces for adults compared to children. No significant interactions were found between *Age Group* and *Touched*, $b = 0.04$, $t = 0.71$, $p = 0.48$, *Age Group* and *Hands*, $b = -0.06$, $t = -0.87$. $p = 0.39$, and *Age Group* and *Bodies*, $b = -0.08$, $t = -1.31$, $p = 0.19$, indicating no significant dwell time differences between children and adults for objects of these categories when controlling for the other dimensions.

Regarding lower level features, we found a significant interaction between *Age Group* and *Eccentricity*, $b = -0.11$, $t = -6.33$, $p < 0.001$, indicating a stronger center-bias for children vs. adults. Finally, we found a significant object *Size* x *Age Group* interaction, $b = -0.07$, $t = -4.10$, $p < 0.001$, indicating a smaller effect of object *Size* on the dwell time of children compared to adults. No significant interaction was found between *Age Group* and *GBVS*, $b = 0.02$, $t = 1.04$, $p = 0.30$.

We also estimated the probability that a given object attracted the first fixation after image onset. Again, the intercept bears no relevant information as it represents the probability of first fixating on an object that lacks any of the incorporated semantic labels, while exhibiting average values across all other continuous variables. The non-significant effect for *Age Group* ($b = -0.08$, $t = -0.60$, $p < 0.55$) indicates a similar probability to first fixate an object across age groups. The probability that *Text* attracted the first fixation was smaller for children vs. adults, $b = -1.35$, $t = -4.83$, $p < 0.001$ and the same was true for *Faces*, $b = -0.46$, $t = -2.91$, $p = 0.004$. Further, a significant interaction was found between *Age Group* and *Touched* objects, $b = 0.65$, $t = 2.99$, $p = 0.003$, indicating a larger probability for *Touched* objects to attract the first fixation in children compared to adults. No significant interactions were found between *Age Group* and *Hands*, $b = 0.16$, $t = 0.75$, $p = 0.45$, and *Age Group* and *Bodies*, $b = 0.31$ $t = 1.84$, $p = 0.07$. Regarding lower level features, larger objects had a higher probability to attract first fixations in adults compared to children (*Age Group* x *Size* interaction; $b = -0.17$, $t = -2.48$, $p = 0.01$, and children showed a stronger center-bias for first fixations (negative effect of object eccentricity) $b = -0.18$, $t = -2.68$, $p = 0.01$. Finally, no group difference was shown for object-based low-level salience (*GBVS*), $b = 0.04$, $t = 0.60$, $p = 0.55$.

Discussion

Adult gaze behaviour towards complex scenes is strongly biased towards semantic object categories[5] and individual observers show large and reliable differences in the magnitude of these biases[10]. Many of these biases appear early on: Infants already show looking preferences for faces[15], hands[25] and (implied) motion[26–28]. At the

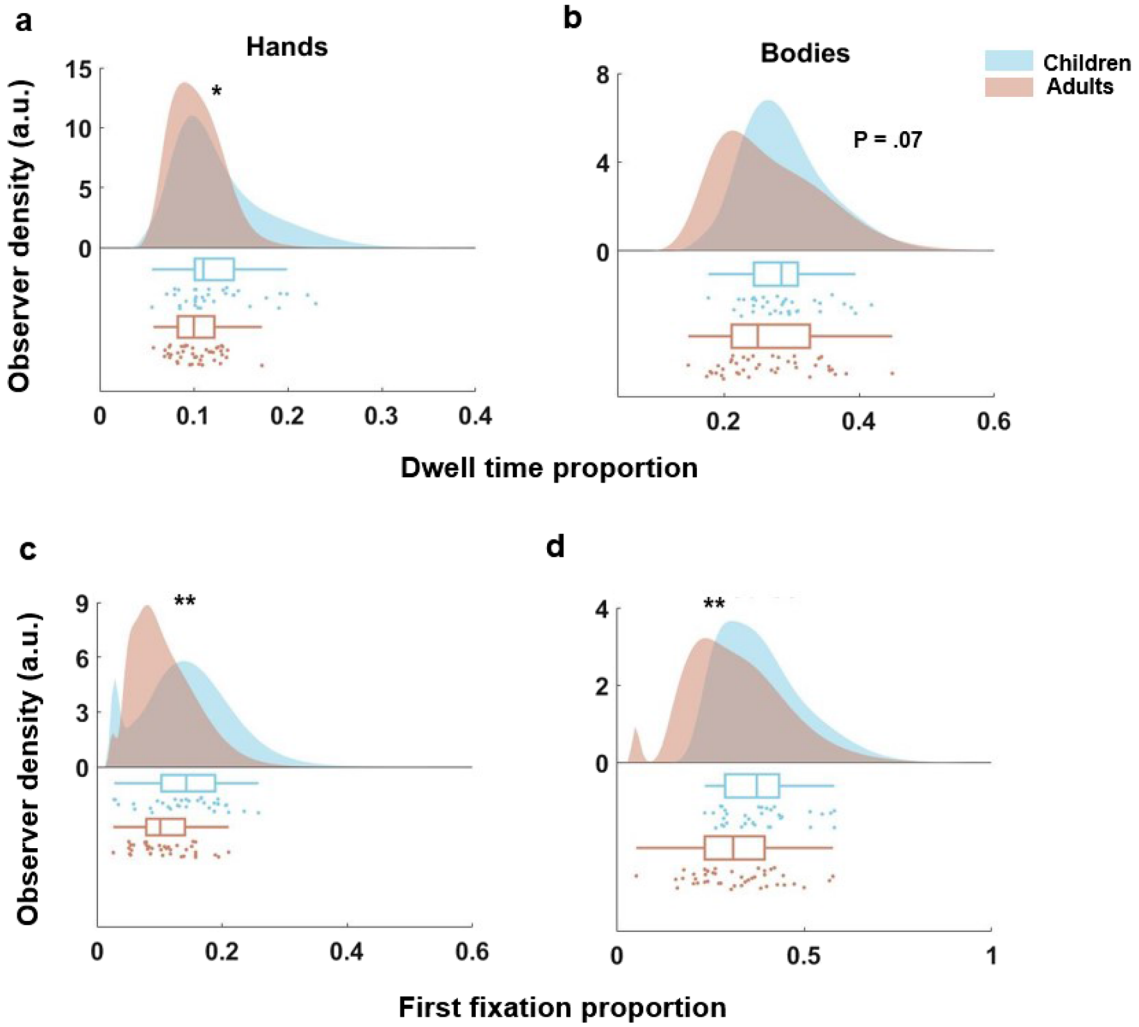

Figure 4. Group differences in gaze behavior towards hands and bodies. Density plots showing the probability distributions of cumulative dwell time (**a**, **b**) and first fixation proportion (**c**, **d**) towards objects of the dimensions *Hands* (**a**, **c**) and *Bodies* (**b**, **d**) for children (blue) and adults (red). Data points depicted below indicate the individual dwell time and first fixation proportion towards the respective dimension. Corresponding box plots above the data points show an overview of the summary statistics for each group and semantic category; the vertical line within a box represents the mean value. The left side of a box indicates the 25th percentile and the right side the 75th percentile. The whiskers represent the minimum and maximum values. $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$ (Holm-Bonferroni corrected; see Methods).

same time, the fixations of adults free-viewing complex scenes are highly drawn to text elements, which pre-literate children may attend less. Here, we investigated the gaze behavior of preschool children towards complex scenes and tested their biases for *Text*, *Faces*, *Hands*, *Touched* objects and *Bodies*.

In our simple inferential statistics, we show substantial differences in attentional biases between children and adults. Compared to adults, children spent a larger proportion of dwell time on *Faces*, *Touched* objects, *Hands* and (albeit only trending) *Bodies* when freely viewing complex scenes. Most significantly, children spent a lot less time fixating *Text*. Moreover, we found that children compared to adults placed significantly fewer first fixations on *Faces* and *Text*, but more on objects being *Touched, Hands* and *Bodies*. To control for the interdependence of gaze biases – in particularly the strong bias differences for *Text*, as well as potential differences regarding the effect of lower-level features (object size, eccentricity and visual salience), we applied follow-up (G)LMM analyses. For the fully specified model predicting dwell time, we saw a substantial reduction in the dwell time attraction of *Text* and a slight reduction in the attraction of *Faces* in children vs. adults. When predicting the probability of first fixating an object of a given category after image onset, we found a reduced attraction of *Text* and *Faces* in children compared to adults, which was accompanied by an enhanced attraction towards *Touched* objects in children. Further, our findings indicate that children tend to spend less time looking at larger objects and are less likely to fixate on them first compared to adults. On the other hand, children have a stronger center bias, both in terms of their dwell time towards objects and their initial fixation than adults. Excluding the *Text* dimension from the models, however, resulted in enhanced weights for almost all person-related features in children than adults, except *Faces* (see supplementary results). This is in line with the idea that many descriptive

Fixed effects	Description	B	SE	*t*	*P*			
Intercept	-	**−0.28**	**0.05**	**−5.42**	**<0.001**			
Age group	Children-adults	**0.21**	**0.05**	**4.17**	**<0.001**			
Text	Main effect	0.15	0.09	1.65	0.10			
Age group:text	Children-adults	**−0.84**	**0.08**	**−10.16**	**<0.001**			
Faces	Main effect	**0.52**	**0.07**	**7.38**	**<0.001**			
Age group:faces	Children-adults	**−0.12**	**0.06**	**−2.05**	**0.04**			
Touched	Main effect	**0.23**	**0.08**	**2.78**	**0.01**			
Age group:touched	Children-adults	0.04	0.06	0.71	0.48			
Hands	Main effect	0.06	0.08	0.70	0.49			
Age group:hands	Children-adults	−0.06	0.07	−0.87	0.39			
Bodies	Main effect	−0.04	0.07	−0.52	0.60			
Age group:bodies	Children-adults	−0.08	0.06	−1.31	0.19			
Eccentricity	Main effect	−0.04	0.03	−1.60	0.11			
Age group:eccentricity	Children-adults	**−0.11**	**0.02**	**−6.33**	**<0.001**			
Size	Main effect	**0.26**	**0.02**	**11.00**	**<0.001**			
Age group:size	Children-adults	**−0.07**	**0.02**	**−4.10**	**<0.001**			
GBVS	Main effect	**0.09**	**0.03**	**3.21**	**0.001**			
Age group:GBVS	Children-adults	0.02	0.02	1.04	0.30			
Random effects								
Groups	**Name**	***SD***	***r***					
Subjects	Intercept	0.18	Intercept					
	Text	0.18	−0.03	Text				
	Faces	0.16	0.01	−0.19	Faces			
	Touched	0.14	−0.25	0.30	−0.59	Touched		
	Hands	0.17	0.04	0.29	−0.30	0.90	Hands	
	Bodies	0.16	−0.27	0.47	0.03	0.76	0.86	Bodies
Objects	Intercept	0.41	0.41					
Scenes	Intercept	0.07	0.07					

Table 2. Results of linear mixed-effects model predicting object dwell time. Differences between groups were tested by including the relevant interaction terms. A beta estimate for a given interaction represents the estimated difference between children and adults. See methods for information regarding the applied coding scheme. Main effects represent the average estimates across both age groups. Significant effects are reported in bold.

gaze differences between children and adults can be explained by the drastically reduced tendency to fixate text in children. Children appear to allocate the fixations adults spent on text on person features and specifically on objects close to the hands.

For the main analyses, we did not include the semantic category *Motion* because there was an almost perfect overlap between body features (*Hands*, *Faces* and *Bodies*) and implied motion. An additional analysis suggests that children devote a significantly larger proportion of their dwell time towards objects with implied *Motion* (see supplementary materials). Future research should aim to test this motion bias in children in a stimulus set that allows to disentangle it from gaze biases towards body features.

We also found an enhanced center-bias in preschoolers, which matches recent evidence showing an enhanced center-bias in younger vs. older children[35]. Finally, we found no group difference in graph-based visual salience (GBVS) when predicting object-directed gaze behaviour. This contrasts previous reports of decreasing effects of visual salience with age[33]. One possible reason for this may be that the apparent group difference in the effects of visual salience may be better explained by semantic features, though this cannot be settled without targeted experiments explicitly disentangling the two (cf.[7]). Future research could test this explicitly.

Together, these findings contribute to our understanding of how human attention changes across the lifespan. Already very early in life our gaze becomes increasingly drawn towards distinct semantic social information[25]. This semantic information still seems to hold weight for preschoolers as well as adults. At the same time, our findings show that semantic gaze tendencies change between preschool childhood and adulthood, to include text, at the expense of attention towards body-related features like hands, bodies and objects being touched. This suggests differences in how adults and children perceive the same complex scenes, which in turn seem tied to the acquisition of reading. Interestingly, most of these changes in semantic attention extend to the first saccade after image onset. One exception are faces—*more* first fixations land on faces for adults compared to children. Whether this change of gaze biases appears as soon as literacy begins to evolve or gradually changes with reading experience is still an open question and represents an interesting subject for future research. Gaze

Fixed effects	Description	B	SE	*t*	*P*			
Intercept	-	**−4.92**	**0.21**	**−23.46**	**<0.001**			
Age group	Children-adults	−0.08	0.14	−0.60	0.55			
Text	Main effect	0.18	0.37	0.48	0.63			
Age group:text	Children-adults	**−1.35**	**0.28**	**−4.83**	**<0.001**			
Faces	Main effect	**2.38**	**0.28**	**8.35**	**<0.001**			
Age group:faces	Children-adults	**−0.46**	**0.16**	**−2.91**	**0.004**			
Touched	Main effect	0.62	0.34	1.80	0.07			
Age group:touched	Children-adults	**0.65**	**0.22**	**2.99**	**0.003**			
Hands	Main effect	0.56	0.34	1.67	0.09			
Age group:hands	Children-adults	0.16	0.21	0.75	0.45			
Bodies	Main effect	0.24	0.29	0.83	0.41			
Age group:bodies	Children-adults	0.31	0.17	1.84	0.07			
Eccentricity	Main effect	**−1.28**	**0.13**	**−10.03**	**<0.001**			
Age group:eccentricty	Children-adults	**−0.18**	**0.07**	**−2.68**	**0.01**			
Size	Main effect	**1.66**	**0.12**	**13.36**	**<0.001**			
Age group: size	Children-adults	**−0.17**	**0.07**	**−2.48**	**0.01**			
GBVS	Main effect	**0.78**	**0.13**	**6.05**	**<0.001**			
Age group:GBVS	Children-adults	0.04	0.06	0.60	0.55			
Random effects								
Groups	**Name**	***SD***	***r***					
Subjects	Intercept	0.18	Intercept					
	Text	0.18	−0.79	Text				
	Faces	0.16	0.72	−0.23	Faces			
	Touched	0.14	−0.57	0.06	−0.98	Touched		
	Hands	0.17	0.84	0.35	−0.81	0.73	Hands	
	Bodies	0.16	−0797	0.27	0.77	0.70	0.99	Bodies
Objects	Intercept	1.56	–					
Scenes	Intercept	0.0002	0.07					

Table 3. Results of binomial generalized linear mixed-effects model predicting first fixations towards objects. Differences between groups were examined by including the relevant interaction terms. A beta estimate for a given interaction represents the estimated difference between children and adults. See methods for information regarding the applied coding scheme. Main effects represent the average estimates across both age groups. Significant effects are reported in bold.

behavior towards semantic features appears to be the result of an interaction between image content, individual differences and experience.

This may explain why heritability estimates for gaze behaviour towards social scenes are substantially higher for infants[12] compared to older children (9–14 years[30]). Individual differences at the beginning of life seem dominated by genes, but may be modulated substantially by later experience, such as the acquisition of reading.

We hypothesize that reduced gaze behaviour towards text in children is due to limited literacy. Our preschool sample had not yet acquired formal education in reading and text features consequentially are uninformative to them. With age and improving reading ability, gaze behaviour may increasingly be tuned towards text at the expense of other semantic information. However, we did not assess reading ability so cannot exclude the possibility that at least some children already acquired reading skills before entering school.

As mentioned above, our basic inferential statistics further suggest that children spent a somewhat higher fraction of their dwell time on faces, but a *lower* fraction of their first fixations. This adds to previous works showing that face-based models are more successful in predicting gaze behaviour in adults compared to children[36], but that children showed enhanced proportions of first fixations towards social information when memorising a scene[50]. Our LMM control analyses, which excluded text as an explanatory variable, did not indicate any significant difference in the dwell time attraction of faces between children and adults. This suggests that any observed preference for faces in children may be attributable to corresponding object features such as eccentricity, size, and GBVS. However, when we controlled for text, our model did show a significantly lower preference for faces in children compared to adults. The finding that children directed a *lower* fraction of first fixations towards faces extended to the GLMM controlling for all other factors and may point to two separate visual processes with dissociable developmental trajectories driving face salience during the earlier and later stages of viewing a scene: one, indicated by first fixation distributions, that is largely bottom-up, guiding attention immediately after stimulus onset and showing a stronger, possibly more matured bias towards faces in adults. This may reflect more accurate and faster face processing abilities in adults[46], enabling them to saccade towards faces in the periphery

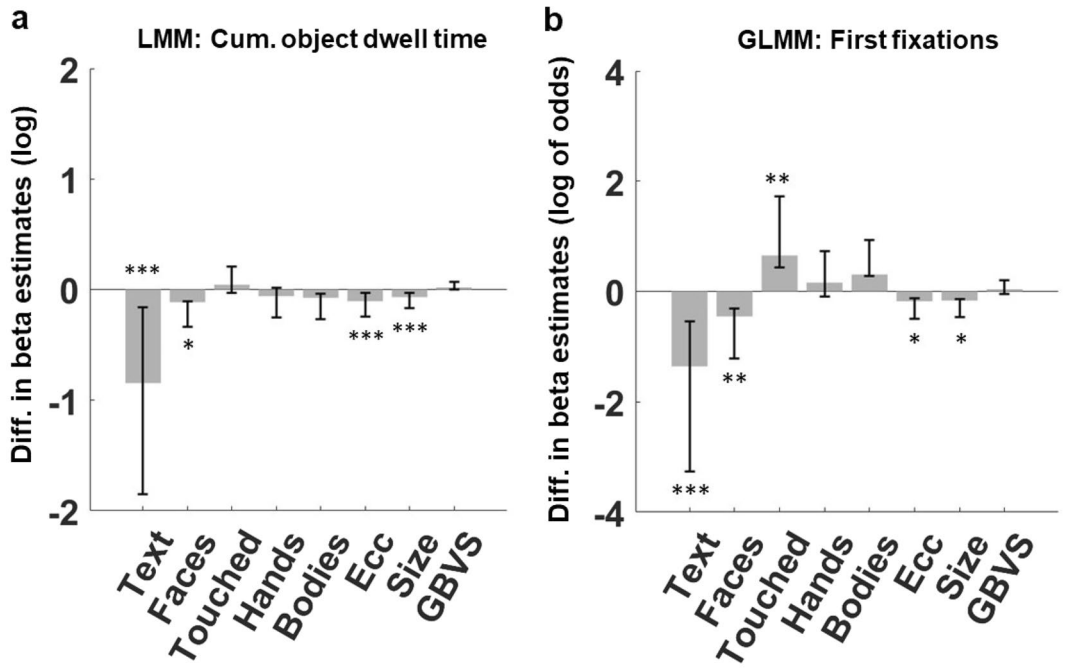

Figure 5. Group differences in predicting object-directed dwell time and first fixations. Bar plots depicting the beta estimates of a given object feature x *Age Group* interaction, which describe the difference in estimates between children and adults for a given object property. Panel a shows the interaction estimates for the LMM predicting object-directed dwell time and panel b depicts the binomial GLMM predicting first fixations towards objects after image onset across the included predictors *Text*, *Faces*, *Touched*, *Hands* and *Bodies* as well as Object-image centroid eccentricity (Ecc), (Object-) Size and graph-based visual salience (GBVS). Asterisks indicate the significance of a given group x predictor interaction. Results of (G)LMMs that did not take into account the effect of Text can be found in the supplementary materials. Error bars represent the 95% confidence intervals. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

very rapidly[46]. And a second one indicated by dwell time distributions, that may be under more top-down control and reflect the stronger competition of text with other semantic categories in adults for our stimuli.

