DOCTOR OF PHILOSOPHY

Development of attention to social interactions in naturalistic scenes

Mihai, Ioana

Award date:
2022

Awarding institution:
Bangor University

Development of attention to social interactions in naturalistic scenes

Ioana Raluca Mihai

Thesis submitted to the School of Human and Behavioural Sciences, Bangor University, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Bangor, United Kingdom | October 2021

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Acknowledgements

I want to thank my committee – Kami Koldewyn, Patricia Bestelmeyer, and Rob Ward – for their encouragement in these four years. I thank my examiners for taking time to read this work.

Kami, thank you for being such an inspiring supervisor and mentor. Thank you for your amazing scientific guidance, for being so patient, kind, for believing in me and for watering that seed of passion that brought me on this path. I feel so grateful and privileged to have had the chance to work and learn from a great scientist like you.

In these four years I have learnt plenty. Aside from the amazing scientific training I received, which I am extremely grateful for, I have also learnt to never forget where I started. I am aware of the long way I've come, and this way is paved with the most amazing of the support.

I thank my family for the unconditional and constant support and for making this possible. I wouldn't have made it without their constant education to dreaming and looking ahead. Grazie per avermi educata a sognare in grande, a crederci, e a stringere i denti. Non sarei dove sono oggi senza il vostro supporto e questa tesi è anche vostra.

I thank my brother Bogdan, for his constant presence. You are the best, kind and loving brother one could desire, and I am lucky to be your sister and I look up to you. I feel safe and held by you, and this thesis is yours as well.

Aoife, you are the most one can ask from a friend. Thank you for holding me and empowering me. I feel so very privileged to have you in my life, and your strength and trust have been incredibly inspiring. Stephen, we've met at a strange time, but I thank you for taking care of me, and for being such an amazing adventure partner. Thank you for laughing with me until my face would hurt. Eva, you are one of the strongest and inspiring people I know. I feel so lucky to have you as a friend, thank you for taking care of me. Paul, thank you for keeping my philosophical spirit alive, for making me laugh and for the soul feeding conversations. Jade, you are one of the sweetest people I know, thank you for checking on me, for giving me space and believing in me. I hope one day to return the favour. Ionela, you have been indeed a constant discovery, thank you so much for reconnecting me with my roots, you really are so inspiring.

Jon and Marco, it's been such a pleasure and privilege to have met you. Thank you Jon for the travels into the best of the humour and Marco, thank you for the colours and for bringing Italy a little bit closer. Julia, we started off as office mates, but now I am so lucky to be able to say I'm your friend, and your support, especially in these last 2 years, has been crucial. Rachel, we have been writing our theses together, supporting each other every day. Thank you for your blunt and unfiltered support, it's been really important to have a partner like you in this journey. Caroline, thank you for feeding me, and for teaching me to be kind. Greg and Luke thank you for being there when I needed it the most. Ciara and Emma, you are two inspiring examples of strong empowering women, and I look up to you.

Alice and Ale Bollini, you really are two strong women and I'm so lucky to be your friend. Pedro – my concert partner –, Fracco and Davide, I've known you for years, and you've never stopped supporting me, and I'm grateful for that.

I want to thank the other lab members, Lois, Laura, Katie and Simona for being such good companions and for welcoming me. Julia, Simona and Jon working with you has been a privilege and a unique learning opportunity. Thank you for your patience.

Prof Toraldo and Dr. Annalisa Parente, thank you for lighting the first flame of the passion that brought me here and for supporting me in the very first steps in this direction.

Finally, I would like to thank a big part of my mental health support team: Naga, Tanuki, Asta, Maia, Pookie, Mims, Tilly, Dexter, Toby, Mac, Eddie and Hector.

Table of contents

Summary

Human visual attention is exquisitely specialized for and captured by social information in naturalistic scenes, and this “social preference” begins as early as infancy. Recent visual perception and neuroimaging research suggest that people also have high visual interest in interacting dyads compared to non-interactors, supporting the idea that social interactions might provide unique social information, above and beyond the mere presence of two people. Thus, interactions might capture visual attention above other social information, a ‘preference’ that could contribute, across development, to social learning processes. However, very little work has directly examined if attention is more biased to social interactions than to other social information in complex scenes, nor how and when such a bias may arise or change across development.

In this work, across three free-viewing eye-tracking experiments, we investigate the development of attention to social information in naturalistic, complex scenarios, in order to better understand the role of social interactions in such processes. In experiment 1, scenes contained dyads who were either interacting or not, while in experiment 2, dyads were presented together with one or two additional non-interactors to evaluate developmental changes in attention to social information when social interactions compete for attention with other social information. In experiment 3, the pictures contained ambiguous social dyads and, after the free viewing session, participants were asked to indicate whether they perceived the two agents as interacting or not. Our aim in that experiment was to investigate age differences in the way pre-existing knowledge about social events (e.g. the concept of what it means to be a social interaction) might influence social attention.

In all three experiments, we compared attentional engagement and capture by social areas of interest (i.e., human information) with non-social information (other scene elements), further contrasted by whether scenes were interactive or not. Results revealed both children and adults manifested a strong human attentional bias in the first two experiments, but a weaker bias in the third, when the social information was ambiguous. This social bias towards human information was moderated by the presence of a social interaction only in the first experiment, but not the second or the third, and was moderated in a similar way across development. In experiment 2, when interacting people were contrasted with non-interacting people in the same scene,

interacting people capture attention more quickly and engage it more strongly than other social targets when there is one other agent in the scene for both adults and children. However, this effect is smaller (and not significant) in children than for adults when an interaction competes with a pair of non-interactors. This suggests interactions can take attentional priority, but that this "interaction bias" increases across development, especially as scenes become more crowded and complex. Finally in the third experiment, we find that adults were more likely to see ambiguous scenarios as interactive compared to children. However, this difference was not reflected in the way attention was oriented to social information, as the social bias was similar in the scenes categorized as interactive or not and not different across development.

The results in this thesis are consistent with the idea that social interactions carry additional information, compared to isolated humans, and even more importantly, that this 'bias' to attend to social interactions is present as early as six. Finally, the implications of these findings for social attention and social development are discussed, followed by a discussion of future theoretical and experimental questions left to explore.

Chapter 1. General Introduction

1. Rationale for the thesis & overview of the chapter

Human beings have evolved as social creatures, equipped since the first hours of life to detect the social information surrounding them, and predisposed to learning how to make sense of it. This bias towards 'socialness' contributes across the lifespan to the development of highly specialized social understanding skills, reflected in high sensitivity to social information, brain regions specialised for detecting and processing social cues, and especially the ability to learn from observing others (i.e., social learning). Research looking into the human ability to perceive faces and eyes has shown, for example, how tuned human visual attention and perception are towards social information. This results in an attentional bias whereby we typically prefer to attend to social rather than non-social information, as demonstrated across a variety of experimental tasks and stimuli (briefly reviewed below). We know less about how human attention filters and makes sense of the constantly changing, noisy, and complex social information that bombards us, such as the wide variety of social interactions we observe between other people. Understanding our bias towards "people watching", and how we process social cues from those we observe has recently received much more scientific attention. There is some evidence that children use the encounters they observe to guide their own social choices (e.g. Skinner et al., 2017) and some evidence that there is an added value of third party encounters (Quadflieg & Westmoreland, 2019). Despite this, there is still relatively little exploration of how visual attention operates in the context of observed social interactions, and there has been little exploration of how the cognitive systems that support such social attentional processes develop.