Touched objects attracted a higher fraction of dwell time and first fixations in children compared to adults. The latter effect persisted in a GLMM controlling for all other modelled high- and low-level features. In a recent study, Nordt et al. showed that from young childhood to teen age, expanding word selective regions in ventral temporal cortex are directly linked to decreases in limb selectivity[43]. These findings are interpreted as evidence of cortical recycling of limb-selective areas for the visual word form area, emerging during reading acquisition. It is tempting to speculate that such effects of cortical recycling may be linked to the matching changes in visual preferences we see here—particularly for first fixations, that is, a shift in visual attention from limbs to text in adults compared to children.

Another candidate hypothesis for why children overall spent more of their dwell-time on touched objects and hands than adults is a possible immaturity of action understanding and social processing abilities. Critical social skills like reasoning about other people's thoughts (theory of mind; ToM) likely continue to develop beyond 5 years of age[51–53]. Children may compensate for limited action understanding and social processing abilities by allocating more dwell time towards hands and touched objects to extract the information necessary for making sense of goals and behaviours of people depicted in the scenes. This may be particularly pronounced for briefly presented complex scenes including multiple objects, people and their interactions. This is underscored by our model-based results which suggests that early visual preference for touched objects is increased in children, even when controlling for the lower visual preference for Text.

Further, there is evidence that children are embodied learners, that is, they seem to build their knowledge more strongly on sensorimotor experiences compared to adults[54]. It is possible, that gaze preferences in children for hands and objects being touched reflect such tuning towards physical and motor experiences.

Finally, children's enhanced gaze biases towards touched objects and hands may be related to their distinct drive to explore why and how things in the world interact as they do. Children seem to have a strong preference for understanding the causal structures surrounding them. For instance, when exposed to events that are inconsistent with their prior knowledge, preschool children seek causal explanations[55]. Moreover, preschoolers most frequently asked questions about functions and causal properties when they encountered novel objects[56,57]. This preference for causal information seems to be particular pronounced in preschoolers[58]. Children's gaze biases towards objects being touched and hands could be a consequence of this early drive to understand cause and effect and the functionality of objects.

Future studies could test this hypothesized relationship between causal inference and corresponding eye movements in preschool children in controlled experiments, juxtaposing stimuli with well-known and novel functions for children. Further, one could compare gaze behaviour towards semantic information in preschoolers and children who have already experienced formal education to test the effect of literacy on attention and specifically whether it is gradual or sudden. Given recent evidence for cortical recycling during childhood[43], future research could also examine whether a shift from cortical selectivity for limbs towards enhanced selectivity for words co-occurs with corresponding gaze biases. Finally, future research could probe gaze behaviour towards semantic dimensions in preschool children using more naturalistic stimuli, like videos of everyday scenes. This could reveal whether present gaze differences—possibly including implied motion—also hold for dynamic scenes.

Taken together, we report several differences in gaze biases between preschool children and adults. Children spent a far smaller proportion of dwell time on *Text* and more on *Faces*, *Touched* objects, *Hands* and a corresponding trend for *Bodies*. For the dimensions *Text*, *Touched* and *Hands*, these biases already show for first fixations after image onset, while children directed *fewer* first fixations on *Faces*. Further analyses controlling for *Text* and other salience factors (GLMM analyses) suggest that most body-related attentional biases in children are due to the lack of competition with *Text* seen in adults. One exception is the higher propensity of *Touched* objects to attract first fixations in children, which persisted even when controlling for *Text* salience.

These findings show that fixation behaviour towards semantic features changes over the lifespan. Children and adults may perceive the same visual environment in different ways, which in turn may be related to the acquisition of reading ability, socio-cognitive development and cortical recycling. Future research examining the relationship between these factors and gaze behavior can inform theories on the development of human visual attention.

Methods

Subjects. In total n = 47 children and n = 46 adults with normal or corrected vision took part in the study. We only included subjects with at least one fixation for more than 80% of the trials (34 trials). Applying this criterion, we removed 17 subjects from the original sample (13 children and 4 adults). This resulted in a final sample of $n = 34$ children and $n = 42$ adults.

Of the remaining sample, children on average showed 2.9% and adults 1.1% missing trials.

All subjects provided written informed consent before participation. For children, parents gave written informed consent. The study was approved by the local ethics committee of Justus-Liebig-University Giessen and adhered to the declaration of Helsinki.

Children ($n = 34$; $M_{age} = 5{,}7$; range = 5.1–5.9; $SD = 0.17$; 19 females) were recruited as part of a larger study based on a local data base of parents, who indicated their interest in child development studies. No child in the sample attended school at the time of the study. All children completed another task which was not related to the present study. Subjects received no financial reimbursement for participation. Children were rewarded with a certificate of participation and a small gift.

Adults (n = 42; $M_{age} = 24.4$; range = 18–59; $SD = 7.05$; 31 females) were recruited as part of a larger study and completed other tasks which were unrelated to the present study. Adult participants were compensated with money (7€/hr) or course credit for participation.

Apparatus. The free viewing task was created and implemented using Psychopy 2020.1.2[48] in Python version 3.6.10[59]. Stimuli were shown on a Lenovo ThinkPad X230 laptop with a screen resolution of 1366 × 768 pixels. The stimuli were presented with a size of 1000 × 750 pixels, roughly corresponding to 21 × 16 degrees visual angle (d.v.a.). Eye movements were recorded from both eyes using a head-free Tobii 4c Eye Tracker (Tobii AB, Danderyd, Sweden) operating in remote mode at 90 Hz and a distance of 50–90 cm. No official information has been released stating the accuracy or precision of the Tobii 4c Eye Tracker. Work evaluating the predecessor model Tobii EyeX has reported an accuracy of < 0.6° and precision of < 0.25°[60].

Stimuli and procedure. We used the OSIE40[5,11] dataset, which includes 40 complex everyday scenes and corresponding pixel masks for 364 objects with binary labels for 12 semantic object dimensions. Additionally, we used pixel masks for bodies and hands recently published by Broda and de Haas[47]. We recently found that individual gaze tendencies towards semantic features could be estimated reliably with 40 images[11]. As in previous analyses[11], the 12 labels for the semantic dimensions were modified in order to reduce overlap between them in the following way: The *Face* label was removed from all object fixations with the *Emotion* label; the *Smell* label was removed from all object fixations with a *Taste* label; and the *Operable* and *Gazed* label were removed from object fixations with a *Touched* label. Objects were labelled as *Touched*, when they were touched by a human or animal in the scene. The *Watchable* label was removed from all object fixations with the label *Text*.

The experiment took place at a laboratory with only the participant and experimenter present. All participants were placed at a distance of 50–60 cm from the screen/eye tracker (the recommended user distance of the eye tracker is between 50 and 90 cm) and reminded to sit as still as possible during the session. To assure an appropriate position, a custom position guide interfaced with the eye tracker and gave online feedback about whether or not a participant was placed within the trackable range. On average, subjects sat at a distance of ~ 54.5 cm from the screen (children = 57 cm, adults = 52 cm). All viewing angles were calculated based on the average position distance of the respective group.

After completing a child-friendly 5-point calibration and validation procedure, subjects were instructed to freely view 40 images. Each image was presented centrally on the screen for 3 s, with a fixation cross in between trials presented centrally on the screen. A trial could only be initiated if a subject's gaze did not deviate 2 d.v.a. from the fixation cross for 1 s. Images were presented in the same order across subjects.

To test eye tracker accuracy across trials, we calculated the deviation of trial onset fixations from the nominal center for each individual and group (i.e. the median distance in d. v. a. between the fixation cross shown between trials and the onset fixation). Results indicate that this deviation was minimal for both groups (M = 0.37, SD = 0.14 for children and M = 0.46, SD = 0.23 for adults). The slightly larger error for adults was not statistically significantly different from that for children, *tobs* = 1.08, $p = 0.07$. To test the estimated drift across the task, we compared the deviation from the nominal center between the first and last trial for each observer and group (again in d.v.a.). We found that for children, there was no significant difference in drift between the first and last trials (M = 0.1, SD = 1.44, *tobs* = − 0.38, $p = 0.34$). For adults, there was a significant difference showing larger drift (M = 0.29, SD = 0.81, *tobs* = 2.38, $p < 0.001$). Finally, we tested the difference in drift between groups and found no significant differences (*tobs* = − 0.74, $p = 0.07$).

Data processing. All pre-processing steps and statistical analyses were computed using MATLAB R2019B (MathWorks). Gaze data from both eyes were averaged. Fixations were extracted from raw eye tracking data by applying an inter-sample saccade velocity threshold of 30 d.v.a/s and a spatial inter-sample distance threshold of 2 d.v.a. Thus, samples were successively marked as fixations if both the corresponding inter-sample velocity was below 30 d.v.a/s and the spatial inter-sample distance was below 2 d.v.a. The latter criterion was applied to account for possible intermittent sample loss. Further, we used the median x and y position of samples deemed to be part of a fixation to determine the respective fixation location. Fixations with a duration < 100 ms were not considered. We further removed all fixations with an onset time (after trial onset) of < 100 ms to exclude fixations and saccades initiated before image onset. Saccade onsets and offsets were detected by applying an inter-sample velocity threshold of 30 d.v.a./s. Only saccades with amplitudes > 2 d.v.a. were considered.

A given fixation can be assigned multiple semantic features. This is because we took the extent of the fovea and accuracy limits of the eyetracker into account by applying a tolerance margin when assigning semantic labels to fixations. Each fixation was assigned the labels of all objects falling within a radius of ~ 0.5 d.v.a. surrounding the pixel coordinates of the fixation. The margin size was selected based on the tested accuracy of the Tobii 4c eyetracker and its previous model. For more information on the eye tracker's accuracy, please refer to the information provided above. Additionally, for further discussion on best practices, please refer to Holmqvist et al.[61] and Orquin et al.[62]. Figure 6 gives an overview of label overlap among all included fixations.

Fixation tendencies across semantic dimensions. First, we tested differences in fixation tendencies along three semantic dimensions (*Faces*, *Text* and *Touched*) between children and adults. We chose these three dimensions, because adult gaze behaviour has been shown to reliably vary along them for the stimuli used in the present study (OSIE40)[11]. We calculated the proportion of individual cumulative dwell time falling on each of the aforementioned categories, which we previously found to be the best predictors of individual and group-level gaze behaviour. For this, the sum of respective fixation durations was divided by the sum of fixation durations on any object labelled in the OSIE40 or the *Body* and *Hand* masks which we derived from Broda & de Haas[47]. We further determined the proportion of all object directed first fixations (following the first saccade after image onset) which landed on *Text*, *Faces*, *Touched* objects for each participant (first fixation proportion. Given that

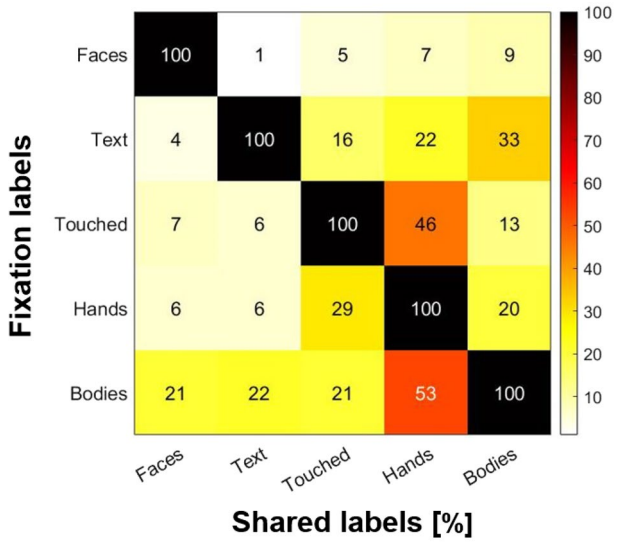

Figure 6. Label overlap across semantic categories. Heatmap showing the percentage of mutual label assignments. Each row shows the shared labels for all fixations bearing a given label. For example, the bottom row shows which percentage of *Bodies* fixations were also labelled as *Face*, *Text*, *Touched* and so on. Note, here we have calculated the proportion of fixations of a given category that is also labelled another category either (1) because the objects have several labels or (2) the fixation fell on several objects of different labels. The heatmap was created using MATLAB R2019B (MathWorks).

some variables of interest were not normally distributed we tested group differences by applying non-parametric permutation tests. For this approach, in each iteration, we first pooled individual fixation data across all included 76 observers and then drew two random subsamples of respectively n = 34 and n = 42 without replacement from the pooled sample. After this, we calculated the mean difference between both samples for a given metric. This was repeated 10,000 times, resulting in a null distribution consisting of 10,000 permutated estimates. The *p*-value was determined as the proportion of permutated estimates that were more extreme than the true mean difference. Finally, we calculated the *t*-test statistic between the observed scores of both groups for a given metric (here referred to as *t*obs). *P*-values were Holm-Bonferroni for the number of included dimensions in each model. All group comparisons (Figs. 2, 3, 4) are visualized using Raincloud Plots[63].

Explaining object-directed fixations via (G)LMMs. Finally, we used (G)LMMs to explain the likelihood and cumulative duration of fixations as a function of object and observer properties. This allowed us to test age group differences in the predictive weights of semantic features, object size, eccentricity, and low-level salience. Our analysis followed the approach by Nuthmann et al., which employed a comparable model-based methodology while also accounting for center bias, object size, and visual salience[48,49]. Cumulative fixation duration towards objects was modeled as a continuous dependent variable of a LMM (estimated via Restricted Maximum Likelihood; REML). Here, cumulative fixation duration is defined as the total dwell time spent on a given object. Further, we only included objects that were fixated at least once. First object-directed fixation probability after image onset was modelled by fitting a binomial GLMM (using Laplace approximation) aiming to explain a binary response variable indicating whether an object was first fixated (1) or not (0) after image onset. Both models included the following fixed-effects terms: object-image *Eccentricity*, expressed as the distance between an object centroid and the image center in d.v.a.; object *Size* as the pixel coverage of an object in % of the corresponding image; the mean of graph-based visual salience (*GBVS*) across pixels of the object; and binary dummy coded variables indicating whether the object was labelled as *Text*, *Face*, *Touched*, *Hands* or *Body*. Finally, we included the predictor *Age Group*, using effect coding (− 0.5/0.5) with adults as the reference group. This has the advantage of retrieving main effects (average estimates across both age groups) for each given predictor variable instead of simple effects. Note that our interest here was limited to group comparisons and estimates have to be interpreted as relative to other types of objects (and differ from pixel-wise salience models in that regard). The continuous variables *Eccentricity*, object *Size*, *GBVS*, and fixation duration were z-standardized across all objects in the dataset. Moreover, we log-transformed object *Size* and cumulative fixation time, because both distributions were positively skewed. To test age group differences, we further included interaction terms between *Age Group* and each remaining predictor.

In our present study design the random factor *Subject* is nested under *Age Group*. Moreover, the random factor *Object* is nested under the random factor *Scene*. We therefore included *Subject* as random intercept as well as by-subject random slopes for all semantic feature predictors (*Text*, *Faces, Touched, Hands and Bodies*). Finally, we included *Object* and *Scene* (i.e. image) as random intercepts.

The formula for the model-based analyses for both the binomial as well as the linear model was:

Response variable ~ 1 + Age Group + Text + Age Group:Text + Faces + Age Group:Faces + Touched + Age Group:Touched + Hands + Age Group:Hands + Bodies + Age Group: Bodies + ObjSize + Age Group:ObjSize + Ecc + Age Group:Ecc + GBVS + Age Group:GBVS + (1 + Text + Faces + Touched + Hands + Bodies |Subject) + (1|Object) + (1|Scene).

Data availability

Anonymized data and code to reproduce the presented findings and figures are available at osf.io/78aqf/.

Received: 14 April 2022; Accepted: 16 July 2023
Published online: 21 July 2023

References

1. Strasburger, H., Rentschler, I. & Jüttner, M. Peripheral vision and pattern recognition: A review. *J. Vis.* **11**, 13 (2011).
2. Rosenholtz, R. Capabilities and limitations of peripheral vision. *Annu. Rev. Vis. Sci.* **2**, 437–457 (2016).
3. Harel, J., Koch, C. & Perona, P. Graph-based visual saliency. *Adv. Neural Inform. Process. Syst* **19** (2006).
4. Itti, L., Koch, C., Itti, L. & Koch, C. A saliency-based mechanism for overt and covert shifts of visual attention. *Vis. Res.* **40**, 1489–1506 (2000).
5. Xu, J., Jiang, M., Wang, S., Kankanhalli, M. & Zhao, Q. Predicting human gaze beyond pixels. *J. Vis.* **14**, 28–28 (2014).
6. Einhäuser, W., Spain, M. & Perona, P. Objects predict fixations better than early saliency. *J. Vis.* **8**, 18 (2008).
7. Stoll, J., Thrun, M., Nuthmann, A. & Einhäuser, W. Overt attention in natural scenes: Objects dominate features. *Vis. Res.* **107**, 36–48 (2015).
8. Wang, H.-C. & Pomplun, M. The attraction of visual attention to texts in real-world scenes. *J. Vis.* **12**, 26–26 (2012).
9. Kümmerer, M., Wallis, T. S. A., Gatys, L. A. & Bethge, M. Understanding low- and high-level contributions to fixation Prediction. In *2017 IEEE International Conference on Computer Vision (ICCV)* 4799–4808 (2017). https://doi.org/10.1109/ICCV.2017.513
10. Haas, B., Iakovidis, A., Schwarzkopf, D. & Gegenfurtner, K. Individual differences in visual salience vary along semantic dimensions. *Proc. Natl. Acad. Sci.* https://doi.org/10.1101/444257 (2018).
11. Linka, M. & de Haas, B. OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *J. Vis.* **20**, 13 (2020).
12. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344 (2017).
13. Anouck, A., Guillaud, E., Bouvard, M. & Cazalets, J.-R. Developmental changes in face visual scanning in autism spectrum disorder as assessed by data-based analysis. *Front. Psychol.* **6**, 989 (2015).
14. Linka, M., Broda, M. D., Alsheimer, T., de Haas, B. & Ramon, M. Characteristic fixation biases in super-recognizers. *J. Vis.* **22**, 17 (2022).

15. Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **40**, 1–19 (1991).
16. Johnson, M. H., Senju, A. & Tomalski, P. The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neurosci. Biobehav. Rev.* **50**, 169–179 (2015).
17. Reid, V. M. *et al.* The human fetus preferentially engages with face-like visual stimuli. *Curr. Biol.* **27**, 1825-1828.e3 (2017).
18. Scheel, A. M., Ritchie, S. J., Brown, N. J. L. & Jacques, S. L. Methodological problems in a study of fetal visual perception. *Curr. Biol.* **28**, R594–R596 (2018).
19. Reid, V. M., Dunn, K., Donovan, T. & Young, R. J. Response to Scheel et al. *Curr. Biol.* **28**, R596–R597 (2018).
20. Reynolds, G. & Roth, K. The development of attentional biases for faces in infancy: A developmental systems perspective. *Front. Psychol.* **9**, 222 (2018).
21. Chien, S.H.-L. No more top-heavy bias: Infants and adults prefer upright faces but not top-heavy geometric or face-like patterns. *J. Vis.* **11**, 13–13 (2011).
22. Gluckman, M. & Johnson, S. Attentional capture by social stimuli in young infants. *Front. Psychol.* **4**, 527 (2013).
23. Andravizou, A., Gliga, T., Elsabbagh, M. & Johnson, M. Faces attract infants' attention in complex displays. *Ann. Gen. Psychiatry* **14**, 550–562 (2008).
24. Kelly, D., Duarte, S., Meary, D., Bindemann, M. & Pascalis, O. Infants rapidly detect human faces in complex naturalistic visual scenes. *Dev. Sci* **22**, e12829 (2019).
25. Frank, M., Vul, E. & Saxe, R. Measuring the development of social attention using free-viewing. *Infancy* **17**, 355–375 (2011).
26. Simion, F., Regolin, L. & Bulf, H. A predisposition for biological motion in the newborn baby. *Proc. Natl. Acad. Sci.* **105**, 809–813 (2008).
27. Shirai, N. & Imura, T. Implied motion perception from a still image in infancy. *Exp. Brain Res.* **232**, 3079–3087 (2014).
28. Shirai, N. & Imura, T. Emergence of the ability to perceive dynamic events from still pictures in human infants. *Sci. Rep.* **6**, 37206 (2016).
29. Guy, N. *et al.* A novel perceptual trait: Gaze predilection for faces during visual exploration. *Sci. Rep.* **9**, 10714 (2019).
30. Kennedy, D. P. *et al.* Genetic influence on eye movements to complex scenes at short timescales. *Curr. Biol.* **27**, 3554-3560.e3 (2017).
31. Yu, C. & Smith, L. B. The social origins of sustained attention in one-year-old human infants. *Curr. Biol.* **26**, 1235–1240 (2016).
32. Yu, C., Suanda, S. & Smith, L. Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Dev. Sci.* **22**, e12735 (2018).
33. Acik, A., Sarwary, A., Schultze-Kraft, R., Onat, S. & König, P. Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Front. Psychol.* **1**, 207 (2010).
34. Helo, A., Pannasch, S., Sirri, L. & Rämä, P. The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vis. Res.* **103**, 83–91 (2014).
35. Krishna, O., Helo, A., Rämä, P. & Aizawa, K. Gaze distribution analysis and saliency prediction across age groups. *PLoS ONE* **13**, e0193149 (2018).
36. Rider, A., Coutrot, A., Pellicano, E., Dakin, S. & Mareschal, I. Semantic content outweighs low-level saliency in determining children's and adults' fixation of movies. *J. Exp. Child Psychol.* **166**, 293–309 (2018).
37. Jones, W., Carr, K. & Klin, A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch. Gen. Psychiatry* **65**, 946–954 (2008).
38. Klin, A. & Jones, W. Altered face scanning and impaired recognition of biological motion in a 15-month-old infant with autism: Report. *Dev. Sci.* **11**, 40–46 (2008).
39. Hanania, R. & Smith, L. Selective attention and attention switching: Toward a unified developmental approach. *Dev. Sci.* **13**, 622–635 (2010).
40. Chouinard, M. Children's questions: A mechanism for cognitive development [Monograph]. *Monogr. Soc. Res. Child Dev.* **72**(1), vii–ix (2007).
41. Gopnik, A. Explanation as orgasm*. *Minds Mach (Dordr)* **8**, 101–118 (1998).
42. Brown, T. & Jernigan, T. Brain development during the preschool years. *Neuropsychol. Rev.* **22**, 313–333 (2012).
43. Nordt, M. *et al.* Cortical recycling in high-level visual cortex during childhood development. *Nat. Hum. Behav.* **5**, 1686–1697 (2021).
44. Parkhurst, D. *et al.* Modeling the role of salience in the allocation of overt visual attention. *Vis. Res.* **42**, 107–123 (2002).
45. Anderson, N., Ort, E., Kruijne, W., Meeter, M. & Donk, M. It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *J. Vis.* **15**, 9 (2015).
46. Crouzet, S., Kirchner, H. & Thorpe, S. Fast saccades toward faces: Face detection in just 100 ms. *J. Vis.* **10**(16), 1–17 (2010).
47. Broda, M. D. & de Haas, B. Individual differences in looking at persons in scenes. *J. Vis.* **22**, 9 (2022).
48. Nuthmann, A., Schütz, I. & Einhäuser, W. Salience-based object prioritization during active viewing of naturalistic scenes in young and older adults. *Sci. Rep.* **10**, 22057 (2020).
49. Nuthmann, A., Einhäuser, W. & Schütz, I. How well can saliency models predict fixation selection in scenes beyond central bias? A new approach to model evaluation using generalized linear mixed models. *Front. Hum. Neurosci.* **11**, 491 (2017).
50. Doherty, B. R., Fraser, A., Nobre, A. C. & Scerif, G. The functional consequences of social attention on memory precision and on memory-guided orienting in development. *Dev. Cogn. Neurosci.* **36**, 100625 (2019).
51. Golarai, G. *et al.* Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat. Neurosci.* **10**, 512–522 (2007).
52. Osterhaus, C. & Koerber, S. The development of advanced theory of mind in middle childhood: A longitudinal study from age 5 to 10 years. *Child Dev.* **92**, 1872–1888 (2021).
53. Saxe, R., Whitfield-Gabrieli, S., Scholz, J. & Pelphrey, K. Brain regions for perceiving and reasoning about other people in school-aged children. *Child Dev.* **80**, 1197–1209 (2009).
54. Wellsby, M. & Pexman, P. Developing embodied cognition: Insight from children's concepts and language processing. *Front. Psychol.* **5**, 506 (2014).
55. Legare, C., Gelman, S. & Wellman, H. Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Dev.* **81**, 929–944 (2010).
56. Greif, M., Nelson, D., Keil, F. & Gutierrez, F. What do children want to know about animals and artifacts?: Domain-specific requests for information. *Psychol. Sci.* **17**, 455–459 (2006).
57. Nelson, D., Egan, L. & Holt, M. When children ask, "What is it?" What do they want to know about artifacts?. *Psychol. Sci.* **15**, 384–389 (2004).
58. Booth, A. & Alvarez, A. Developmental changes in causal supports for word learning. *Lang. Learn. Dev.* **11**, 80–92 (2014).
59. van Rossum, G. & Drake, F. L. Python 3 Reference Manual (2009).
60. Gibaldi, A., Vanegas, M., Bex, P. J. & Maiello, G. Evaluation of the Tobii EyeX eye tracking controller and Matlab toolkit for research. *Behav. Res. Methods* **49**, 923–946 (2017).
61. Holmqvist, K. *et al.* Eye tracking: Empirical foundations for a minimal reporting guideline. *Behav. Res. Methods* https://doi.org/10.3758/s13428-021-01762-8 (2022).
62. Orquin, J. L., Ashby, N. J. S. & Clarke, A. D. F. Areas of interest as a signal detection problem in behavioral eye-tracking research. *J. Behav. Decis. Mak.* **29**, 103–115 (2016).

63. Allen, M., Poggiali, D., Whitaker, K., Marshall, T. & Kievit, R. Raincloud plots: A multi-platform tool for robust data visualization. *Welcome Open Res.* https://doi.org/10.7287/peerj.preprints.27137v1 (2018).

Acknowledgements
We would like to thank Amelie Benner and Nele Zscherper for data collection. ML and BdH are supported by ERC Starting Grant 852885 INDIVISUAL. BdH and GS are further supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 222641018–SFB/TRR 135 TP C9 and ÖS is supported by DFG Forschergruppe 2253 Crossing the Borders, and BdH and GS by "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. Finally, we would like to thank Xu et al. as well as Broda & de Haas for making their stimuli, pixel masks and metadata publicly available[5,47].

Author contributions
M.L. and B.d.H. designed and implemented the experiment in collaboration with H.K., Ö.S. and G.S. Ö.S. conducted the experiment. M.L. analysed the data and prepared the initial draft of the manuscript in collaboration with B.d.H. G.S. and B.d.H. administered and supervised the project. All authors provided comments and/or revisions and approved the final article.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-38854-8.

Correspondence and requests for materials should be addressed to M.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Materials

Methods

Group differences in basic oculomotor metrics

To test whether children and adults differ in their gaze behaviour along basic oculomotor measures, we computed the mean fixation duration, saccadic amplitude, fixation duration of first fixations, saccadic amplitude of first saccades and first saccade latency respectively for each individual and group. After this, we applied two sampled *t*-test for group comparisons. The corresponding *p*-values were Holm-Bonferroni corrected for all 5 tests. We recommend interpreting these results with caution, given the limited temporal resolution of the eyetracker we used (90 Hz).

Group differences in gaze behaviour towards semantic dimensions

Our main analyses excluded background fixations. To control for possible effects of this, we ran a control analysis testing group differences in the proportion of *overall* dwell time and first fixations towards *Text*, *Faces*, *Touched* objects, *Hands* and *Bodies*. All group differences were tested for statistical significance via permutation tests. *P*-values were Holm-Bonferroni corrected for the dimensions included in a given model.

(G)LMM analyses not controlling for text

To illustrate the influence of the group difference in *Text* fixations on gaze differences towards any other semantic dimension, we have additionally fitted trimmed (G)LMMs. These models predicted object-directed dwell time (LMM) and the probability of first object-directed fixation (binomial GLMM) based on all previously added predictors excluding *Text*.

Group difference in dwell time towards objects with implied motion

Finally, we tested group differences in cumulative dwell time towards objects with implied *Motion*. We did not include *Motion* in our main analyses, because it almost perfectly overlaps with *Bodies*, *Hands*, and *Faces*. Here we nevertheless explored the proportion of

object-directed cumulative dwell time for *Motion* between children and adults and tested the group difference via a permutation test.

Results

Group differences in basic oculomotor metrics

Table S1 provides an overview of the mean values and standard deviations regarding all five included metrics for children and adults as well as test statistics and *p*-values for each group comparison. We found no significant group differences between children and adults for the variables mean fixation duration, mean saccadic amplitude, mean first fixation duration and first saccade latency. However, we found that children exhibited significantly larger first saccades compared to adults.

Table S.1. Group differences in basic oculomotor metrics

	Adults		Children				
	M	*SE*	*M*	*SE*	*t*	*df*	*P*
Fix. duration (ms)	302.6	11.18	318.41	9.48	-0.05	74	0.63
Sacc. ampl (DVA)	5.87	0.12	5.68	0.11	-1.11	74	0.27
First fix. duration (ms)	297.57	18.12	300.14	11.23	-0.11	74	0.91
First sacc. amplitude (DVA)	3.96	0.08	4.46	0.16	2.91	74	0.004
First sacc. latency	357.7	25.5	377.12	11.77	-0.64	74	0.52

Group differences in gaze behaviour towards semantic dimensions including background fixations

Fig. S1 shows the distributions and mean values of dwell time (a) and first fixation proportions (b) for each semantic dimension and group when including background fixations in the analysis. The pattern of results was highly similar to that observed when excluding background fixations (cf. Fig. 3 and 4, main text).

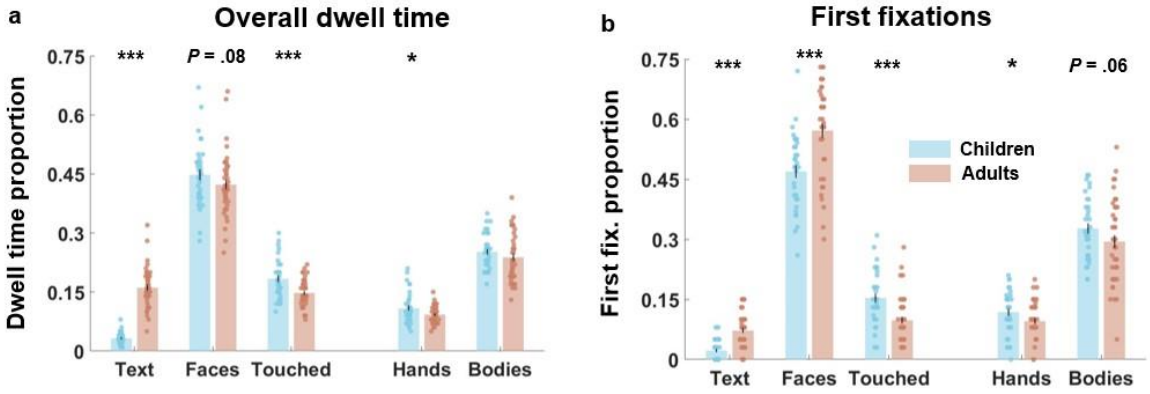

Figure S1. **Group differences in attentional biases towards semantic dimensions**. Bar plots indicate mean values for the proportion of overall dwell time (a) and proportion of overall first fixations (b) respectively for children (blue) and adults (red). Data points indicate the individual mean values for a given dimension and group. Error bars represent the SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Holm-Bonferroni corrected; see Methods)

Our findings show that children spent a larger proportion of their overall dwell time towards *Touched* objects, $t_{obs} = 3.76$, $p < 0.001$, *Faces* (only trending; $t_{obs} = 1.37$, $p = 0.08$) and *Hands* ($t_{obs} = 2.25$, $p = 0.03$). Further, children spent a lot less time fixating *Text* objects compared to adults, $t_{obs} = -14.36$, $p < 0.001$. No significant group difference was found for the dimension *Bodies*, $t_{obs} = 1.1$, $p = 0.14$. Considering group differences in first fixation proportions, we found that adults compared to children devoted a significantly larger proportion of their first fixations towards *Faces*, $t_{obs} = -4.22$, $p < 0.001$ and *Text*, $t_{obs} = -5.98$, $p < 0.001$. Further, children directed a significantly larger proportion of first fixations towards *Touched* objects, $t_{obs} = 3.92$, $p < 0.001$. Finally, children showed a larger proportion of first fixation on *Hands*, $t_{obs} = 2.03$, $p = 0.047$ and *Bodies*, $t_{obs} = 1.53$, $p = 0.06$ (only trending).

(G)LMM analyses not controlling for text

Model coefficients, standard errors, *t*-values and *p*-values for results are reported in Tables S2 and S3 respectively. Fig S2 shows the estimates of each included Age group x object feature interaction for both models.

Table S2. Results of Linear Mixed-Effects Model predicting object dwell time without Text. Differences between groups were tested by including the relevant interaction terms. A beta estimate for a given interaction represents the estimated difference between children and adults. Main effects represent the average standardized estimates across both age groups Significant effects are reported in bold.