Investigation of visual processing of observed social interactions has made some recent progress. Indeed, recent research suggests that social interactions are processed differently than individual figures; in other words, that interactions are treated differently than just the 'sum of their 'arts' (Walbrin & Koldewyn, 2019). Additionally, it seems that the human brain might be highly tuned to the processing of interacting dyads, perhaps even to a similar extent as it is to face, body and motion processing (Isik et al., 2017; Walbrin et al., 2018). Although it is clear that there is early visual sensitivity to social interactions in infants and that young children can learn from 3rd-party encounters, it has recently been shown that the social brain is not yet fully 'adult-like', even at 12 years of age when processing social interactions.

Understanding how social interactions are perceived and processed in typical childhood is particularly important in understanding adult social processing and how the 'social brain' develops. Although understanding these processes in neurodevelopmental disorders will also be of great importance, social attention in neurodevelopmental disorders is complicated by reported non-social differences in attention and executive function, particularly when social information is embedded in naturalistic, cluttered, and complex scenes. Here, we focus on the development of social attention across typical childhood, specifically looking at how the presence of a social interaction may modify social attention in complex, naturalistic scenes.

Therefore, the purpose of this work has been to explore how spontaneous attention to social information in complex scenarios develops across childhood and investigate the potential role of social interactions in these processes. This is one of the first attempts to investigate these questions using naturalistic scenes.

Study I investigates the influence of a social interaction on attentional orienting to social information in complex scenes across pre-adolescent childhood. Study II introduces competition between social interactions and other social information in the same scene, and Study III investigates the role of pre-existing top-down social knowledge on social orienting of attention in ambiguous social scenarios.

Overview of the chapter

In this chapter I will briefly outline some developmental aspects of visual attention, with a focus on the tripartite model of visual attention proposed by Posner (Petersen & Posner, 2012; M. I. Posner & Petersen, 1990), scene exploration, and eye movements research. This will provide a theoretical developmental context for the social attention research presented in this thesis. I will then proceed with briefly reviewing the evidence for an attentional social bias in development, and the research surrounding visual sensitivity to social interactions. Finally, I will briefly discuss social orienting in scenes in neurodevelopmental disorders and discuss how that literature informs our understanding of typical development.

2. Developmental aspects of visual attention

2.1 Attention

Attention is an umbrella term for a multi-componential process of allocation of brain resources to accomplish behavioural tasks and goals (Atkinson & Braddick, 2012). Several models of attention have been proposed, focusing on different features and functions (for reviews see Carrasco, 2011; Chun et al., 2010; Kanwisher & Wojciulik, 2000; Knudsen, 2007; Petersen & Posner, 2012; Posner & Petersen, 1990), but the most common aspect amongst models is the conceptualization of attention as a filter selecting or limiting the processing of information beyond perception so that only a small subset enters memory, or is used to support learning and action (Rueda et al., 2015; Scerif, 2020). In the tripartite model of attention proposed by Posner et al. (Petersen & Posner, 2012; M. I. Posner & Petersen, 1990; Michael I. Posner & Boies, 1971), attention is conceptualized as an activation state preparing the individual for action. Attention is a tool for selecting information from the environment based either on goals and priorities or by 'bottom-up' stimulus-driven factors (e.g., sudden movement, salience). At the same time, attention is also a fundamental instrument of action control and conflict monitoring when we are *doing* as well as perceiving (Rueda et al., 2015). Therefore, in this model, attention subserves three different functions: orienting, alerting, and executive control. In particular, the orienting system flexibly engages and orients attention – first by orienting to the most 'important' inputs in the environment, then by disengaging and shifting attention between different regions of visual space, depending on internal goals and representations (endogenous) or based on stimulus saliency (exogenous) (Atkinson & Braddick, 2012; Colombo, 2001). The alerting network, on the other hand, has the role of maintaining high sensitivity of the system to incoming information, and is crucial in tasks that require sustained attention (Atkinson & Braddick, 2012). Finally, the executive control network is involved in the regulation of responses in a goal-directed effortful mode (Petersen & Posner, 2012; Pozuelos et al., 2014).

To assess these components and functions of attention, the most used task is the Attentional Network Task (ANT) (Fan et al., 2002) and its variants, where the participant has to decide the direction of the middle arrow out of a row of five. The three attentional networks are evaluated through the measure of the response times under influence of warning cues, spatial validity of the cues, and congruency of the flankers (Figure 1).

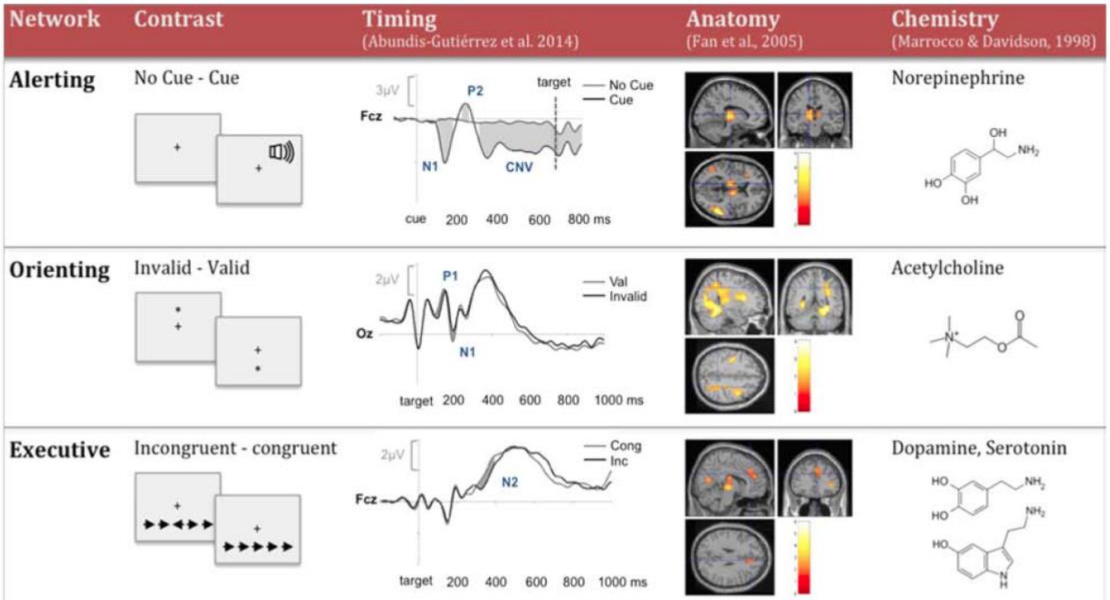

Figure 1. Attention Networks as investigated by different methodologies (figure taken from Rueda et al., 2015).

2.2 Development of visual attention and processing speed

Investigation of developmental processes of attention not only reveals how children attend to and learn from their environment, but has the potential to offer important insights into how adult attention works and how the adult cognitive state has been achieved over time (Amso & Scerif, 2015; Astle & Scerif, 2009). One way that developmental work provides insights is through disentangling which attentional functions have a similar developmental trajectory. Do all attentional functions develop *together* or do some skills develop before others? In addition, understanding what is intelligible and relevant for cognitive systems in development is important in order to understand what information has priority for learning and in the development of the systems themselves, and to better understand neurodiverse functioning (Amso & Scerif, 2015; Scerif, 2010, 2020). Indeed, attention is also an important tool to understand what pieces of knowledge from the environment are selected based on their priority and usefulness for the individual's goals (Henderson, 2003) and what information is most important for individuals at specific moments in development (Amso & Scerif, 2015; Soto-Icaza et al., 2015).