Fixed Effects	**Description**	**B**	**SE**	***t***	***P***
Intercept	**Main effect**	**-0.20**	**0.05**	**-4.18**	**<0.001**
Age Group	Children-Adults	0.04	0.05	0.68	0.49
Faces	**Main effect**	**0.43**	**0.07**	**6.30**	**<0.001**
Age Group: Faces	Children-Adults	0.07	0.06	1.21	0.23
Touched	Main effect	0.14	0.08	1.76	0.08
Age Group: Touched	**Children-Adults**	**0.22**	**0.06**	**3.87**	**<0.001**
Hands	Main effect	-0.04	0.08	-0.50	0.61
Age Group: Hands	**Children-Adults**	**0.15**	**0.07**	**2.29**	**0.02**
Bodies	Main effect	-0.13	0.07	-1.86	0.06
Age Group: Bodies	**Children-Adults**	**0.07**	**0.06**	**1.33**	**0.18**
Eccentricity	Main effect	-0.05	0.03	-1.77	0.08
Age Group: Eccentricity	**Children-Adults**	**-0.10**	**0.02**	**-5.94**	**<0.001**
Size	**Main effect**	**0.25**	**0.02**	**10.56**	**<0.001**
Age Group: Size	Children-Adults	-0.03	0.02	-1.80	0.07
GBVS	**Main effect**	**0.08**	**0.03**	**3.05**	**0.002**
Age Group: GBVS	Children-Adults	0.02	0.02	1.04	0.30

Random effects

Groups	**Name**	***SD***	***r***
Subjects	Intercept	0.19	Intercept

	Faces	0.19	-0.20	Faces			
	Touched	0.13	-0.32	-0.37	Touched		
	Hands	0.16	-0.03	-0.14	0.90	Hands	
	Bodies	0.15	-0.30	0.14	0.77	0.88	Bodies
Objects	Intercept	0.42	-				
Scenes	Intercept	0.06	-				

Table S3. Results of binomial Generalized Linear Mixed-Effects Model predicting first fixations towards objects without text. Differences between groups were examined by including the relevant interaction terms. A beta estimate for a given interaction represents the estimated difference between children and adults. Main effects represent the average standardized estimates across both age groups Significant effects are reported in bold.

Fixed Effects	**Description**	**B**	**SE**	**t**	**P**
Intercept	**Main effect**	**-4.83**	**0.21**	**-22.94**	**<0.001**
Age Group	**Children-Adults**	**-0.48**	**0.12**	**-4.09**	**<0.001**
Faces	**Main effect**	**2.27**	**0.26**	**8.73**	**<0.001**
Age Group:Faces	Children-Adults	-0.10	0.14	-0.70	0.48
Touched	Main effect	0.53	0.33	1.62	0.11
Age Group:Touched	**Children-Adults**	**0.98**	**0.21**	**4.79**	**<0.001**
Hands	Main effect	0.49	0.32	1.53	0.12
Age Group:Hands	**Children-Adults**	**0.54**	**0.20**	**2.75**	**0.01**
Bodies	Main effect	0.07	0.27	0.25	0.80
Age Group:Bodies	**Children-Adults**	**0.59**	**0.16**	**3.82**	**<0.001**
Eccentricity	**Main effect**	**-1.34**	**0.13**	**-10.49**	**<0.001**
Age Group:Eccentricity	**Children-Adults**	**-0.20**	**0.07**	**-3.10**	**0.002**
Size	**Main effect**	**1.78**	**0.13**	**13.40**	**<0.001**
Age Group:Size	Children-Adults	-0.09	0.07	-1.39	0.16
GBVS	**Main effect**	**0.78**	**0.13**	**5.99**	**<0.001**

Age Group:GBVS	Children-Adults	0.05	0.06	0.86	0.39	

Random effects

Groups	**Name**	***SD***	***r***				
Subjects	Intercept	0.04	Intercept				
	Faces	0.29	0.88	Faces			
	Touched	0.55	-0.80	-0.99	Touched		
	Hands	0.41	-0.99	-0.82	0.73	Hands	
	Bodies	0.35	-0.98	-0.77	0.68	1	Bodies
Objects	Intercept	1.44	-				
Scenes	Intercept	0.67	-				

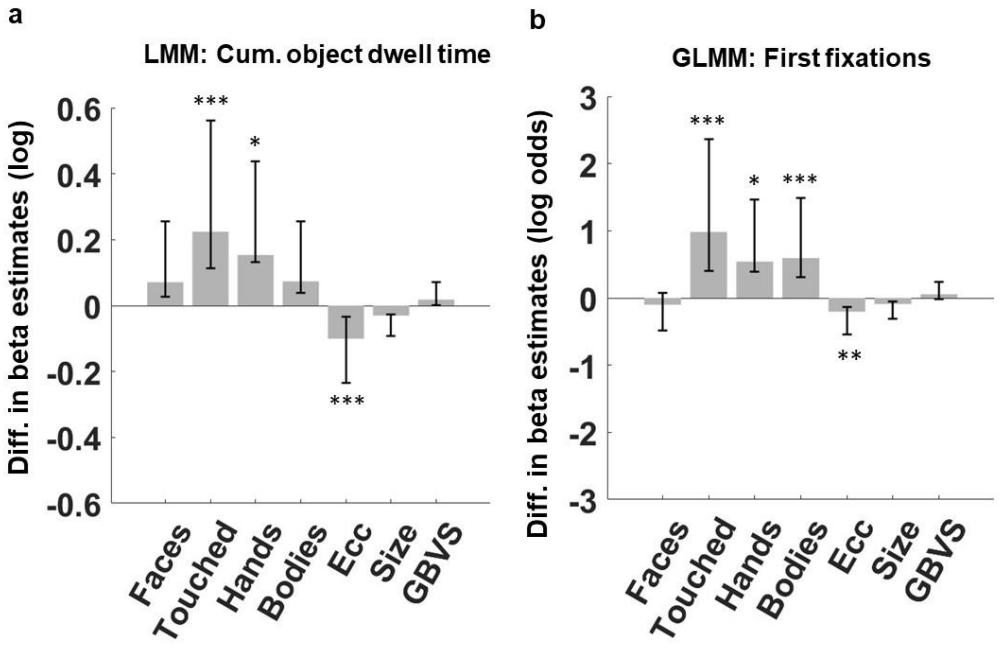

Figure S2. **Group differences in predicting object-directed fixations and dwell time without Text.** Bar plots depicting the beta estimates of a given object feature x Age interaction, which describe the difference in estimates between children and adults for a given object property. Panel a shows the interaction estimates for the LMM predicting object-directed dwell time and panel b depicts the binomial GLMM predicting first fixations towards objects after image onset across the included predictors Text, Faces, Touched, Hands and Bodies as well as Object-image centroid eccentricity (Ecc), (Object-) Size and graph-based visual salience (GBVS). Asterisks indicate the significance of a given group x predictor interaction. Results of model-based analyses that did not take into account the effect of text can be found in the supplementary materials. Error bars represent the 95% confidence intervals. *p < 0.05, **p < 0.01, ***p < 0.001.

In the analysis predicting object-directed dwell time, significant interactions were found between *Age* and *Touched* and *Age* and *Hands*, indicating that when not controlled for *Text*, children devoted more dwell time towards *Touched* objects, $b = 0.22$, $t = 3.87$, $p < 0.001$ and *Hands*, $b = 0.15$, $t = 2.29$, $p = 0.02$, compared to adults. Furthermore, a significant interaction between *Age* and *Eccentricity* showed a stronger central bias for children than adults, $b = -0.10$, $t = -5.94$, $p < 0.001$. No significant differences in estimates were found for *Faces*, $b = 0.07$, $t = 1.21$, $p = 0.23$. Finally, we found no significant interactions between *Age* and object *Size* and *Age* and *GBVS*, indicating no significant differences between children and adults for the predictors *Size*, $b = -0.03$, $t = 1.80$, $p = 0.07$, and *GBVS*, $b = 0.02$, $t = 1.04$, p = 0.30.

When predicting whether an object was fixated immediately after image onset (first fixation), results showed significant interactions of *Touched*, *Hands*, and *Body* with *Age*. These interactions indicate that the probability of an object being fixated immediately after image onset was higher for *Touched*, $b = 0.98$, $t = 4.79$, $p < 0.001$, *Hands*, $b = 0.54$, $t = 2.75$, $p = 0.01$ and *Body*, $b = 0.59$, $t = 3.82$, $p < 0.001$, in children compared to adults. Further, a significant interaction between *Age* and *Eccentricity*, $b = -0.20$, $t = -3.10$, $p = 0.002$ indicated a stronger center bias for first fixations in children relative to adults. Finally, we found no significant interactions between *Age* and *Face*, $b = -0.10$, $t = -0.70$, $p = 0.48$, object Size, $b = -0.09$, $t = 1.39$, $p = 0.16$, and *GBVS*, $b = 0.05$, $t = 0.86$, $p < 0.39$.

These findings are in line with the stronger tendency to fixate *Touched* objects, *Hands* and *Bodies* observed for children (see simple inference statistics, Fig. 3 & 4, main text). At the same time, they suggest that these differences are mostly driven by the drastically reduced gaze bias towards *Text* in children. Non-significant interactions between *Age* and *Faces* when predicting object-directed dwell time and first fixations are not line with our earlier findings and hence might be explained by the additional control predictors included here, e.g. the stronger center bias in children. Also note, that the GLMM controlling for *Text* showed a

stronger attraction of *Faces* for first fixations in adults, indicating that this effect is robust to low-level salience differences when controlling for the competing attraction to *Text* in adults.

Group difference in gaze towards objects with implied motion

Here we tested group differences in cumulative dwell time proportion towards objects with implied *Motion* using a permutation test. Fig S4 shows the distributions of cumulative dwell time proportion spent on objects with implied *Motion.*

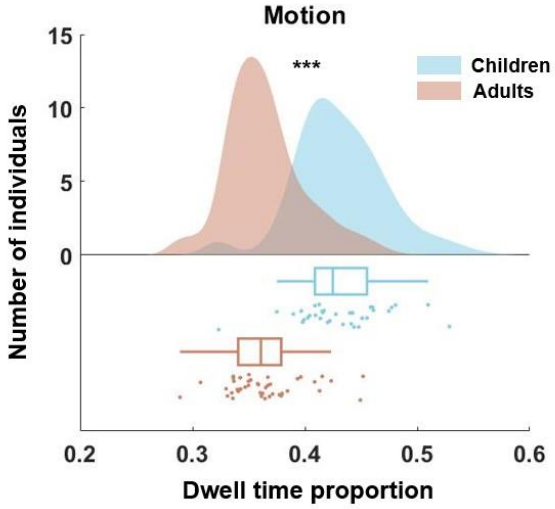

Figure S3. **Group differences in gaze behaviour towards objects with implied *Motion***. Density plot showing the probability distributions of cumulative dwell time towards objects of the dimension *Motion* for children (red) and adults (blue). Data points depicted below indicate the individual dwell time proportion for *Motion*. Corresponding box plots above the data points show an overview of the summary statistics for each group and semantic category; the vertical line within a box represents the mean value. The left side of a box indicates the 25th percentile and the right side the 75th percentile. The whiskers represent the minimum and maximum values. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Holm-Bonferroni corrected; see Methods)

The findings showed that children devoted a significantly larger amount of their dwell time towards objects with implied *Motion* compared to adults, $t_{obs} = 8.01$, $p < 0.001$. Note, for this analysis we calculated the dwell time proportion of fixations towards all objects, which are labelled *Motion.* As 92% of objects with implied *Motion* are also labelled *Hands, Bodies* or *Faces*, enhanced attentional biases towards objects of these dimensions in children may explain group differences shown for *Motion*.

Study 4

Linka, M., Karimpur, H., & de Haas, B. (2024, February 14). Protracted development of gaze behaviour. *Psyarxiv*

Protracted development of gaze behaviour

[1]Marcel Linka*, [1]Harun Karimpur, [1]Benjamin de Haas

[1]Experimental Psychology, Justus-Liebig Universität Gießen, Germany

***Corresponding author:**

Marcel Linka

Experimental Psychology

Justus-Liebig Universität Gießen

Otto-Behaghel Straße 10f

35394 Giessen

Germany

marcellinka54@gmail.com

Abstract

How does natural gaze behaviour develop? Here, we present data from > 6,500 subjects from 5-72 years of age, freely viewing 40 natural scenes. We find that the development of scene viewing is surprisingly protracted. Semantic salience for social features continuously changes until late adolescence and text salience increases until the third decade of life. Basic oculomotor biases towards the image centre and along the horizontal meridian develop until early adulthood, matching developmental changes in visual sensitivity and cortex. Finally, while the tendency for visual exploration continuously increases, fixation patterns become less idiosyncratic and more canonical throughout adolescence. These findings show that fundamental aspects of adult gaze take decades of continuous development and push individuals towards more canonical viewing patterns. We suggest that development is key to understanding the general mechanisms of active vision.

Introduction

Visual attention towards natural scenes is drawn to low-level features like luminance, colour, and orientation contrast[1,2], as well as to higher-level features like objects[3] and their meaning[4]. When predicting natural gaze behaviour, higher-level features, like Faces, Text, Touched objects and Taste[5,6] carry more weight than their lower-level counterparts[3,5]. At the same time, the attraction to specific semantic features varies across observers[7,8] and is influenced by genetic factors[9,10]. Adult active vision is further marked by spatial biases, like a tendency to fixate central image regions[11–13] and a predominance of horizontal saccades[14,15], along with a corresponding horizontal (and lower visual field) advantage in visual sensitivity[16–20].

A central question in visual perception is whether and how oculomotor biases and attentional preferences develop. Although gaze preferences for faces and social information are present in infants[21–25] and early childhood[26], it remains poorly understood how they evolve throughout the subsequent stages of development. Some semantic gaze preferences differ between preschoolers and adults, with an increase of text salience, at the expense of attention towards hands and objects being touched[27]. This shift in attention matches a competition for cortical resources between text and limb preferring patches in the ventral stream, which starts with the onset of reading and lasts into adolescence[27–29]. More basic patterns of oculomotor behaviour also seem to develop. Saccade amplitudes are higher and fixation durations are shorter at later stages of childhood compared to age 2[30] and the tendency to fixate on central image regions is weaker at age 8 than 4[31]. The horizontal bias seems to be present in infants, albeit likely to a lesser extent than in adults[32]. Overall, the understanding of gaze development is severely limited by a scarcity of data, with most studies comparing to two or three age groups and rarely covering adolescence.

At the same time, recent studies have demonstrated that visual field asymmetries typical for adults are comparatively diminished in both, children and adolescents.

Specifically, adults show a stronger enhancement of visual performance along the horizontal compared to the vertical meridian and for the lower compared to the upper vertical meridian[33,34]. The diminished lower visual field advantage in adolescents goes along with a more balanced V1 representation of the lower and upper vertical meridian in terms of surface area[35]. Moreover, recent work has shown that the success of salience model predictions for free-viewing can hinge on the age of participants[36]. Together, these studies suggest that basic aspects of visual processing continue to evolve beyond early childhood, and raise the possibility that this extends to oculomotor biases and attentional preferences. Recent findings on interindividual variability in adult gaze[7,8] further pose the question whether individual differences develop from a more canonical starting point, or *vice versa* diminish with age.

Here, we aimed to trace the developmental trajectories of semantic salience and basic oculomotor biases continuously and across a wide age range. In collaboration with a local science museum, we gathered gaze data from over 10,000 participants ranging from 5 to 72 years of age. Participants sat in an eyetracking booth and freely viewed 40 everyday scenes, which were carefully selected to maximise sensitivity for individual gaze biases[8] and contained hundreds of semantically annotated objects. To probe the development of semantic salience, we calculated the proportion of dwell time and of first fixations after image onset towards several semantic features, individually for age bins ranging from 5-6 to 71-72 years. We then fitted polynomial developmental trajectories to the resulting curves. Our dataset is particularly sensitive to developmental trends between 6-56 years, a period covered by over 100 observers per age bin (cf. Methods, Figure 5). Similarly, we traced the developmental trajectories of the central and horizontal bias, of visual exploration, fixation frequency, and of inter-observer similarities. We find strong evidence that the development of attentional preferences and basic oculomotor biases is protracted into adulthood and that viewing patterns become more canonical with age. The size of our dataset further allowed us to probe a

potential relationship between text salience and the horizontal bias independently of age, which turned out to be negligible.

Results

Protracted development of semantic salience

To trace the development of semantic salience, we calculated the proportion of dwell time and of first fixations towards four types of targets for each age bin: *Text*, *Faces*, *Touched* objects, and *Bodies*. Next, we applied polynomial curve fitting and the Akaike Information Criterion (AIC) to determine the best-fitting developmental trajectories. To differentiate periods of broad linear change from stability we further applied a sliding window to continuously assess the evidence for linear slopes across 5 contiguous age bins. Figure 1 shows the developmental trajectories of semantic salience for dwell time (A) and for first fixations (B), as well as pixel masks and the fixations of two age groups for an example image (C).

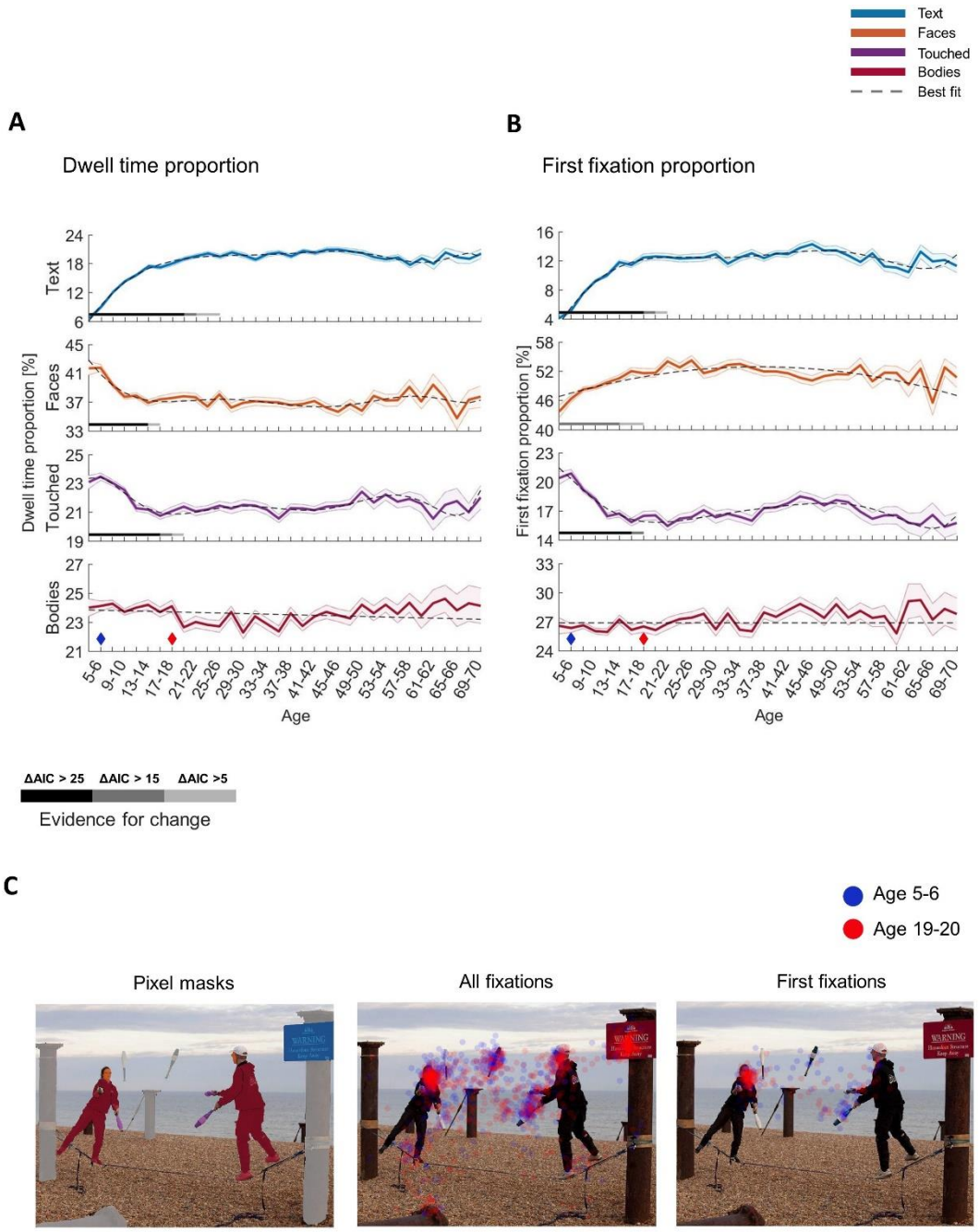

Figure 1. Protracted development of semantic salience. Line plots display the mean proportion of dwell time (A) and first fixations (B) directed towards four semantic features across age bins. Semantic categories are indicated by colours, as shown in the inset. Error shades indicate +/- 1 standard error of the mean (SEM). Dotted lines show the corresponding best fitting polynomials. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. main text and methods). The left-hand side of panel (C) displays object pixel masks for an example image, with colors corresponding to categories in A and B (cf. inset, objects outside these categories shown in light gray). The middle and right-hand images display alpha-blended fixation overlays for the same image, showing all fixations (middle) or only first fixations after image onset (right-hand side) for two age groups. Fixation overlays shown in blue and red correspond to ages 5-6 and 19-20, respectively. These age groups are marked correspondingly with blue and red rhombuses above the x-axes in Panels (A) and (B).

Text salience steeply and continuously increased, with dwell time and first fixation proportions on text more than doubling between ages 5-6 and ages 15-16 (from about 7-17% for dwell time and 4-12% for first fixations). This was followed by a shallower increase before finally stabilising in the twenties. The sliding-widow analysis yielded ΔAIC >5 for a positive slope until age 25-26 for dwell time and until age 21-22 for first fixations. The corresponding developmental trajectories were best described by 9^{th} and 6^{th} degree polynomials, respectively. These findings proved robust across images with different types of text elements and across participants who indicated German or English as their preferred language, suggesting they are largely independent of language proficiency (see supplementary materials).

Face salience showed a differentiated developmental pattern, with dwell time proportions continuously dropping between ages 5-6 and 15-16 (from 42% to 37%; ΔAIC for decrease >5), whereas the proportion of first fixations on faces *increased* between ages 5-6 and 17-18 (from 44% to 52%; ΔAIC for increase >5). The corresponding developmental trajectories were best fitted by 8^{th} and 2^{rd} degree polynomials, respectively.

The salience of *Touched* objects continuously decreased with dwell times dropping between ages 5-6 and 19-20 (from 23% to 21%; ΔAIC for decrease >5) and the proportion of first fixations dropping between ages 5-6 and 17-18 (from 20% to 16%; ΔAIC for decrease >15). Both developmental trajectories were best described by 7^{th} degree polynomials.

The salience of *Bodies* showed no clear developmental changes. and no significant periods of change in the sliding window analyses. Overall, the developmental trajectory of dwell time proportions was best fit by a linear model with a slight negative slope, while that of first fixations was bet fit by an intercept model.

Protracted development of the horizontal bias

Adults make far more saccades along the horizontal than the vertical[14,15], which matches anisotropies in sensitivity across the visual field[16,17,20]. Age-specific polar histograms

(Fig. 2A) and the developmental trajectory of the proportion of horizontal saccades (< 45 degrees from the horizontal meridian; Fig. 2B) revealed a strong and prolonged increase in the horizontal bias. The proportion of horizontal saccades increased from 57% to 62% from ages 5-6 to 15-16 (ΔAIC > 15) and the best fitting developmental trajectory (3rd degree polynomial) reached a plateau from about age 29-30.

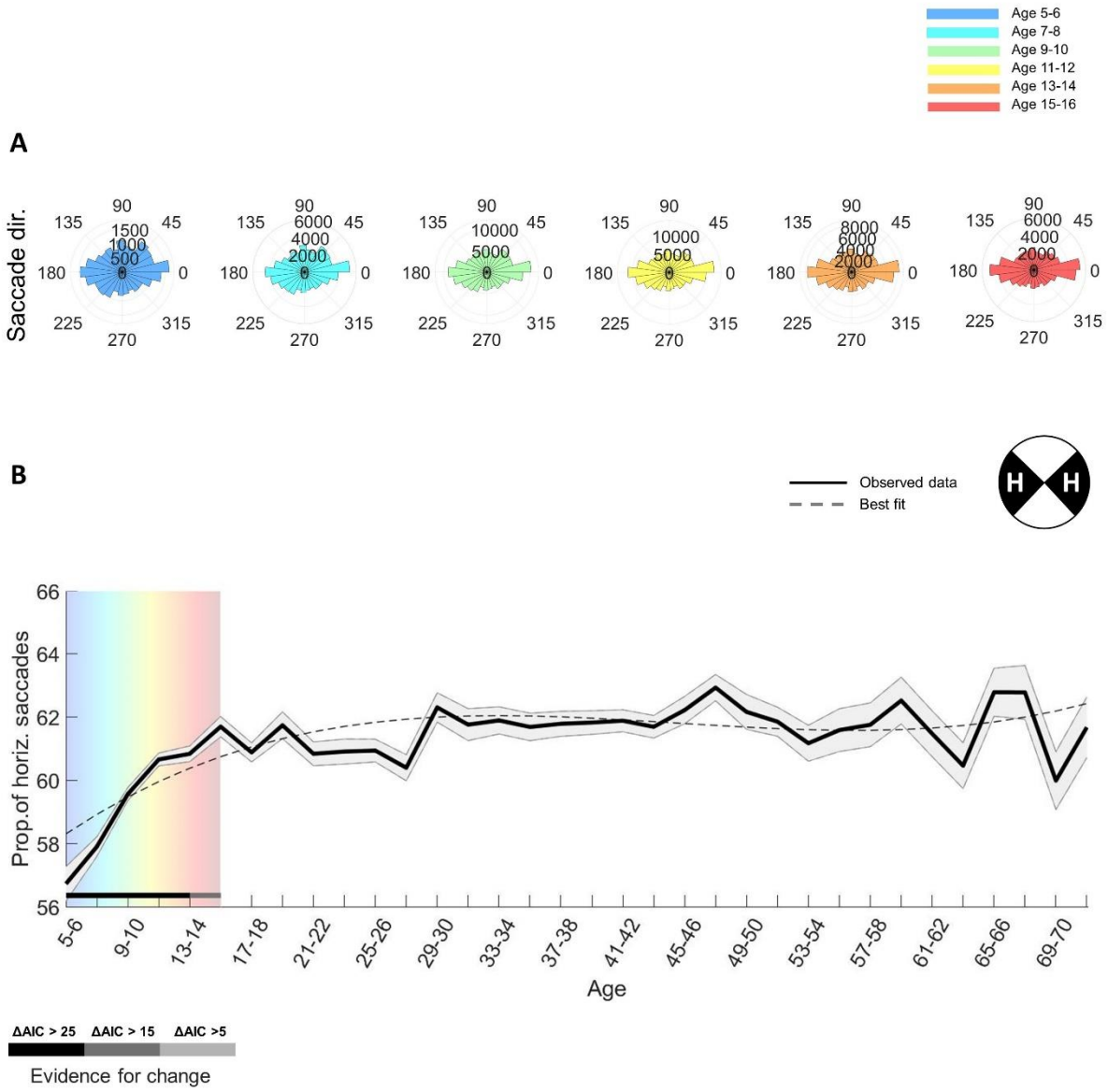

Figure 2. Protracted development of the horizontal bias. Panels A and B show polar saccade plots across 36 direction bins for ages 5-6 to 15-16. Age is indicated by color as shown in the inset. Panel C shows the average proportion of horizontal fixations across age bins +/- 1 SEM (solid line and error shade, respectively). The dotted line shows the best fitting polynomial. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. main text and methods).

The parallel protracted development of the horizontal bias and text salience raises the possibility of a causal connection between the two. Reading requires mostly horizontal saccades and this acquired behaviour may generalize to scene viewing. We first tested that the developmental pattern we observed for the horizontal bias generalizes to images without text elements, which could be confirmed (supplementary results). We then tested the partial correlation of individual differences in the horizontal bias and text salience when controlling for age, which was statistically significant but negligible ($r(6734) = .07$, $p < .001$, 95% CI [.05, .1]). Given that the normality assumption was met for the horizontal bias and text salience, but not for age, we conducted an additional control analysis testing the same partial correlation with log-transformed age. This confirmed a negligible correlation between the horizontal bias and text salience when controlling for age ($r(6734) = .04$, $p = .002$, 95% CI [.01, .06]). This is counter to the hypothesis of a strong causal relationship between the horizontal bias and individual text salience.

Protracted development of the center bias and of visual exploration

Next, we examined developmental changes in the center-bias, as indicated by the mean eccentricity of fixations from the image centroid (Figure 3A), as well as changes in visual exploration, as indicated by fixation frequency (Figure 3B) and the absolute number of objects fixated (Figure 3C).

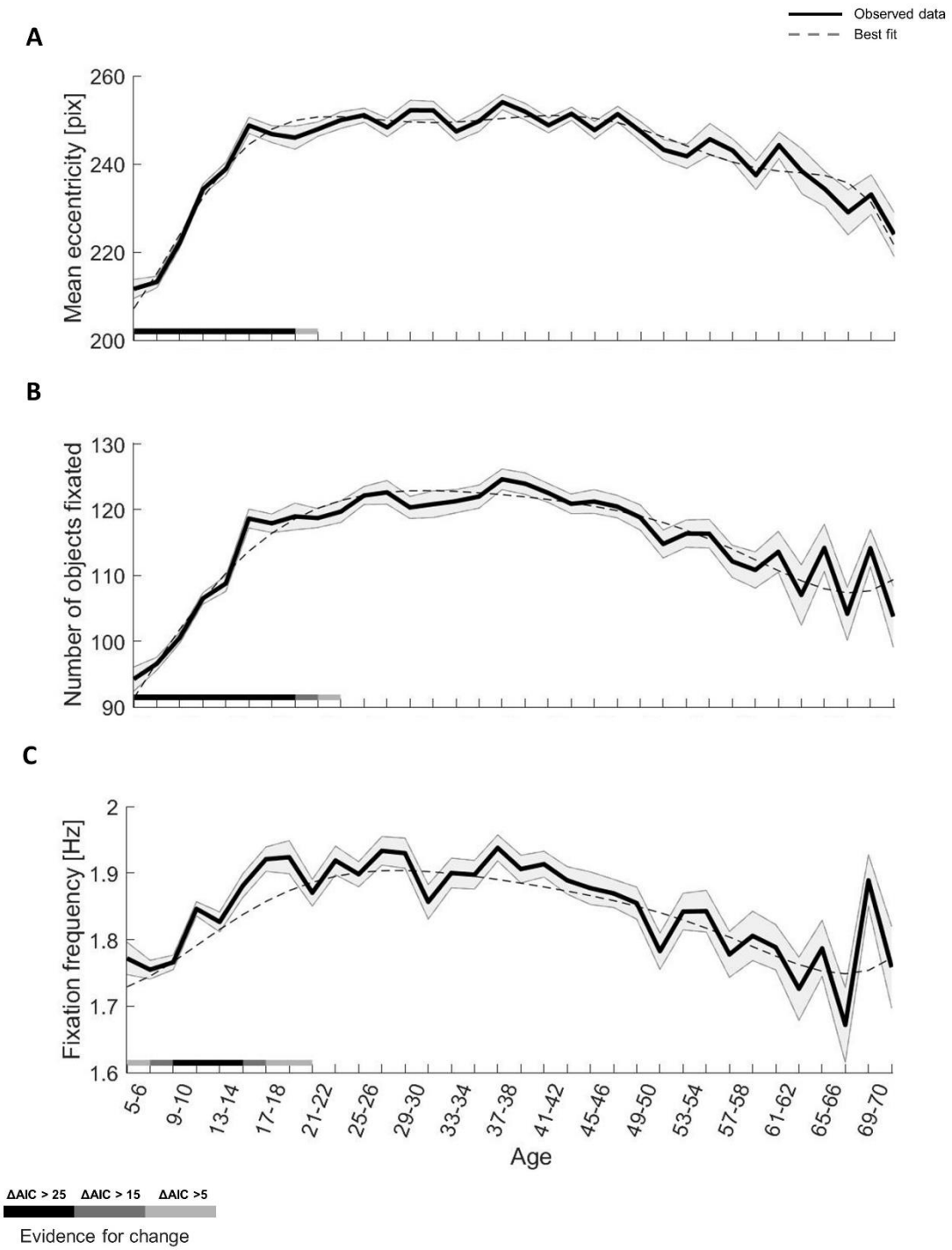

Figure 3. Protracted development of visual exploration. Line plots illustrate the mean distance to the image centroid (in pixels; A), the average number of objects fixated (B), and average fixation frequency (Hz; C) for age bins ranging from ages 5-6 to 71-72 (x-axis). Error shades indicate +/- 1 standard error of the mean (SEM). Dotted lines show the best fitting polynomials. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. main text and methods).

169 The development of average fixation eccentricity from the image center was best fitted

170 by a broadly U-shaped 7th degree polynomial. Average eccentricity increased between ages

5-6 and 19-20 average (from 212 to 246 pixels; ΔAIC >15) and slowly decreased from about age 49-50. The number of fixated objects across all images increased between ages 5-6 and 21-22 (from 94 to 119; ΔAIC > 5), while fixation frequency increased between ages 5-6 and 19-20 (from 1.77 to 1.92 Hz; ΔAIC > 5). Both indicators of visual exploration slowly decreased from the late thirties or early forties and followed developmental trajectories best fitted by broadly U-shaped 6th degree polynomials.

The sliding window analyses using a window size of five age bins yielded strong evidence for the increase in all indicators of visual exploration over the first two decades of life, but not for the later decline. However, a supplemental analysis using a larger window size of 10 bins confirmed significant change for both periods (supplemental results). This reflects the more protracted and gradual decline of visual exploration with older age, compared to the steeper increase early in life.

Protracted development of entropy and inter-observer similarity

Finally, we traced the development of fixation entropy and individual differences in gaze behavior. Adults show systematic individual differences in gaze behaviour towards complex natural scenes[7,8]. Here, we first computed the Shannon entropy of smoothed fixation maps for each image and age group. This group-level entropy showed a steep early increase, peaked at age 11-12, after which it decreased until age 25-26, followed by a second, more prolonged decrease from age 47-48. At age 71-72, group-level entropy was at its lowest, 34% below its peak. This developmental trajectory was best described by a 5th-degree polynomial. (Fig. 4 A, see markers above the x-axis for periods of significant linear change in the sliding-window analysis).

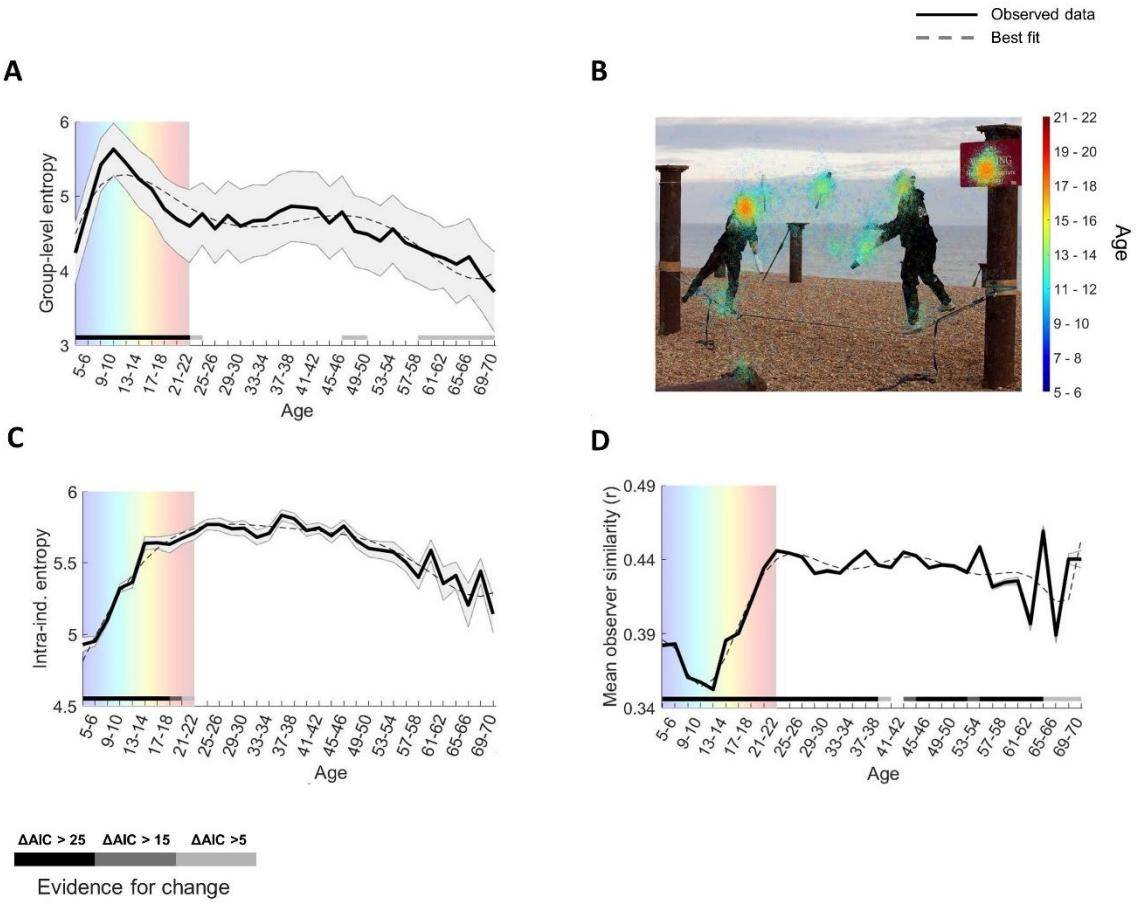

Figure 4. Protracted development of inter-observer similarity. Panel A shows the average entropy of group-level fixation maps across age bins. Panel B shows fixation data from observers aged 5-22 overlaid as colored circles on one example image. Colors correspond to age groups as indicated by the colorbar. Note the lower dispersion of hotter colors, indicating less dispersion on the group level for adult participants. Panel C shows the average entropy of observer-specific fixation maps across age bins. Panel D shows the mean observer similarity for each age bin, as indicated by the average pairwise correlation of observer-specific dwell-time distributions across objects. All error shades indicate +/- 1 standard error of the mean (SEM) and the dotted lines the best fitting polynomial. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. main text and methods).