Research into the development of attention has been crucial in informing models of how infants and children learn. To start with, as infants have limited motor abilities early in life, visual attention is a fundamental tool for exploration and understanding of the environment, giving infants an active role in their own development. The first signs of attentional development (for reviews Colombo, 2001; Plude et al., 1994; Reynolds & Romano, 2016) is measured by the way eye movement control develops. This control is already highly efficient at 4 months (Colombo, 2001; Rueda et al., 2015; Scerif, 2020), when infants already show the ability to control how they orient their eyes (and attention) towards peripheral stimuli (Atkinson & Braddick, 2012). At this stage, infants transition from the "sticky" type of attention they show prior to this, characterised by a difficulty in disengaging their attention from something they are looking at and, thus, in shifting attention between locations and to disengage it (Johnson, 2001; Plude et al., 1994; Scerif, 2020), to more controlled eye-movements that allow them to voluntarily and flexibly shift their attention. During the first year of life, infants' attention is increasingly driven by internal factors, such as interest, including by a recognition of the novelty of a stimulus (Scerif, 2020). Starting at 4-months, they can voluntarily shift attention between different positions in space, and between that age and into the school years, executive control of attention slowly increases and gets refined to allow goal-oriented management of the attentional tools (Amso & Scerif, 2015; Colombo, 2001; Plude et al., 1994; Rueda et al., 2015).

While during adulthood visual attention is guided by endogenous and exogenous factors in a balanced way depending on goals and task requirements (Chun et al., 2010; Henderson, 2003; Knudsen, 2007; Petersen & Posner, 2012; M. I. Posner & Petersen, 1990), during childhood exogenous and endogenous factors drive attention following separate developmental trajectories, with exogenous effects emerging earlier compared to endogenous (for a review see Scerif, 2010). In childhood, speed of processing improves as well, resulting in an enhanced precision of information processing which settles and becomes 'adult-like' in middle adolescence (Kail, 1991; Luna et al., 2008; Miller & Vernon, 1997). This research suggests that in childhood many changes occur in the attentional tools.

Studies using the Attentional Network Task (Fan et al., 2002) based on the Posner model of attention (Petersen & Posner, 2012) have investigated developmental changes in the three attentional networks (Konrad et al., 2005; Mullane et al., 2016; Pozuelos et al., 2014; Rueda et al., 2004) and have shown that their development shows separate trajectories. In particular, the executive side of the attentional networks has received a lot of scientific attention and many studies suggest that attentional control continues to develop until adulthood, showing change throughout adolescence and young adulthood (Amso & Scerif, 2015; Crone, 2009). Less is known, however, about the orienting and alerting systems, also termed the spatial and temporal aspects of attention. Rueda et al. 2004 (similarly to Mullane et al., 2016) showed no important qualitative changes in the covert orienting of attention between age 6 and 10, but performance improved in speed and accuracy with age, and adults scored significantly better in attention altering tasks compared to children (Rueda et al., 2004). More difficult variants of the ANT show changes in both the alerting and executive networks between 6 and 11 years of age (Cowan et al., 2010; Konrad et al., 2005; Mullane et al., 2016; Pozuelos et al., 2014), with the youngest children benefitting the most from a warning auditory cue before target appearance, indicative of younger children having greater difficulty in maintaining tonic alertness when monitoring targets (Pozuelos et al., 2014). No significant changes were seen in the orienting attentional network, although when comparing the youngest children with the older children (11-12-years-old), some substantial differences were seen in performance, with slower and less accurate responses in the youngest children when they had to reorient attention to the target, and increasing ability to orient and reallocate attention with age (Pozuelos et al., 2014). Across these studies a common factor that emerges is age related increases in both speed and accuracy, changes that go on until adulthood in the executive network, but that seem pretty much ‘adult-like’ around 11-12 years of age for the orienting and the alerting networks. This difference in trajectory suggest that the orienting and alerting networks develop first.

Interestingly, attention and mechanisms regulating eye movements have overlapping neural systems (Amso & Scerif, 2015), and develop concurrently. Indeed, it seems that basic eye motor mechanisms are already adult-like at 8 years of age (Karatekin, 2007; Luna et al., 2008), as indicated by children and adults showing similar peak velocities of saccades once the movement has been initiated (Fukushima et al., 2000; see Luna et al., 2008 and Karatekin, 2007 for reviews), and by the fact that children of at least 8-years are as able as adults to adjust and re-orient attention following task requirements (for a review, Plude et al., 1994). However, initiating a voluntary oculomotor response to a visual target, independently of any cognitive task, is much slower in children and only becomes mature during the teenage years (Fukushima et al., 2000; Luna et al., 2008; Plude et al., 1994). This finding is consistent with a later development of frontal cortex areas involved in the control of eye movements, and adult-like executive control of saccades is not yet mature until adolescence, as indicated by a progressive decrease of errors in anti-saccade tasks (e.g. Fukushima, 2000; see Karatekin, 2007 for a review).

Finally, these three attentional networks, despite working together, have separate (but overlapping) brain networks (M. I. Posner & Petersen, 1990). Konrad et al. (2005) have investigated developmental changes in the neural circuits involved in these three key aspects of attention. Brain data seemed to mirror behavioural findings of higher sensitivity to interference and a weaker reorienting system in children, with decreased brain activity in the a priori defined regions of interest for the three networks, and more distributed activation outside the pre-defined regions, indicating less specificity of activation in the developmental sample. In particular, adults showed greater activation in the alerting fronto-parietal network, orienting right inferior frontal gyrus, temporo-parietal junction, and bilateral superior parietal cortex. Finally, the control network induced greater activation in a frontal circuit including the anterior cingulate gyrus and the dorsolateral prefrontal cortex (Konrad et al., 2005). This suggests that the networks subtending executive control and reorienting of attention are not yet developed and tuned in 8-12-year-old children, and might undergo further specialization towards more focal and less distributed circuits during adolescence.

The work presented in this thesis investigates the development of visual attention through free-viewing paradigms where participants freely explore complex, although static, naturalistic scenes. Research showing slower voluntary oculomotor responses in children suggests our child participants may be slower to orient towards information compared to the adults, but in general the attentional tools necessary to orient to and explore social scenes are already in place in the age-group we considered in this thesis.

2.3 Scene processing and exploration

In this section I will discuss briefly important aspects of scene processing and scene exploration, with a focus on the developmental changes in these patterns, before briefly describing the social attention literature.

As explained in the General Methods chapter, high quality visual perception happens only in a small portion of the retina, the fovea, and the resolution of visual information decreases the further away from the fovea (Holmqvist et al., 2011). Therefore, the eye has to move across a scene in order to parse the visual information it contains in detail. As mentioned above, eye movements and attention are highly linked: eye movements are the overt phenomenon of attention allocation (Corbetta et al., 1998; Henderson, 2003), meaning that the portion of space we gaze at is, under most circumstances, also the focus of our visual attention. Thus, many researchers analyse eye-movements to better understand visual attention (see General Methods chapter for details on the eye-tracking methodology).

When exploring a scene, people shift their eyes, and attention, to informative regions (Buswell, 1935; Henderson, 2003; Yarbus, 1967). Attention and eye movements are therefore also a tool to understand what is selected by the individual from the environment based on its relevance (Henderson, 2003). When there are no task requirements and participants are allowed to freely explore a scene, what they orient to and explore most in a scene can reveal which aspects of the visual world are intrinsically compelling. Additionally, if an individual has more time to explore the scene, typically the eye will return to the most informative regions of the scene rather than exploring the whole area of the picture (for a review see Karatekin, 2007). Humans can be very quick to understand a scene: we can get the gist of a many scenes from just one (well-placed) single fixation (Henderson, 2003) and we can be very fast – as quick as 100 ms (Greene & Oliva, 2009) or even 40ms (Castelhano & Henderson, 2008) – in categorizing objects or even the context of the scene (Joubert et al., 2007; Rousselet et al., 2005).