Group-level entropy reflects both, intraindividual fixation dispersion and interindividual differences. To disentangle the two, we first computed the intraindividual entropy of individual fixation maps and averaged it for each age group (Fig. 4 C). Echoing our findings on visual exploration, intraindividual entropy first increased until age 37-38, before the best fitting 6^{th}-degree polynomial steadily declined again. Strong evidence of linear change in the sliding window analysis ($\Delta AIC > 5$) was limited to the steepest phase of the

initial rise between ages 5-6 and 23-24, but extended to the steady decline when using a larger window size of 10 bins (supplemental results). Just as the other indicators of individual visual exploration, this reflects the more protracted and gradual decline in older age, compared to the steep increase early in life (see above).

Finally, to pinpoint the development of *inter*individual similarity specifically, we directly compared pairs of observers regarding their individual distributions of dwell time across objects in the images. Specifically, we quantified the average pair-wise correlation of object fixation patterns for each age bin. As shown in Fig 4 D, pair-wise observer similarities slightly decreased from $r = .38$ to $r = .35$ between ages 5-6 and 13-14, before steeply increasing to $r = .45$ at age 23-24 ($\Delta AIC > 25$) onwards. This developmental trajectory was best described by a 9^{th} degree polynomial.

Taken together, visual exploration and intraindividual fixation dispersion steeply increased over the first two decades of life. This increase of individual exploration went along with a small initial drop and subsequent rise of observer similarity during adolescence. That is, from age 13-14, the distribution of gaze became both, more dispersed on the individual level and more canonical across observers.

Discussion

To probe the developmental trajectory of free-viewing behaviour towards natural scenes, we analysed the gaze of thousands of individuals from 5 to 72 years of age. We found surprisingly protracted development for the semantic salience of text, faces, and touched objects, as well as for visual exploration and basic oculomotor biases, such as the central and horizontal bias, all of which continuously developed into adulthood. This suggests that typical adult gaze behaviour takes decades to unfold and may require prolonged visual experience. At the same time, the inter-individual similarity of gaze patterns steeply rises during adolescence and reaches its maximum at age 22-23 onwards. This suggests that the protracted development of gaze ultimately results in more canonical ways of exploring our surroundings.

Protracted development of semantic salience

We found a continuous increase in the proportions of dwell time and first fixations directed towards text, which only plateaued in the third decade of life. This matches previous results showing drastically enhanced text salience in adults compared to pre-schoolers[27] and reveals that the development of text salience extends far beyond reading acquisition. Control analyses of a subsample indicating English as preferred language or using an image subset with text elements common in German suggest that this effect is unrelated to language proficiency (see supplementary materials). Adult text salience appears to unfold across most of the educational history and may be mediated by levels of literacy and reading experience. An increasing share of text in the foveal diet may have a self-reinforcing effect on text salience. Future work could test this hypothesis directly, using literacy tests and mobile eye-tracking to estimate changes in the visual diet and in gaze behaviour longitudinally. Similarly, longitudinal studies combining functional neuroimaging and eyetracking could probe a potential bidirectional relationship between the development of text salience and word-preferring patches of cortex in the ventral stream[7,27,29]. Cross-sectional and longitudinal studies have shown that word-preferring ventral regions develop from reading acquisition until at least the age of 17[28,37,38].

The salience of objects being touched continuously decreased until early adulthood. We speculate that this may be related to an increasing understanding of (inter-)actions among people and objects. Accumulated experience with scenes and actions (e.g., two people playing basketball) may enable adults to categorise a given action from scene gist[40] whereas children and adolescents may have a greater need to foveate hands and the things they touch for action understanding c.f.[41]. This hypothesis could be tested in future studies comparing action understanding from scene gist across different age groups and levels of expertise. Interestingly, a recent longitudinal study revealed that after reading onset, limb-preferring cortical regions shift their preference towards text (and faces). This shift lasted until ages 15-

17, suggesting a prolonged process of cortical recycling[28,29], which may be related to corresponding changes in visual diet. The longitudinal neuroimaging studies suggested above could track the push-pull relationship between preferences for hands and text in both, gaze behaviour and cortical tuning to test whether they happen in tandem or one is preceding the other.

Face salience continuously developed over the first two decades of life, a relevant finding for a longstanding debate focussing on its potentially innate nature and very early development[42,43]. Interestingly, face salience showed a differential developmental pattern, with a protracted *de*crease of dwell time on faces and a parallel *in*crease of first fixations directed towards faces. This finding is consistent with previous work comparing pre-schoolers and adults[27]. While the individual proportions of first fixations and dwell time on faces were highly correlated with each other in all age groups, their opposing developmental trajectories may point to dissociable contributing mechanisms. The tendency to saccade towards faces immediately after image onset appears stronger and more mature in adults and may primarily be driven bottom-up. A second, top-down mechanism may guide dwell time towards faces and suffer stronger competition with text salience and visual exploration in adults. Notably, the proportion of adult dwell time on text is considerably higher than the proportion of first fixations on text (~20% and ~12%, respectively), indicating the stronger competition of text for dwell time than for first saccades. As mentioned above, recent developmental studies suggest that ventral cortical territory selective for limbs may be recycled for both, text and face preferences[28,29]. Future longitudinal research could track and test a relationship between the development of face salience and 'ultrarapid' saccades towards faces[44–46], which presumably present bottom-up mechanisms. Future studies may also test whether the neural signature of immediate and short latency saccades towards faces is different from that of later and more prolonged fixations towards faces.

For body salience (beyond faces and hands), we found a very subtle negative trend for dwell time across the whole age range tested, but no strong developmental changes. This matches studies finding no changes in cortical activation for bodies along the ventral stream throughout adolescence and young adulthood[28,47], underscoring the dissociable and special nature of faces and hands for the visual system and its development.

Protracted development of visual exploration and saccadic anisotropies

Previous research has indicated a more pronounced central bias in infants and children aged 4-8[31] compared to adults, as well as a less prominent horizontal bias in infants[32]. Here, we found that visuo-spatial biases continuously develop until adulthood. Specifically, we observed a continuous reduction of the central bias until age 20-21 and an increase in the horizontal bias until age 15-16. Given that reading is characterized by horizontal saccades, we explored a potential link between text salience and the horizontal bias, which could not be confirmed when controlling for age. This suggests that the visual processes underlying enhanced text salience and the horizontal bias develop in parallel but may be distinct from each other. The protracted changes in saccadic anisotropies match recent findings, showing that children and adolescents have a lower sensitivity advantage for the horizontal meridian compared to adults[33,34]. Again, future longitudinal studies could disentangle the relationship between the protracted development of this horizontal sensitivity advantage and of the saccadic horizontal bias. Furthermore, the development of visual exploration and the horizontal bias may be due to the prolonged formation of spatial priors for typical scene layouts[14]. Additionally, the visual field distribution of informative features may vary with age. Small children probably encounter relevant stimuli in their upper visual field more often than adults[34].

Protracted development of inter-observer similarity

Intraindividual fixation dispersion consistently increased until at least age 21-22. Initially, this was accompanied by a slight increase of individual differences, both in terms of

object directed dwell time distributions and group-level entropy. However, from age 13-14 this trend reversed and inter-individual similarity steeply increased. Visual exploration became increasingly canonical, despite an ongoing trend for enhanced intra-individual fixation dispersion. It is remarkable that children show larger inter-individual differences despite their lower tendency for exploration stronger central bias. This suggests children explore fewer parts of a scene, but which parts they select is highly individual. Adult gaze is more widely dispersed, but in a comparatively canonical way. This more canonical nature of adult gaze may reflect a slow accumulation of expertise for everyday scenes, which may in turn reflect increasingly similar visual experience. Individual visual diets may become less idiosyncratic and richer from the onset of out-of-home care, including a larger number of faces and places encountered[48]. Future studies could use mobile eyetracking to test whether the intraindividual variance of visual diets increases with age, while the *inter*-individual variance decreases.

Taken together, the data of thousands of individuals from ages 5-72 show that sematic salience, visual exploration and spatial saccadic biases unfold over at least the first two decades of life. The data further suggest that a crucial result of this protracted development is increased interobserver similarity of viewing patterns. Many of the underlying biological and environmental causes are unclear. Studying them appears key for understanding the general mechanisms of gaze.

Methods

Subjects

From February 2022 to November 2023 a total of 13984 datasets were collected at the German science museum Mathematikum, Giessen. 10632 of these stemmed from individual participants, of whom 10624 proceeded to the free viewing task (cf. Figure 6 B). 10.7% of these were excluded due to incomplete data for more than half of the 40 presented images and a further 28% because their average calibration error was above 1 degree visual angle. This

resulted in a final sample of 6843 subjects, of which 6736 subjects were 5 to 72 years of age (M = 25.78, SD = 18.30; See Fig. 5). 55.43% of this included sample identified as male, 41.09% as female, and 3.48% as having diverse gender identities. The study was approved by the Justus Liebig Universität Fb06 Local Ethics Committee (lokale Ethik-Kommission des Fachbereichs 06 der Justus Liebig Universität Giessen) and participants provided informed consent vis a touchscreen.

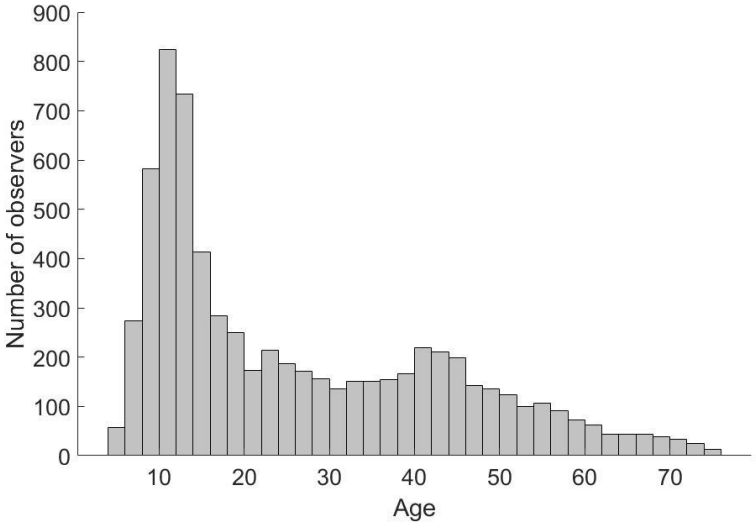

Figure 5. Age distribution of analysed free viewing data. Histogram displaying the distribution of ages of included observers using a bin width of two.

Apparatus

The free viewing task was programmed with Psychopy version 2020.1.2[49] in Python 3.6.10[50]. Stimuli were displayed on a Iiyama ProLite T1931SR-B5 touchscreen with a 1280 × 1024 pixel screen resolution and at a size of 1000 × 750 pixels, which approximately corresponds to a visual angle of 37 × 28 degrees (d.v.a.). Eye movements were recorded from both eyes using a head-free Tobii 4c Eye Tracker (Tobii AB, Danderyd, Sweden) operating in remote mode at a distance of 50–90 cm and with a sampling rate of 90 Hz. Previous research assessing the predecessor model, Tobii EyeX, reported an accuracy of < 0.6 d.v.a. and a precision of < 0.25 d.v.a. Practice trials confirmed comparable performance for the 4c Eye Tracker and the median validation accuracy for the present study was 0.6 d.v.a. (SD = 2.88).

Procedure

Mathematikum visitors took part in in our study from February 2022 to November 2023. The software interface was designed to enable participants to navigate the menu without external assistance. Panel A in Figure 6 displays an image of the setup, panel B displays the order of events during the experimental procedure. For a more detailed description, please refer to Supplementary Figure S1.

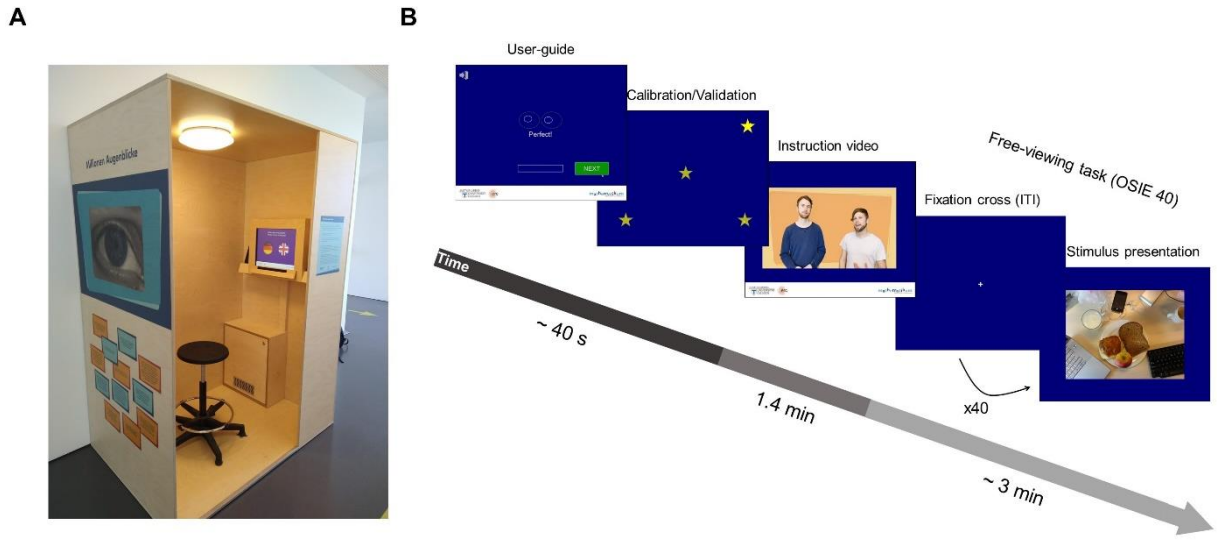

Figure 6. Eyetracking booth and procedure. Panel A shows an image of the eyetracking booth exhibited at the museum Mathematikum. Panel B provides an overview of the experimental procedure. For a more detailed description of the software, please refer to Supplementary Figure S1.

Participants seated themselves in front of a touchscreen and initiated the process by selecting their preferred language (German or English). Subsequently, they provided informed consent, indicated their age, gender, prior participation (y/n), and whether they wore glasses (y/n). Following this, participants underwent a 5-point calibration and validation procedure. If the average validation error was above 0.7 degrees visual angle (d.v.a.), the procedure was repeated once more. To ensure correct positioning, a customized position guide interfaced with the eye tracker and provided real-time feedback on whether the participant was within the trackable range. On average, participants sat at a distance of 59 cm from the screen. All

viewing angles for event detections and error margins for AOI analyses (see below) were computed based on the individual average distance.

Subsequently, participants viewed an instructional video in which they were welcomed to the exhibit and instructed to 'look at the images in any way they want'. Then, 40 images were displayed at the center of the screen for 3 seconds each, with a fixation cross displayed centrally between trials. A trial could only commence if a participant's gaze did not deviate by more than 2 d.v.a. from the fixation cross for 1 second. The presentation order of images was constant for all participants.

Stimuli

We used the OSIE40[8] dataset, which includes 40 everyday scenes and corresponding pixel masks for 364 objects with binary labels for 12 semantic object features. Additionally, we used pixel masks for bodies recently published by[51]. Individual gaze tendencies for several semantic features can be estimated reliably with this stimulus set[8].

Data Analyses

Data processing. All pre-processing steps and statistical analyses were performed using MATLAB R2019B (MathWorks). Raw eye tracking data were converted to fixation and saccadic events using the event detection algorithm NH2010 by Nyström and Holmqvist[52]. The NH2010 was specified to identify fixations using a threshold on minimum fixation duration of 100 ms and a maximum velocity criterion of 1,000º/sec as well as a minimum saccade duration of 11.11 ms (please refer to Figure S4 for example trace plots of (raw) eye tracking data). For a given trial, we removed fixations below 100 ms onset time to exclude intertrial fixations and saccades initiated before trial onset.

Statistical approach

To examine the developmental trajectory of a given measure across age groups, we fitted polynomials from 0-10 degrees to each curve. Among these candidate models, we then computed the AIC differences as $\Delta i = AICi - AICmin$, where $AICi$ stands for the AIC of the i th

model, and $AICmin$ represents the minimum AIC among all models. Subsequently, we selected the least complex model with $\Delta i < 4$ as the best-fitting choice. This approach helped us to trace broad developmental trends. To pinpoint periods of developmental change more specifically, we additionally used a sliding window approach. Here, we computed the AIC difference (ΔAIC) between first- and zero-degree polynomials, that is the evidence for linear change over a constant for each window of 5 consecutive age bins. We then computed the mean evidence for change for all windows containing a given age bin and projected the value onto this bin. This way we evaluated the evidence for continuous linear change in a specific but smooth fashion. To capture broader and less steep developmental changes, we report the results of the same analysis using a window size of 10 consecutive age bins in the supplement.

Semantic salience. To quantify semantic salience, we used two metrics of gaze towards objects with a given semantic feature: (1) the proportion of dwell time and (2) the proportion of first fixations after image onset among all fixations landing on labelled objects[7]. These were computed for each individual and then averaged for each age bin.

Oculomotor behavior across development. To examine oculomotor behavior across age groups, we calculated the following metrics for each observer: (1) Mean fixation eccentricity from the image centroid (in pixels); (2) visual exploration, indicated as the mean number of objects fixated and (3) the mean fixation frequency in Hz. Additionally, (4) we investigated the distribution of saccades across 60 directional bins and the proportion of saccades within ±45° of the horizontal meridian. Again, we averaged these metrics separately for each age bin to trace developmental trajectories.

Inter-observer similarity. To test intra-individual fixation entropy, we computed the mean Shannon entropy of observer- and image-specific fixation maps and averaged the resulting values for each age bin. Similarly, we computed the *group-level* entropy based on the Shannon-entropy of joint fixation maps for all members of a given age group (averaged across images). Finally, to determine (dis-)similarities in object-directed gaze, we first

compiled the cumulative dwell time devoted to each labelled object across all scenes for every observer (a vector with 366 entries per observer). We then computed all pairwise Pearson correlations between these individual gaze patterns in a given age bin and used the average of these correlations to determine typical pairwise similarity within the age bin.

Data Availability

Anonymized data are freely available at osf.io/ycrav.

Code Availability

Code to reproduce the presented findings are freely available at osf.io/ycrav.

Acknowledgements

ML and BdH are supported by ERC Starting Grant 852885 INDIVISUAL. BdH is further supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 222641018–SFB/TRR 135 TP C9 and by "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. Finally, we would like to thank Xu et al. as well as Broda & de Haas for making their stimuli, pixel masks and metadata publicly available.

Contributions

M.L. and B.d.H. designed and implemented the experiment in collaboration with H.K. M.L. analysed the data and prepared the initial draft of the manuscript in collaboration with B.d.H, who administered and supervised the project. All authors provided comments and/or revisions and approved the final article.

References

1. Itti, L. & Koch, C. A saliency-based mechanism for overt and covert shifts of visual attention. Vision Res. 40, 1489-1506. *Vision Res* **40**, 1489–1506 (2000).

2. Harel, J., Koch, C. & Perona, P. *Graph-Based Visual Saliency*. *Adv. Neural Inform. Process. Syst* vol. 19 (2006).

3. Stoll, J., Thrun, M., Nuthmann, A. & Einhäuser, W. Overt attention in natural scenes: Objects dominate features. *Vision Res* **107**, 36–48 (2015).

4. Henderson, J. M. & Hayes, T. R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat Hum Behav* **1**, 743–747 (2017).

5. Xu, J., Jiang, M., Wang, S., Kankanhalli, M. & Zhao, Q. Predicting human gaze beyond pixels. *J Vis* **14**, (2014).

6. Wang, H.-C. & Pomplun, M. The attraction of visual attention to texts in real-world scenes. *J Vis* **12**, (2012).

7. de Haas, B., Iakovidis, A., Schwarzkopf, D. & Gegenfurtner, K. *Individual differences in visual salience vary along semantic dimensions*. (2018). doi:10.1101/444257.

8. Linka, M. & de Haas, B. OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *J Vis* **20**, 13 (2020).

9. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344 (2017).

10. Kennedy, D. P. *et al.* Genetic Influence on Eye Movements to Complex Scenes at Short Timescales. *Current Biology* **27**, 3554-3560.e3 (2017).

11. Tatler, B. W., Baddeley, R. J. & Gilchrist, I. D. Visual correlates of fixation selection: effects of scale and time. *Vision Res* **45**, 643–659 (2005).

12. Ruddock, K. H., Wooding, D. S. & Mannan, S. K. The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spat Vis* **10**, 165–188 (1996).

13. Parkhurst, D. J. & Niebur, E. Scene content selected by active vision. *Spat Vis* **16**, 125–154 (2003).

14. Foulsham, T., Kingstone, A. & Underwood, G. Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Res* **48**, 1777–1790 (2008).

15. Tatler, B. W. & Vincent, B. T. Systematic tendencies in scene viewing. *J Eye Mov Res* **2**, (2008).

16. Talgar, C. P. & Carrasco, M. Vertical meridian asymmetry in spatial resolution: Visual and attentional factors. *Psychon Bull Rev* **9**, 714–722 (2002).

17. Barbot, A., Xue, S. & Carrasco, M. Asymmetries in visual acuity around the visual field. *J Vis* **21**, 2 (2021).

18. Corbett, J. E. & Carrasco, M. Visual Performance Fields: Frames of Reference. *PLoS One* **6**, e24470- (2011).

19. Cameron, E. L., Tai, J. C. & Carrasco, M. Covert attention affects the psychometric function of contrast sensitivity. *Vision Res* **42**, 949–967 (2002).

20. Carrasco, M., P.Talgar, C. & Cameron, E. L. Characterizing visual performance fields: effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spat Vis* **15**, 61–75 (2001).

21. Johnson, M. H., Senju, A. & Tomalski, P. The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neurosci Biobehav Rev* **50**, 169–179 (2015).

22. Andravizou, A., Gliga, T., Elsabbagh, M. & Johnson, M. Faces Attract Infants' Attention in Complex Displays. *Ann Gen Psychiatry* **14**, (2008).

23. Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **40**, 1–19 (1991).

24. Gluckman, M. & Johnson, S. Attentional capture by social stimuli in young infants. *Front Psychol* **4**, 527 (2013).

25. Kelly, D., Duarte, S., Meary, D., Bindemann, M. & Pascalis, O. Infants rapidly detect human faces in complex naturalistic visual scenes. *Dev Sci* **22**, (2019).

26. Frank, M., Vul, E. & Saxe, R. Measuring the Development of Social Attention Using Free-Viewing. *Infancy* **17**, (2011).

27. Linka, M., Sensoy, Ö., Karimpur, H., Schwarzer, G. & de Haas, B. Free viewing biases for complex scenes in preschoolers and adults. *Sci Rep* **13**, 11803 (2023).

28. Nordt, M. *et al.* Cortical recycling in high-level visual cortex during childhood development. *Nat Hum Behav* **5**, 1686–1697 (2021).

29. Kubota, E., Grill-Spector, K. & Nordt, M. Rethinking cortical recycling in ventral temporal cortex. *Trends Cogn Sci* (2023) doi:https://doi.org/10.1016/j.tics.2023.09.006.

30. Helo, A., Pannasch, S., Sirri, L. & Rämä, P. The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Res* **103**, 83–91 (2014).

31. Krishna, O., Helo, A., Rämä, P. & Aizawa, K. Gaze distribution analysis and saliency prediction across age groups. *PLoS One* **13**, e0193149- (2018).

32. Van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E. J. & Visser, I. Do infants have the horizontal bias? *Infant Behav Dev* **44**, 38–48 (2016).

33. Carrasco, M., Myers, C. & Roberts, M. Visual field asymmetries vary between adolescents and adults. *bioRxiv* 2023.03.04.531124 (2023) doi:10.1101/2023.03.04.531124.

34. Carrasco, M., Roberts, M., Myers, C. & Shukla, L. Visual field asymmetries vary between children and adults. *Curr Biol* **32**, R509–R510 (2022).

35. Himmelberg, M. M. *et al.* Comparing retinotopic maps of children and adults reveals a late-stage change in how V1 samples the visual field. *Nat Commun* **14**, 1561 (2023).

36. Strauch, C. *et al.* Saliency models perform best for women's and young adults' fixations. *Communications Psychology* **1**, 34 (2023).

37. Dehaene-Lambertz, G., Monzalvo, K. & Dehaene, S. The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biol* **16**, e2004103- (2018).

38. Saygin, Z. M. *et al.* Connectivity precedes function in the development of the visual word form area. *Nat Neurosci* **19**, 1250–1255 (2016).

39. Shi Chen and Ming Jiang and Qi Zhao. What Do Deep Saliency Models Learn about Visual Attention? Preprint at (2023).

40. Zelinsky, G. Understanding scene understanding. *Front Psychol* **4**, (2013).

41. Vanmarcke, S. *et al.* Gist perception in adolescents with and without ASD: Ultra-rapid categorization of meaningful real-life scenes. *Res Autism Spectr Disord* **29–30**, 30–47 (2016).

42. Simion, F. & Di Giorgio, E. Face Perception and Processing in Early Infancy: Inborn Predispositions and Developmental Changes. *Front Psychol* **6**, (2015).

43. Wilkinson, N., Paikan, A., Gredebäck, G., Rea, F. & Metta, G. Staring us in the face? An embodied theory of innate face preference. *Dev Sci* **17**, 809–825 (2014).

44. Crouzet, S., Kirchner, H. & Thorpe, S. Fast saccades toward faces: Face detection in just 100 ms. *J Vis* **10**, 16.1-17 (2010).

45. Vanmarcke, S. & Wagemans, J. The influence of age and gender on ultra-rapid categorization. *Vis cogn* **23**, 894–916 (2015).

46. Geringswald, F., Afyouni, A., Noblet, C. & Grosbras, M.-H. Inhibiting saccades to a social stimulus: a developmental study. *Sci Rep* **10**, 4615 (2020).

47. Peelen, M. V, Glaser, B., Vuilleumier, P. & Eliez, S. Differential development of selectivity for faces and bodies in the fusiform gyrus. *Dev Sci* **12**, F16–F25 (2009).

48. Jayaraman, S. & Smith, L. B. Faces in early visual environments are persistent not just frequent. *Vision Res* **157**, 213–221 (2019).

49. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav Res Methods* **51**, (2019).

50. van Rossum, G. & Drake, F. L. Python 3 Reference Manual. in (2009).

51. Broda, M. D. & de Haas, B. Individual differences in looking at persons in scenes. *J Vis* **22**, 9 (2022).

52. Holmqvist, K. *et al.* Eye tracking: empirical foundations for a minimal reporting guideline. *Behav Res Methods* (2022) doi:10.3758/s13428-021-01762-8.

Supplementary Materials

Supplementary Methods

Procedure

Figure S1 gives an overview of all phases of the experimental procedure. Participants took their seats in front of a touchscreen panel and selected their preferred language (either German or English). Following this, they provided informed consent and indicated their age, gender, whether they had participated before (yes or no), and whether they wore glasses (yes or no). To ensure accurate positioning, a customized position guide offered real-time feedback to confirm that the participant was within the trackable range. Subsequently, participants completed a 5-point calibration and validation procedure. If the average validation accuracy was less than 0.7 degrees of visual angle (d.v.a.), the procedure was repeated once more. After this, participants watched an instructional video welcoming them to the exhibit and instructing them to 'look at the images in any way they want'. Following these instructions, 41 images were displayed sequentially, each presented at the center of the screen for 3 seconds, with a central fixation cross displayed between trials. The image set consisted of the OSIE40[1], which was the basis for all analyses and an additional image of the museum which was used for gaze illustration later (see below). A trial would not commence before the participant's gaze was within 2 d.v.a. from the fixation cross for 1 second. The presentation order of images remained the same for all participants. Upon completion of the free viewing task, participants saw a heatmap of their personal gaze on the image of the museum and information about how their average fixation duration compared to that of others. Finally, they saw a display with a certificate of participation. After that they could chose to engage in another 'eyetracking game' or end participation.

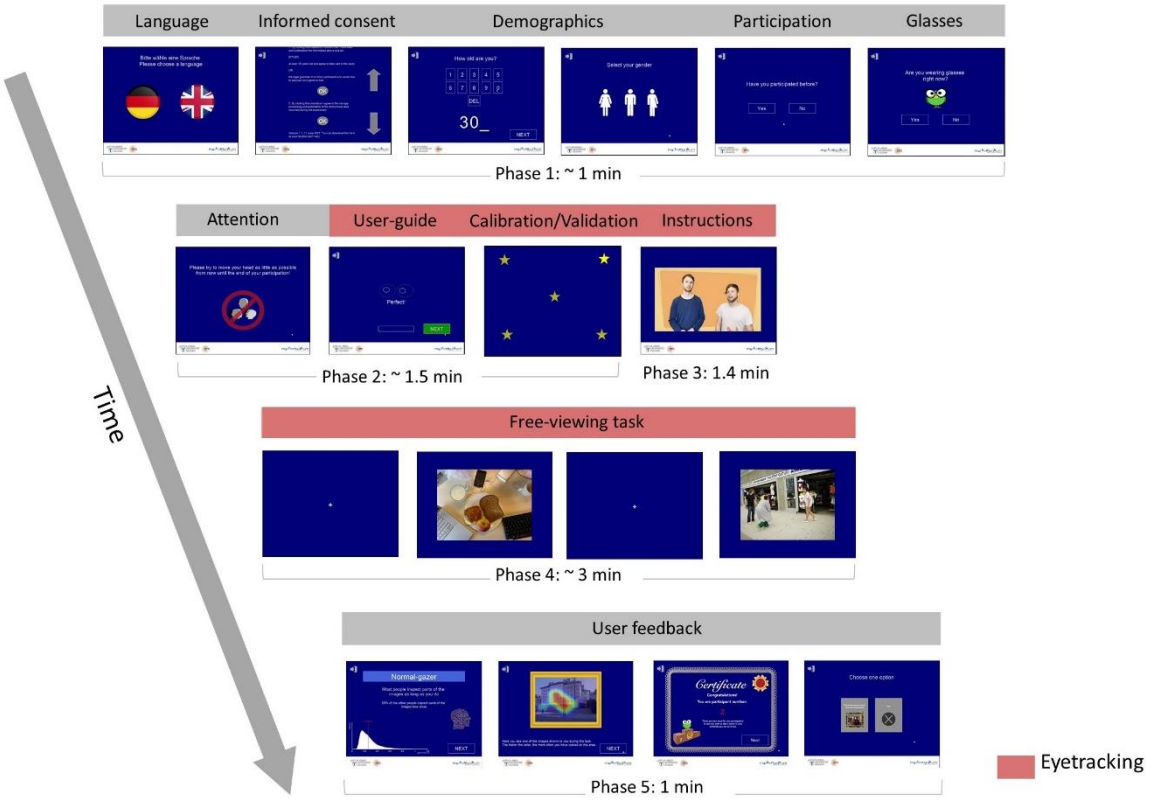

Figure S1. Phases of experimental procedure. Each frame shows a screenshot from a different phase of the experiment. Red bars indicate phases during which eye-tracking data was recorded. Feedback screens were based on online analyses of data and participant-specific.

Protracted development of text salience independent of English proficiency

To test whether the protracted development of text salience in our mostly German-speaking sample hinges on (lacking) English language proficiency, we conducted two additional analyses. First, we selected the subsample of participants who chose English as their preferred language (n = 412) and determined their dwell time proportion towards text across age bins. Next, we computed the dwell time proportion towards text in the full sample, but limited it to two images that included words and numbers common in both languages. As in the main analysis, we computed the corresponding averages for each age bin, fitted polynomials to the resulting developmental trajectory and adopted the sliding window approach to pinpoint periods of change (see main methods).

Protracted development of horizontal bias in images without text

To test whether the protracted development of the horizontal bias is tied to text salience, we replicated our analyses of saccadic directions for the subset of 17 images that contained no text elements (see main methods).

Protracted development of visual exploration using a larger window size

To probe a more gradual decline of visual exploration with age, we again used a sliding window approach to test periods of broad linear change for the center bias, number of objects fixated, fixation frequency and intra-individual entropy across age, however this time using a window size of 10.

Supplementary Results

Protracted development of text salience independent of English proficiency

Figure S2 shows the developmental trajectory of dwell time on text (1) when only including images containing text elements common in the German and English language and (2) when testing the English-preferring subsample. Both analyses replicated strong evidence for a strong and protracted increase of text salience until age 17-18 (ΔAIC >5). The development of text salience for the two images with numbers and words common in German was best fitted by a 4th degree polynomial. The development of text salience in participants indicating English as their preferred language by a 2nd degree polynomial.

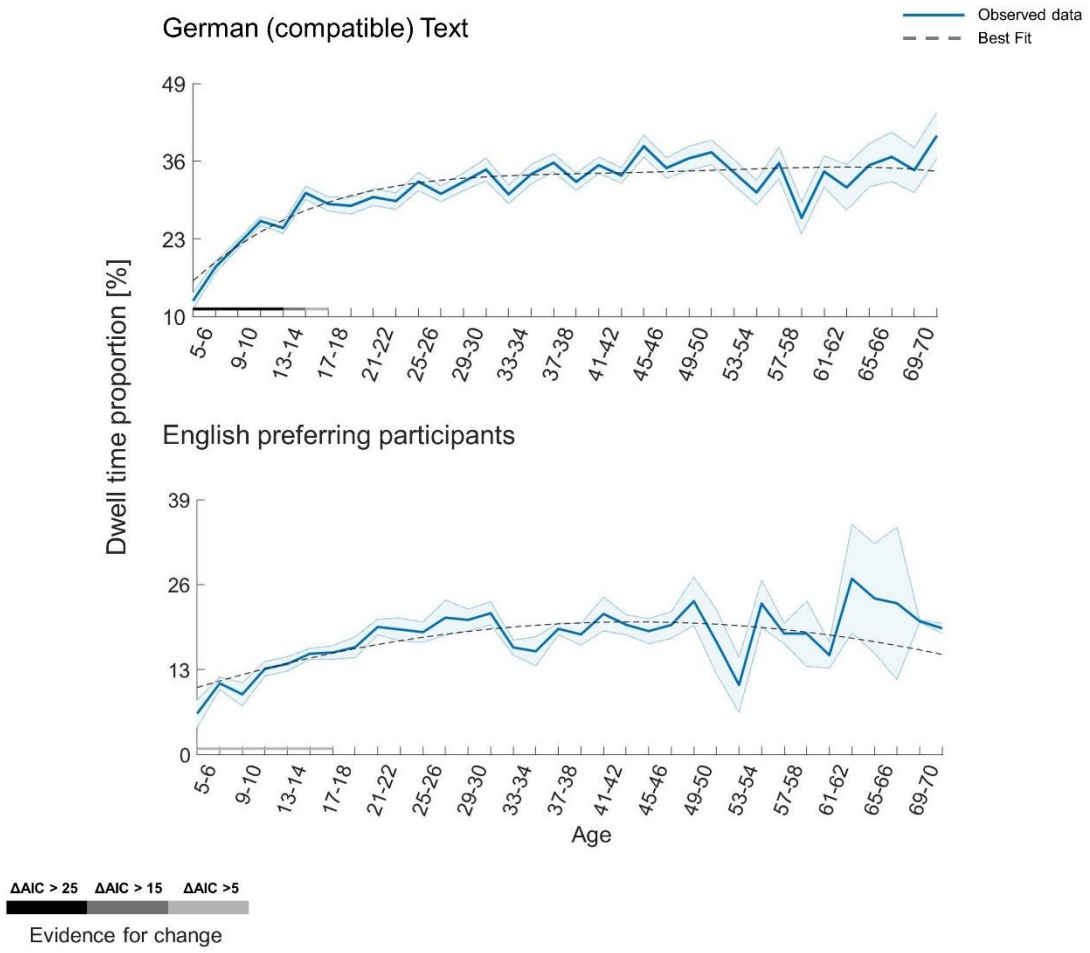

Figure S2. Protracted development of semantic salience. Line plots display the mean proportion of dwell time towards text elements (y-axis) across age bins from ages 5-6 to 71-72 (x-axis) for images containing text elements common in German and English (upper panel) and for the English-preferring sample (lower panel). Error shades indicate +/- 1 standard error of the mean (SEM), dotted lines best polynomial fits. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. main text and methods).