Regions within scenes can be informative and guide attention and eye movements from two perspectives: from a stimulus-driven (bottom-up) perspective and from a knowledge/goal based (top-down) perspective (Henderson, 2003; Henderson & Hollingworth, 1999). There are different models of stimulus-driven gaze control (see Henderson, 2003 for a review), but the bottom line is that contrast, colour, intensity, and edge orientation can be informative about the contents of a scene and are likely to capture and hold attention. However low-level features and bottom-up models are not enough to explain gaze allocation in scene exploration (for reviews, Henderson, 2003; Kaspar, 2013). Indeed, top-down factors can, and do, override bottom-up drivers of attention, depending on the scene and the task. The top-down factors mostly considered when investigating gaze allocation in scenes are task requirements (e.g., searching for particular items), memory (e.g. familiarity with the stimulus), spatial and semantic information (e.g., cups are usually on top of rather than beneath tables) and the observer's goals (DeAngelus & Pelz, 2009; Henderson, 2003; Kaspar, 2013). Important for the research presented in this thesis is the fact that human vision highly benefits from the use of context in object search and detection in a scene (Henderson & Hollingworth, 1999; Oliva et al., 2003; Torralba et al., 2006), suggesting that scene viewing and exploration is not only guided by low level features, but also and especially (Henderson & Hayes, 2018) by the semantics (i.e., the predicted relations between the objects in a scene) and meaning of the scene.

The development of scene exploration has not received much research focus, but much of the work that has been done has focused more on social scenes. This work generally shows that if a social element is in the scene, it will grab attention (see below for detailed discussion of this literature) and that social cues are more powerful than other factors that guide children's attention, especially when using naturalistic stimuli and free viewing tasks (for a review Karatekin, 2007). However, there are several differences between children and adults in how scenes are explored, as assessed through eye-movements. To start with, as mentioned above, initiating a saccade is slower for children than for adults (Fukushima et al., 2000; Luna et al., 2008; Plude et al., 1994; Rayner, 1998), but otherwise eye movements are similar across ages. Helo et al. (2017) have investigated the role of semantics (i.e., the predicted relations between the objects in a scene) and perceptual features of the scene on visual attention in 2-year-olds, showing that toddlers were more influenced by the perceptual features of the scene compared to adults, despite both groups showing effects of the semantics of the scene, indexed by detection of object placement that was inconsistent with the typical semantics of such scenes (e.g., a tea-pot in the bath or toothpaste on the stove; Helo et al., 2017). One of the few studies to investigate scene exploration across development showed that 7-9-year-old children, compared to adults, focus more on local details, and are more influenced by bottom-up processes, like the saliency characteristics of the scene (Açık et al., 2010), while adults display more influence by top-down processes, such as looking to task relevant aspects of scenes. Additionally, it would seem that variability in gaze allocation across a scene decreases with age, suggesting that individuals increase their alignment with group systematic viewing tendencies across development (Kirkorian et al., 2012).

2.4 Summary of the section

Attention is multi-component construct, referring to the processes involved in allocation of cognitive resources, information filtering and selection, together with maintenance of an activation state of the individual and control of action, as conceptualized coherently by the tripartite model of attention. The three functions of alerting, orienting and control have been shown to be separate in both behaviour and in the brain.

The childhood years see many cognitive changes as indicated by a general improvement in the speed and accuracy of processing information. However, basic attentional orienting skills are already in place by 8 years of age, and peak saccade velocity is similar to adults at that age. However, processes involving executive control of attention and the initialisation/control of eye movements continue to develop throughout adolescence, changes that are mirrored in brain imaging data that suggests late development of frontoparietal circuits involved in attention, changes that continue into young adulthood.

Scene exploration is regulated by both low-level features such as salience, and top-down processes, such as task, motivation, memory and knowledge about relations between objects. Little research has investigated developmental changes in scene exploration, but there is some evidence that attention is more influence by low-level features in the scene during childhood, and is increasingly driven by meaning and other top-down processes across middle-childhood and adolescence.

3. The social bias in attention and the development of social attention

In this section I will provide context for social attention research and review some evidence of social tuning and social bias in human visual attention. I will proceed with a description of developmental changes in social attention, and describe how my thesis research is positioned relative to the current state of art.

3.1 Social vision

Humans have evolved as exquisitely social beings: the complexity typical of living in big social groups has contributed immensely to the development of large brains in humans and the need for the complex cognitive skills that are characteristic of humans (Dunbar & Shultz, 2007). The social skills needed to guide social behaviour and way in which the human brain is tuned to detect and understand social cues are both described by the field of ‘social vision’ (Nakayama, 2011; Papeo, 2020).

People’s eyes, faces, and body cues give away a great deal of information about their goals and intentions (Nummenmaa & Calder, 2009), and it is generally agreed that human beings become experts, since early in life, at looking to other people to catch relevant social cues, including cues to relationships and personality in order to make social inferences and decisions (Quadflieg & Westmoreland, 2019). This developing expertise contributes, across the lifespan, to the development of a wide range of complex social skills (Quadflieg & Westmoreland, 2019; Soto-Icaza et al., 2015). Given this, a great deal of research has looked into the processes that drive human vision towards social information, how the extraction of social information happens mechanistically, and how this 'social bias' contributes to the understanding of the complex social environment.

Faces have long been the focus of social perception research, and a variety of methods and paradigms have shown that thanks to its range of movements and expressions, the face is a very efficient social communication tool (for a review, Jack & Schyns, 2017). Indeed, a face can provide interesting insights about one's direction of attention (Langton et al., 2000), emotions (Langton et al., 2000), goals and intentions (Quadflieg & Westmoreland, 2019), or even social status (Cloutier & Gyurovski, 2014; Quadflieg & Westmoreland, 2019). What's more, a great deal of research has shown that human vision is driven to detect and understand other people's gaze, perhaps for its potential to communicate other people's attentional state and thus interests, together with their mental and emotional states (Birmingham & Kingstone, 2009; Langton et al., 2000; Risko et al., 2016).

Interestingly, simple perception of human information seems to not need awareness, as assessed by methods similar to continuous flash suppression (Gobbini et al., 2013) and even more, a whole human figure in a naturalistic scene is granted privileged access to awareness (Gray et al., 2018), suggesting the strength of the specialization of human vision for social information.

Furthermore, people are highly skilled in extracting human information – like a body form – even from impoverished dynamic stimuli like point-light walkers, created from the digital capture of the motion of a person with lights attached to the main joints, while walking or doing a variety of other actions (Neri et al., 1998; Pavlova & Sokolov, 2000). Additionally, we can recognize features such as sex, age, emotions from these impoverished stimuli (Ma et al., 2006).

This evidence suggests how action and human movement recognition might be crucial for one's survival as it allows ready prediction of others' intentions, and as a consequence, a better understanding of one's environment (Thompson & Parasuraman, 2012). Additionally, it suggests and brings a strong support to a deeper investigation of attentional processes to complex social information presented in this work.

3.2 Social brain

This social tuning of vision is mirrored in the function of the brain and its specialized structures for processing of social information (C. D. Frith, 2007). Similar to the behavioural research, a great deal of brain research has been dedicated to the investigation of how humans process faces and face identity. Indeed, there is enough evidence to support the existence of specialized brain networks for extraction of information from faces and that allow humans to promptly recognise familiar faces and crucial features like emotions, attentional state, and focus of attention. The central hub of this face perception network is in the fusiform gyrus, the fusiform face area FFA (for a review, Kanwisher & Yovel, 2006), but also includes regions in the inferior occipital gyrus (occipital face area, OFA; Hoffman & Haxby, 2000), and in the superior temporal sulcus (STS; Deen et al., 2015). As seen earlier in this chapter, a fair amount of research has been dedicated to gaze perception and detection, and this includes brain research, which has shown that a wide network is recruited in the detection and understanding of eye-gaze, with the STS at the core of this network (Birmingham & Kingstone, 2009; Langton et al., 2000; Nummenmaa & Calder, 2009), but including also some regions in the amygdala, fusiform gyrus, and medial prefrontal cortex (for a review, Birmingham & Kingstone, 2009).

Body and body parts perception and processing seem to recruit a set of specialised areas such as the extrastriate body area (EBA) and the fusiform body area (FBA) (for a review, Downing & Peelen, 2016). Indeed, neuropsychological evidence and TMS studies have shown EBA to be recruited not only in response to isolated human figures but also in the detection of people in more cluttered scenes and in other tasks involving the human body, such as body shape discrimination, as well as movement and action perception (Downing & Peelen, 2016). In addition, there is some evidence that EBA is involved in detecting and coding for facing bodies, as opposed to non-facing dyads, and thus may also be involved in at least some aspects of social interaction processing (Abassi & Papeo, 2020).

When taking a wider look at the person perception and social cognition literature, the suggestion is that the STS may be a central hub for the processing of important cues for social understanding and social cognition (Allison et al., 2000; Deen et al., 2015; Zilbovicius et al., 2006). Indeed, evidence shows that this region has multiple sub-areas that respond selectively to certain types of social information, such as the human voice and language processing, dynamic faces (Deen et al., 2015), gaze direction (E. A. Hoffman & Haxby, 2000; Nummenmaa & Calder, 2009; Puce et al., 1998), mouth movements (Puce et al., 1998), moving hands and body movement (Allison et al., 2000), biological motion and action processing (Deen et al., 2015; Thompson & Parasuraman, 2012), and socially interacting dyads (Isik et al., 2017; Walbrin et al., 2018; Walbrin & Koldewyn, 2019). Accurate face, body and action recognition and the extraction of inferences about other that can be guided by face and body language (Quadflieg & Westmoreland, 2019) is thought to principally be supported by the network of social perception regions described above, though this interpretation is not uncontroversial. However, human social skills also involve higher order attributions about other humans, including understanding the unstated contents of others' minds, often called Theory of Mind (ToM). Such mentalising processes are thought to be supported by a network of regions that are specifically involved in such mentalizing processes. The main structures recruited in tasks that require either implicit and explicit inferences about other people's mental states include the dorsomedial prefrontal cortex (dmPFC) (U. Frith & Frith, 2003; Spunt & Adolphs, 2015) – which would seem to be able to distinguish a mental state representation from physical events –, a sub-region of the STS when participants are performing false belief tasks (Deen et al., 2015) – potentially involved in determining agency –, other regions in the temporal poles (Castelli et al., 2000; U. Frith & Frith, 2003) – likely recruiting semantic social knowledge – and the temporoparietal junction (TPJ) (R. Saxe et al., 2006).

Thus, it would seem that there are three main interconnected systems involved in social perception and social processing. Indeed, the structures involved in face and body information analysis could be considered a person perception network including regions involved in face, body and motion perception – occipital face area (OFA), FFA, FBA, pSTS –, that works hand in hand with an action perception network that recruits regions in the inferior parietal lobule (IPL) and inferior frontal gyrus (IFG), and higher order mentalizing processing in a network connecting the dmPFC, TPJ, the precuneus and the anterior temporal lobe (aTL) regions (Quadflieg & Koldewyn, 2017). Furthermore, some would suggest that at the crossroad of these networks the STS plays a central role (Yang et al., 2015), integrating social information from auditory and visual domains (Redcay, 2008), bringing together form information from the ventral visual stream with motion information from the dorsal stream, and through its links to frontal, parietal, auditory, and visual regions (Zilbovicius et al., 2006).

3.3 The question – social attention

In this section I will first provide a clarification of some terms used throughout this work. To start with, the scientific literature is various about what "social attention" is. Authors often use 'social attention' to refer to the orienting to and capture of visual attention by targets that are being attended by other humans (Birmingham et al., 2009b; Birmingham & Kingstone, 2009; Smilek et al., 2006). For example, when looking at a picture of someone looking at an object, we are likely to follow their gaze and *also* look at that same object. When used in this way, 'social attention' is used as attention to the attentional states of other social targets, as indicated by their orienting of gaze, head, and body orientation (Birmingham & Kingstone, 2009; Nummenmaa & Calder, 2009). Indeed, Birmingham & Kingstone (2009) refer to social attention as the orienting to and selection of the gaze as a means to understand what is important in the world, by taking the focus of attention of other humans as a point of reference.

However, 'social attention' has also been used much more broadly to refer to attention to any social target. In this definition, 'social attention' would include attention to faces, bodies, emotional expressions, actions, etc. Given the communicative value of social targets and the number of important social cues to goals and intentions given through body posture, and facial expressions (Nummenmaa & Calder, 2009), together with the high sensitivity of human vision to picking up these cues (Quadflieg & Westmoreland, 2019), attention to social targets is a vital first stage of visual information processing that eventually leads to the understanding and analysis of the intentions and mental states of other individuals (Allison et al., 2000). Indeed, *this* idea of social attention – a tool for visual exploration of human information – is how the term 'social attention' is used in the current work. The goal of the research included in this dissertation was to investigate the initial stages of attentional orienting, and later attentional engagement, to social information like human beings in a scene. In particular, the focus of the current work is the attentional selection of human information compared to non-human information, referred to as a "social or human bias", and how this selection operates on different levels of complexity at different stages in development, depending on the moderation by the presence of social interactions, social content (i.e. number of people in the scene) or social ambiguity.

3.4 A social bias

Research investigating the human attention ability that contributes to the extraction of visual information generally supports the idea that humans have an attentional bias towards faces, and especially eyes (Birmingham & Kingstone, 2009), as well as bodies. For example, covert attention paradigms have investigated the strength of social information to act as distractors in visual search tasks, by slowing down search times. Indeed, in a visual search task where participants had to look for a non-social stimulus – i.e. a butterfly – in a 6-element array, search time was greater when a face was among the distractors compared to trials where there were only non-social distractors (Langton et al., 2008). Additionally, when the social elements became the search target, the non-social distractors were not as efficient in capturing attention and slowing down search times. Similarly, Ro et al. (2007) show this interference effect is also shown when distractors are body parts rather than just faces and, importantly, is valid also in naturalistic scenarios (Ro et al., 2007). Indeed, Doherty et al. (2017) show that a human figure can slow down search times for an object in a scene, suggesting human information receives attentional priority even in more cluttered scenarios (Doherty et al., 2017). Downing et al. (2004) show that unexpected and task-irrelevant human body silhouettes can be resistant to inattentional blindness by being detected more frequently compared to other stimuli, like objects or scrambled versions of the body figures (Downing et al., 2004). Finally, in a visual search task where participants search for a scene depicting humans in an array of scenes containing machines, and vice-versa, the search for human scenes had shallower search slopes and the on target fixations were on average shorter for people compared to machines (Mayer et al., 2015), suggesting that human information might be easier to process.