Protracted development of horizontal bias for images without text

Polar histograms replicated a strong increase in horizontal bias from ages 5-6 to 15-16 for the subset of images excluding text (Fig. S3 A). The same was true for the developmental trajectory of the proportion of horizontal saccades, which steeply increased from 51% to 59% between ages 5-6 and 15-16 (ΔAIC > 25), before increasing more

moderately to 60% at age 19-20 (ΔAIC > 5) and finally stabilizing (Fig. S3 B). The corresponding developmental trajectory was best fitted by a 3rd degree polynomial.

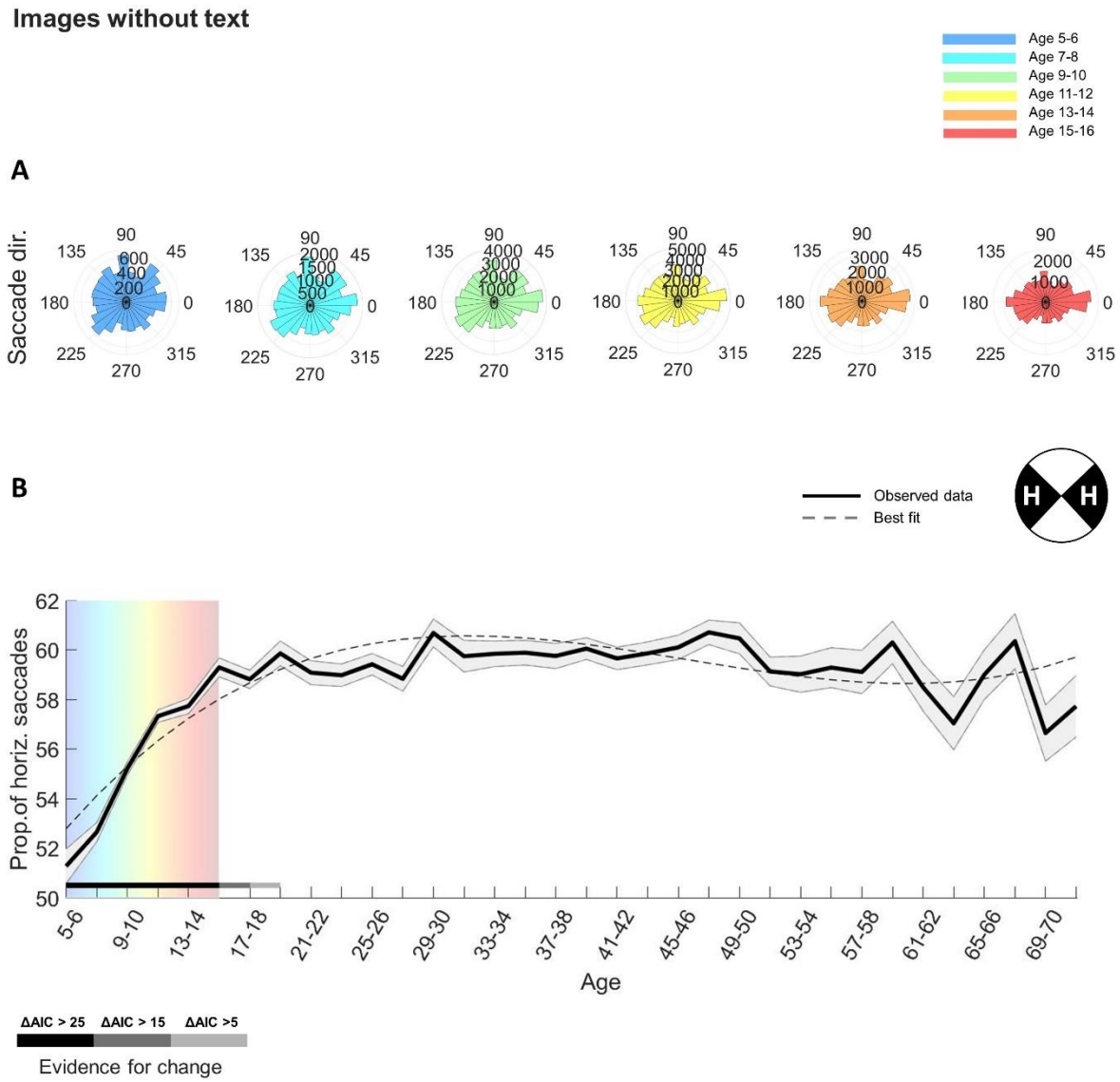

Figure S3. Protracted development of the horizontal bias for images excluding text. Panel A shows polar plots of the distribution of saccades across 36 direction bins from ages 5-6 to 15-16, as indicated by the colors shown in the inset. Panel B plots the developmental trajectory of the average proportion of horizontal fixations across age bins (solid line). The error shade indicates +/- 1 standard error of the mean (SEM) and the dotted grey line the best fitting polynomial. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. main text and methods).

Protracted development of visual exploration using a larger window size

Next, we tested linear developmental changes in the center-bias, indicated by the mean distance to the image centroid (Figure S4 A), visual exploration as indicated by fixation frequency (Figure S4 B), the absolute number of objects fixated (Figure S4 C) and intra-individual fixation entropy (Figure S4 D) via a sliding window approach using a window size of 10 bins.

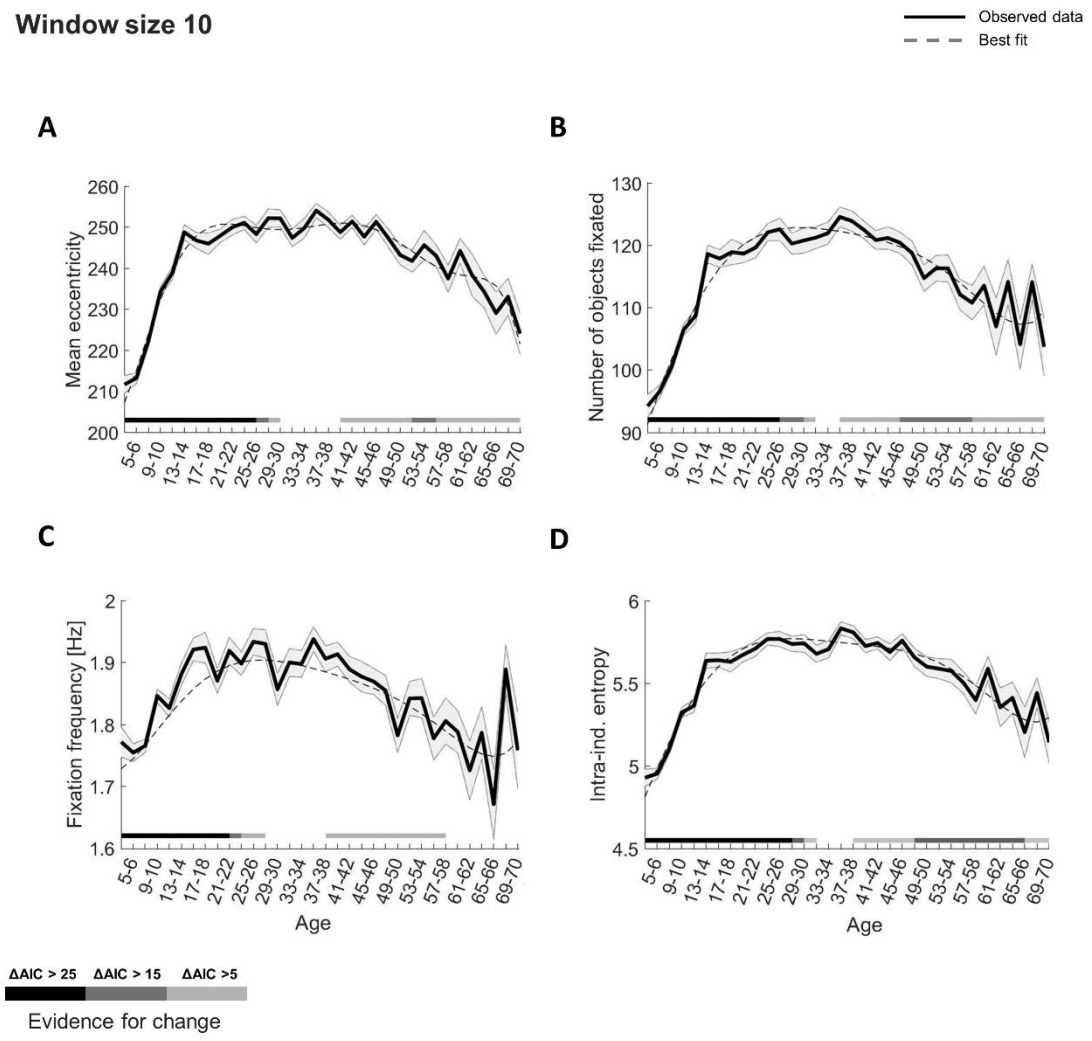

Figure S4. Protracted development of visual exploration. Line plots illustrate the mean distance to the image centroid (in pixels; A), the average number of objects fixated (B), the average fixation frequency (Hz; C) and the average entropy of observer-specific fixation maps (C) across age bins ranging from ages 5-6 to 71-72 (x-axis). Error shades indicate +/- 1 standard error of the mean (SEM), dotted lines best polynomial fits. The lines above the x-axes indicate the level of evidence for a given bin to be part of broad linear change via shades of grey, as shown in the inset (sliding window analysis, cf. main text and methods).

The average distance of fixations from the image center increased from 212 to 252 pixels between ages 5-6 and 31-32 ($\Delta AIC > 5$) and decreased from 249 to 224 pixels between ages 41-42 until 71-72 ($\Delta AIC > 5$). The number of fixated objects across all images increased from 94 to 121 between ages 5-6 and 21-22 ($\Delta AIC > 5$) and decreased from 125 to 104 between ages 37-38 and 71-72 ($\Delta AIC > 5$). Similarly, fixation frequency increased from 1.77 Hz to 1.93 Hz between ages 5-6 and 23-30 ($\Delta AIC > 5$) and declined from 1.91 Hz to 1.81 between ages 39-40 and 59-60 ($\Delta AIC > 5$). Finally, intraindividual entropy increased by 15% between ages 5-6 to and 37-38 ($\Delta AIC > 5$) and declined by 12% between ages 39-40 and 71-72 ($\Delta AIC > 5$).

Example overlay and trace plots for eye tracking data

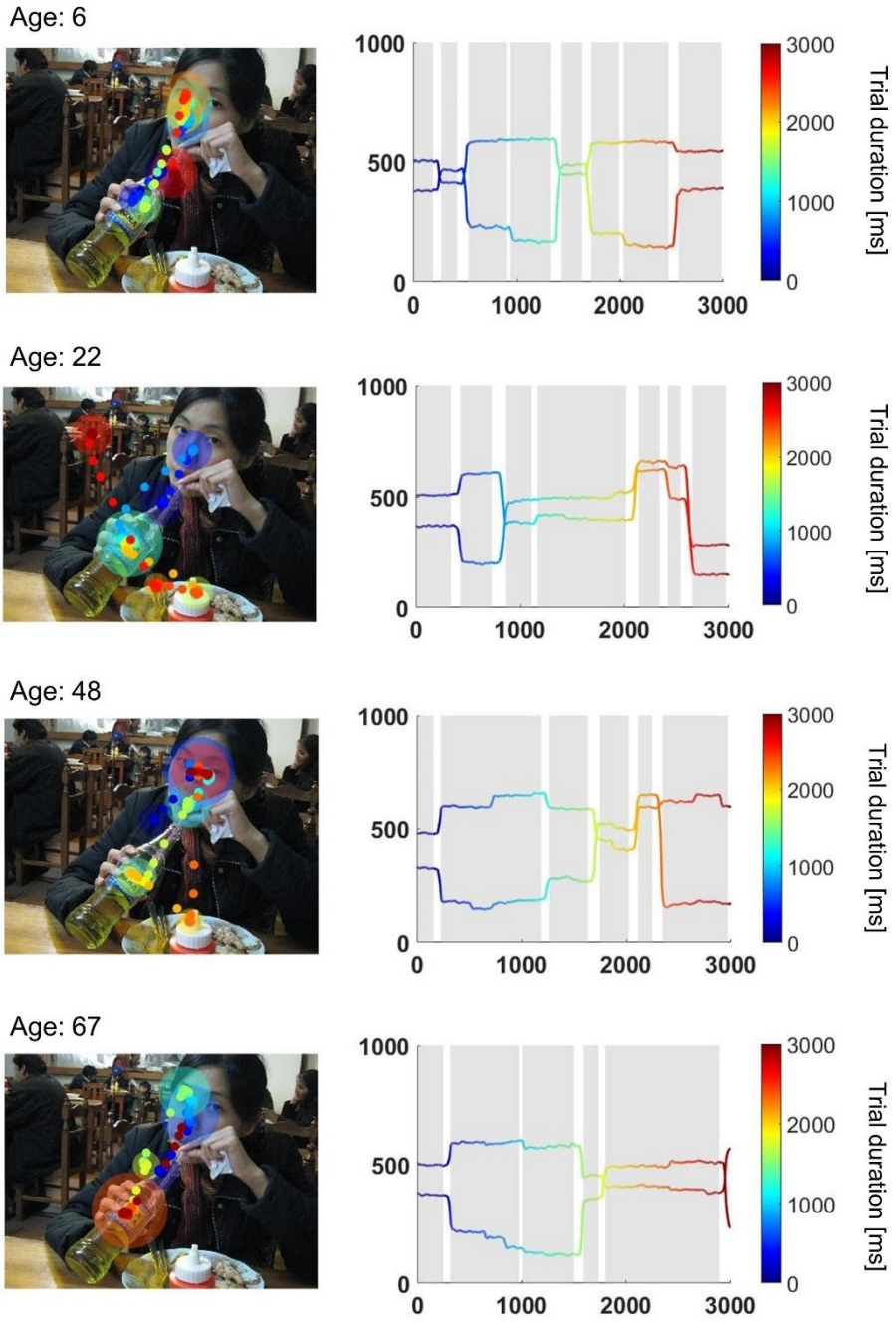

Figure S5. Example trace plots. The overlay plots on the left show the raw eye-tracking data as scatter points (90Hz). Transparent disks indicate fixation events and are scaled according to the corresponding fixation duration. Each row shows the data of an example participant from a different age group. The plots on the right show the corresponding horizontal and vertical position traces, with grey overlays marking periods classified as fixations. Each data point is color-coded to represent the trial time, ranging from 0-3000 ms, as indicated by the color bar on the right.

References

1. Linka, M. & de Haas, B. OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *J Vis*, **20**, 13 (2020).

5. List of all publications

Linka, M., Karimpur, H., & de Haas, B. (2024, February 14). Protracted development of gaze behaviour. https://doi.org/10.31234/osf.io/x7rkw

Linka, M., Sensoy, Ö., Karimpur, H., Schwarzer, G., & de Haas, B. (2023). Free viewing biases for complex scenes in preschoolers and adults. *Scientific Reports*, *13*(1), 11803. https://doi.org/10.1038/s41598-023-38854-8

Linka, M., & de Haas, B. (2023, March 30). Detection, Inspection, Return: A functional classification of fixations in complex scenes. https://doi.org/10.31234/osf.io/bqfdy

Linka, M., Broda, M. D., Alsheimer, T., de Haas, B., & Ramon, M. (2022). Characteristic fixation biases in Super-Recognizers. *Journal of Vision*, *22*(8), 17. https://doi.org/10.1167/jov.22.8.17

Linka, M., & de Haas, B. (2020). OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *Journal of Vision*, *20*(9), 13. https://doi.org/10.1167/jov.20.9.13

van Hooijdonk, M., Ritter, S. M., **Linka, M**., & Kroesbergen, E. (2022). Creativity and change ofcontext: The influence of object-context (in)congruency on cognitive flexibility. Thinking Skills and Creativity, 45, 101044. https://doi.org/https://doi.org/10.1016/j.tsc.2022.101044

Keck, J., Honekamp, C., Gebhardt, K., Nolte, S., **Linka, M**., de Haas, B., Munzert, J., Krüger, K., & Krüger, B. (2024). Exercise-induced inflammation alters the perception and visual exploration of emotional interactions. *Brain, Behavior, & Immunity - Health*, *39*, 100806. https://doi.org/https://doi.org/10.1016/j.bbih.2024.100806

Broda, M. D., Borovska, P., Kollenda, D., **Linka, M**., de Haas, N., de Haas, S., & de Haas, B. (2024). *Estimating the memory bottleneck for contact tracing*. PsychArchives. https://doi.org/10.23668/psycharchives.14175

6. Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmitel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde, oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.

_____ _____

Datum Unterschrift