A great deal of support for a social bias also comes from overt paradigms, where orienting, capture, and engagement of attention to social information as measured by eye movements is the main experimental question. Indeed even early work on scene exploration (Yarbus, 1967) has demonstrated an attentional interest towards a human figure when viewing scenes, such as paintings or pictures. Most of the work on attentional engagement and capture of human information has focused on the importance of gaze as a source of information about others' attentional states and targets (Birmingham & Kingstone, 2009), although other research suggests an interest in the human body as a whole (Bindemann et al., 2010; Sue Fletcher-Watson et al., 2008). When presented with two scenes, one containing a person and the other without, the majority of the fixation time goes to the scene containing the human, and in particular to the human area of interest (Sue Fletcher-Watson et al., 2008). Furthermore, when participants freely explored naturalistic scenes with one or three people, human information – eyes, heads, bodies – were the main objects of interest, while non-social items – foreground and background objects – were looked at later and less frequently (Birmingham et al., 2008). Similarly, when viewing scenes containing multiple people having social interactions, typically developing participants showed more looking time to the people in the scene compared the background (Riby & Hancock, 2008).

As in covert attentional paradigms, a human figure in a scene can act as a distractor also by capturing the first fixation to the scene even when the participant is not searching for a social target (Doherty et al., 2017). One would argue that this early attentional capture by human information could be explained by 'low level' visual features that have influence over early perceptual processes, but research suggests that this process is not entirely dependent on factors like saliency in the scene (Birmingham et al., 2009a; Rösler et al., 2017). Additionally, either when searching for a human scene among machine scenes, or vice-versa, the human target captured on average more first fixations than the non-social target (Mayer et al., 2015). Other studies have looked into exploration of more complex scenes depicting multiple people and even in such cluttered scenes the data demonstrates an attentional bias to human information compared to non-social information, with attention to eyes, faces and heads as the regions that receive the most attentional focus, followed by bodies and background elements (Flechsenhar & Gamer, 2017; Klin et al., 2002; Riby & Hancock, 2008). Although these are important insights into attentional processes to more complex and cluttered information, these studies don't disentangle how increasing the level of complexity of the social scene might affect social attention, nor whether the content of the social information influences the degree of 'social bias'.

3.5 Bottom-up and top-down aspects of social attention

Importantly, many models of scene perception suggest that low-level, stimulus-driven saliency, such as contrast, luminosity, and edges will capture attention (Henderson, 2003 for a review). While one might argue that early attentional capture by social objects might be influenced by lower-level elements such as the visual saliency in the scene, it has been shown that actually the presence of social information is a better predictor of early attentional orienting to a certain position in space compared to high saliency, indicating that attentional orienting to social information might be reflexive (Flechsenhar & Gamer, 2017; Rösler et al., 2017), or at the very least that attention to social information overrides visual saliency driven attentional mechanisms (Birmingham et al., 2009a; End & Gamer, 2017; Langton et al., 2000; Rösler et al., 2017). This research clearly suggests that the human attentional bias is at least partially automatic in nature and also independent of low level features (Ristic & Kingstone, 2005; Rösler et al., 2017). On the other hand, there is also considerable evidence suggesting that top-down knowledge can influence perception and attention (Abdel Rahman & Sommer, 2008; for a review see Collins & Olson, 2014), and this can go as far as top-down increase of category specific response in specialised brain structures after participants are exposed to new visual information which helps to disambiguate initially meaningless stimuli (Dolan et al., 1997). The influence of top-down knowledge has not yet been much explored in social attention but for example, feature attribution to the same ambiguous stimulus – namely thinking the stimulus is a car with wheels or a face with eyes – can influence social orienting, by inducing a face-like cuing effect only when the stimulus is thought to be a face (Ristic & Kingstone, 2005). Similarly, this attentional cuing effect was found to be strongest when the viewed stimuli – a face or a robot – were thought to have intentionality compared to just being inanimate objects (Wiese et al., 2012). This research suggests an interesting interplay between automatic processing and semantic knowledge about visual information, strengthening the idea that visual attention and visual perception are not only guided by bottom-up processes related to low-level features of the scene, but also by the relevance of what is being observed.

3.6 Development of social attention

In this section I will not treat in detail the development of social perception, since much of this research has been carried out in infancy or has focused on face perception, and both are beyond the scope of this dissertation. I will, instead, focus on discussing the literature around social orienting in childhood.

Given the developed end state of the adult 'social brain', which is very well equipped to select human information and social categories (Papeo, 2020), it is natural to expect that at least some attentional processes are, since early in life, operative to select human information. As a consequence, a great deal of developmental research in area has focused on infancy in order to determine whether there is a 'social bias' to attention even at the very beginning of life. This research has shown that a bias to social information, at least for some stimuli, is already in place since the first few months of life (Bertenthal & Boyer, 2015; Gliga & Csibra, 2007; Soto-Icaza et al., 2015). This bias potentially allows early social detection and facilitates early social learning, which has been shown to be fundamental for the development of the social brain (Soto-Icaza et al., 2015) and later language skills. Indeed, the detection of faces and eyes, as well as orienting to face-like stimuli and the ability to orient to intact biological motion seem to develop within the first year of life, although refinement of many of these skills continues throughout childhood (for details see Bertenthal & Boyer, 2015; Mondloch et al., 2003; Reynolds & Roth, 2018; Soto-Icaza et al., 2015; Taylor et al., 2004). In particular, since the very first days of life, infants show signs of an ability to perceive faces, as suggested by their gaze following behaviour, and especially of gaze processing, as shown by a preference for direct gaze compared to averted gaze (for a review, Gliga & Csibra, 2007). Thus, the building blocks of social vision and person perception are already in place very early, including a preference for and bias towards social stimuli.

However, research on how social orienting may change across development, and especially during childhood is very limited. Unsurprisingly, research agrees that there is a general preference in typical developing children for social information that is seen across ages (Chita-Tegmark, 2016). It is, however, unclear whether and how these preferences change across the childhood years. We know for example that 9-months-olds, similarly to adults, prefer fixating a face compared to the rest of a social scene (Frank et al., 2009) and during the first year of life they very quickly become good at discriminating different gaze directions (Gliga & Csibra, 2007). Additionally, during the first year of life attention to faces increases rapidly. This increase continues but slows down in childhood before once again showing an increase again in early adolescence (Amso et al., 2014). The handful of studies that have investigated social attention in childhood generally show the presence of a social attentional bias, for example as indicated by a preference in pre-schoolers (2-5 years old) to attend to a face or a dancing human instead of an object or a moving shape in paired preference paradigms (Pierce et al., 2011; Sasson &

Touchstone, 2014). Similarly, at 9-years, children's attention is attracted by a human figure depicted in a scene (Van Der Geest et al., 2002) and this can also act as stronger distractor compared to non-social distractors in the search for non-social targets (Doherty et al., 2019). Interestingly, in this study, although both children and adults were sensitive to the social distractor, children's first fixation was captured more often by the social element compared to the adults, suggesting that children actually have a stronger social bias than do adults. Also Elison et al. (2012) show an age related increase in attention to social elements in an array containing faces/people and objects in a group of 2-18 year olds (Elison et al., 2012).

Furthermore, it seems like younger groups might be more sensitive to the manipulation of social content (e.g., the number of people depicted in a scene) compared to adults (Stoesz & Jakobson, 2014), as children shift their attention from faces to bodies as the number of the people increases, a shift in attention that isn't seen for adults. Although this data would suggest an increase in attention towards social information across middle childhood, Amso et al. (2014) show only a very *mild* increase from 6 to 12 years old in the proportion of social information (in this case faces) that children attended in scenes (Amso et al., 2014). This set of studies suggests that conclusions around sensitivity and attention to social information across childhood, especially in more complex scenarios, are still unclear. Additionally, given all the behavioural and neural changes in social brain and social communication skills happening during adolescence (Blakemore, 2017; Choudhury et al., 2006), extending this research to more thoroughly investigate what skills and tools children have for social understanding to support further and more complex social development as they enter adolescence seems necessary. There are also many aspects of social perception and attention that have not yet received much, if any developmental attention. For example, as they enter school, children's social world begins to extend beyond the family nucleus and their social experience becomes more independent. Both bring many cognitive changes relevant to engaging with and understanding observed social interactions (Carpendale & Lewis, 2004; Eccles, 1999), therefore it seems crucial to understand how attentional processes support and interact with this development.

Finally, studies about theory of mind in development show that implicit mentalizing processes happen already at 18 months of age – e.g. pretend play –, at 3 years old children already start distinguishing between a physical story and a mental state story, and between 4 and 6 years of age children are able to make more explicit inferences, including being able to verbalize the mental reasons behind false beliefs (for a review, U. Frith & Frith, 2003). All this means that some pretty complex social processing beyond simple perception is already taking place in early childhood, but it is unclear how attention supports these processes. Electrophysiology evidence in infancy shows that some neural structures in the posterior temporal regions specialised for face and eye detection are already in place in the first months of life (Gliga & Csibra, 2007; Soto-Icaza et al., 2015). Although the infant brain in general is continuously changing throughout childhood and beyond (Mills et al., 2014), some regions are 'in place' earlier than others. Indeed, cortical thickness in some structures in the social brain (mPFC, pSTS and TPJ) seem to peak around 9/10 years of age and then this cortex, with some exceptions, thins until early adulthood, where it plateaus until late adulthood (for details see Mills et al., 2014; Soto-Icaza et al., 2015). Such results suggest that these 'social brain' structures are developing across childhood before going through refinement and consolidation during adolescence. The exact correspondence between these structural changes and behavioural skills and abilities is not yet clear, however. Some structures involved in social reasoning and social perception are already highly functionally distinct in early childhood (3 years of age) (Richardson et al., 2018), while others continue to specialize throughout childhood and adolescence until adulthood, including FFA (Peelen et al., 2009), STS (Ross et al., 2014; Walbrin et al., 2020), EBA (Ross et al., 2014), and regions involved in theory of mind like the TPJ (R. R. Saxe et al., 2009).

Taken together, this set of research suggests indeed that humans and their brains come furnished with a set of tools and mechanisms that allow early interpretation and understanding of social information, particularly important in predicting other people's actions – something that is important for survival. At the same time, refinement of these mechanisms and the brain regions that support them continues throughout childhood as higher order social inferences and skills become necessary (Soto-Icaza et al., 2015).

3.7 Summary of the section

Human vision and brain are especially tuned and specialized for social information processing, and equipped for a proper understanding and information extraction from the social world. These tools have been very well investigated in infancy, especially for what regards perception of and attention to faces and eyes, but very little is known about how social information is attended throughout childhood, and more importantly, how it is selected in a typically noisy environment, such as naturalistic scenarios. Despite knowing that there is in general a preference for social information compared to non-social stimuli also in pre-adolescent groups, and some changes and specializations happen in the social brain in this period, it is unclear how selection of this information changes compared to adulthood.

The research discussed in this section brings strength to the idea that compared to an adult social brain, the developing social skills system and brain is not yet fully tuned to the social environment, therefore suggesting that some differences in the way children extract and orient towards social information should be expected.

4. Social interactions

In this section I will provide a discussion of the current evidence for a specialized mechanism for social interaction perception and review the limited literature regarding attentional orienting and selection of social interactions in adult vision. I will then also review what we know about the development of social interaction perception, describe why third party encounters may be important in development as well as the work that has been done to investigate this in childhood. Finally, I will discuss some of the evidence suggesting that social interaction perception may be specifically altered in some neurodevelopmental disorders.

4.1 Third party encounters and social learning

Before discussing the research regarding third party encounters, I will describe some evidence as to why social interactions might be important especially in development. As explained above, humans look at other humans, are interested in other people's behaviour and interests, and, importantly, since early in life they learn from other people (Bertenthal & Boyer, 2015; Quadflieg & Westmoreland, 2019; Soto-Icaza et al., 2015). While research on isolated single figures and social stimuli has resulted in important insights about person perception and the social brain, humans in the natural world are often seen while engaged in social interactions, and certainly are always perceived as having the potential to interact with other humans (Papeo, 2020). Visual systems may be specialized to perceive relations between entities more generally (Hafri & Firestone, 2021), but humans seem to be particularly good at people watching and especially tuned to interactions between conspecifics. Social interactions are very rich constructs that involve a wide variety of verbal and non-verbal behaviours, as well as a wide range of social cues that signal or predict such interactive behaviours. In addition, there are a variety of contexts and settings in which interactions occur, as well as a wealth of emotional exchanges and purposes that drive them (De Jaegher et al., 2010). Given the amount of rich information being exchanged during a social interaction, it is intuitive that when we observe social interactions from a third party perspective, we observe important exchanges that convey unique information that can be a crucial source of social learning (Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017; Quadflieg & Westmoreland, 2019). As we will see in a deeper look at past research, humans seem to be biologically drawn to and interested in making judgments and inferences about social interactions, and this might be because other people's interactions are a great source of scenarios where social rules, social dynamics, relationships and intentions of other future interaction partners are applied and demonstrated (Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017). Additionally, although we are not always accurate in impression formation, we automatically use interactive information to draw inferences about other people's personalities and characters by observing them interacting, forming relationships, and navigating in the social world (Quadflieg & Westmoreland, 2019).

Crucially, past research has shown how interactions can be a rich stage for social learning, especially early in life. Some research into observational learning in infants (e.g. Lee & Rutherford, 2018) demonstrates that even young children are pretty good at imitating others after observing them, but as they develop, children are able to extract information also from more complex situations that involve more than just single individuals. In particular, when 4/5 year-olds observed a video of an experimenter exhibiting positive or negative behavioural biases towards as interaction partner, children later expressed preference for (though pointing at their choice) and exhibited prosocial behaviour (giving a toy) towards the person that was favoured by the adults in the video (Skinner et al., 2017). Additionally, in the same experiment, the authors created a group membership manipulation, demonstrating that children were also able to generalize the bias of the experimenter to other people belonging to the same group. What's more, already at 5 years of age, children are able to understand social status (Brey & Shutts, 2015; Over & Carpenter, 2015) and can correctly attribute friendship to two interactants (Nurmsoo et al., 2012) just by observing nonverbal behaviour.

4.2 People watching

As mentioned earlier in the chapter, there is a recent wealth of social perception literature that shows that adults are very skilled at perceiving and detecting interacting dyads, suggesting a potential perceptual bias to perceive social interactions compared to other social targets, such as single agents or not interacting agents. A wide variety of behavioural paradigms and stimuli have been used to investigate this, from impoverished point-light motion stimuli, simplified 3D mannequins, and animated simple shapes to naturalistic figures of whole humans, together with the manipulation of different social cues (for a review, Quadflieg & Koldewyn, 2017).

To start with, the human ability to extract communicative information from very impoverished stimuli it is well established. Indeed, while Bente et al. (2001) show that mental attributions are similar for real life interacting individuals and 3D animations of the same individuals (Bente et al., 2001), Manera et al. (2010) show that we are able to recognize communicative actions and interactions from simple brief presentations of point-light agents, suggesting that human vision might not need much visual information to detect or recognize a social interaction from a stimulus. Point-light agents have also been used to show that the visual system preferentially selects a meaningfully interacting dyad to reach awareness, when the two eyes are presented with competing information (Su et al., 2016). This would suggest that visual sensitivity to social interactive information may happen in an automatic and reflexive way and in response to minimal information. Indeed, the recognition of an event happening between two actors (e.g. kicking or pushing) in a naturalistic picture, can happen in a very brief period of time – seeing such a picture for just 73ms was enough for participants to be able to recognize both the action and the actors' roles (agent vs patient) in the scene (Hafri et al., 2013).

Facing direction is a particularly important cue for determining whether two agents are interacting or not. Moreover, it is a cue that is relatively simple to manipulate and thus has been widely used in work probing social interaction perception. In particular, we seem to have visual sensitivity to pairs of bodies (with and without the head depicted) facing each other compared to agents not facing each other in detection and recognition tasks, an effect that does not apply to pairs of facing objects or humans facing objects (Papeo et al., 2017; Papeo & Abassi, 2019). The facing direction of a partner can also improve emotion recognition. In particular, the emotional expression of one face is modulated by the emotion of the another face only when the two were facing each other but not when they were facing away (Gray et al., 2017).

Another cue that has been investigated in social interaction research is the synchrony between the actions of two dynamic figures. Indeed, in the attempt to disrupt the perception of social interaction between two point-light fighters, Neri et al. (2006) manipulated different levels of synchrony between the two agents. They showed that when presented with one unobscured fighter and one embedded in noise, it was easier to correctly infer the presence of the second agent when the two actions were synchronised than when they were not, suggesting that the synchrony typical of a social interaction facilitates person and action perception (Neri et al., 2006). Additionally, it has been shown that when fluency – the smoothness of action – and contingency of movements between two interactants (two kinematic characteristics typical of social interactions) were disrupted, participants rated the actions as unnatural (Georgescu et al., 2014). With a similar approach, Manera et al. (2011) used dyads of point-light agents and noise masking to investigate the processing of communicative behaviour. Indeed, they show that when an agent was paired with another portraying a "coherent" communicative gesture, the latter was easier to discriminate among visual noise, compared to a non-communicative pairing (Manera et al., 2011). Furthermore, two shapes moving in a contingent, synchronous, and mutual way are often seen as having a social interaction, which participants providing rich narratives about the perceived encounter (Castelli et al., 2000; Santos et al., 2010; J. Schultz et al., 2005; R. T. Schultz et al., 2003). Indeed, we can distinguish between different types of interactions between shapes (Isik et al., 2017; Walbrin et al., 2018), suggesting human attention and vision is strongly tuned to capture this kind of information even from simple motion and action cues, and even when human information is missing. What's more, it has been shown that two interacting agents seem to be processed and perceived as more than the sum of the two agents, and are grouped more as a unique chunk of meaningful information (e. g. Adibpour et al., 2021; Ding et al., 2017) than as separable objects that must be individually recognised and processed.

This research shows clearly how cues typically indicating social interactions enhance and facilitate social perception and attention, supporting once again their high biological value.

Very few studies have investigated attentional orienting to social interactions, and even fewer paradigms have manipulated interactive content across stimuli. One paradigm that has been used to investigate visual attention to interactions are visual search tasks, where there seems to be a search asymmetry for upright body dyads, meaning that facing bodies are found faster and with higher accuracy when searched for among non-facing bodies than when non-facing pairs are searched for among interacting dyads (Papeo et al., 2019). Facing dyads seem to have a perceptual and attentional advantage (Papeo & Abassi, 2019; Vestner et al., 2019) when they are looked for among same facing direction pairs, although some researchers suggest that this attentional advantage for facing dyads might be produced by an attentional cuing hot-spot effect that is created by facing agents compared to non-facing figures (e.g. Vestner et al., 2020, 2021). Interestingly, while facing direction does not work in this way for some objects (i.e., chairs; Papeo & Abassi, 2019) a similar advantage is seen for facing pairs of objects when those objects direct spatial attention in a Posner-style attentional cueing paradigm (Vestner et al., 2020, Vestner et al, 2021). In visual search tasks, then, attentional cuing mechanisms may partially explain the search advantage for social interactions. However, another finding suggests that interactions do not only have an influence on attentional orienting/capture but also on the engagement of attention. In a pilot study later used with ASD and TD children, typical adult participants saw pairs of real facing or non-facing agents at the same time on a single grey background, in a 10 seconds free viewing paradigm. Despite having more than sufficient time to fully explore both dyads, participants spent significantly more time looking at the facing pairs compared to the non-facing pairs (Stagg et al., 2014). Furthermore, although not naturalistic, this is the only study in our knowledge to have shown a preference in looking times for social interactions compared to non-interactions in a multiple people scenario.

However, most of these studies have investigated attention towards social interactions using tightly controlled schematic stimuli, and the few studies investigating social attention in naturalistic scenes have not explored the role of dyadic social interactions compared to non-interacting dyads. To our knowledge there is only one study manipulating the social content and the interactive element in observed scenes, with the aim to investigate social attention in more complex scenes, and in particular attention to the eyes. During a 15 seconds free viewing paradigm, Birmingham et al. (2008) used scenes depicting either one or 3 people, either being active (doing something independently), inactive (doing nothing independently) or interacting (the 3 people in the scenes were doing something together). Interestingly, the authors show that attention to the eyes increases when the number of agents involved increased, but they show no difference in the fixations proportions across the AOIs (eyes, heads, bodies, foreground and background objects) between the scenes depicting 3 people interacting and the “independently active” scenarios, suggesting no influence of the presence of an interaction on attention within the scene (Birmingham et al., 2008). In contrast, Kujala et al., 2012 show that when adults freely viewed pairs of humans interacting in a friendly manner or facing away, their attention was driven to the heads (rather than bodies) more in the facing condition compared to the facing away condition (Kujala et al., 2012) Although attention specifically to the head vs body is not a measure used in the current work, this research is worth mentioning because it suggests that social attention patterns change when a social interaction is introduced into the scene, but also that it might be attention *to the* dyad where these differences occur.

The neural basis of social interaction perception has been investigated with a variety of paradigms, methods and stimuli (Quadflieg & Koldewyn, 2017). Indeed, increasing evidence suggests the posterior superior temporal sulcus (pSTS) as a central hub for the visual perception of social interactions (Isik et al., 2017; Quadflieg et al., 2015; Walbrin et al., 2018; Walbrin & Koldewyn, 2019), and not only does it distinguish between interacting dyads and non-interacting dyads or single agents, but also between different kinds of social interactions (Walbrin & Koldewyn, 2019), even when explicit social inferences are not a required by the in-scanner task (Walbrin et al., 2018). Additionally, it seems like the pSTS region might be tuned also for extraction of interactive content in social scenarios without explicitly human content (J. Schultz et al., 2005, Isik et al., 2017, Walbrin et al., 2018).

Although the pSTS has been shown to be selective in the visual perception of interacting dyads compared to pairs of independent actors when participants were not performing specific tasks, other structures are recruited, depending on the experimental task utilised. Even when using more naturalistic scenarios and free viewing paradigms it is shown that a set of structures involving the amygdala, dorsomedial prefrontal cortex, IFG, aSTS and precuneus are recruited for the processing of relations between humans even when no specific task is performed (Iacoboni et al., 2004; Kujala et al., 2012; Wagner et al., 2016). Simlarly, during categorization of point-light displays as acting together or independently, a wide network was recruited including the pSTS, TPJ, anterior superior temporal sulcus (aSTS), dorsal mPFC, and inferior frontal gyrus (IFG) (Centelles et al., 2011). Additionally, the IFG, STG and pSTS among other areas, were more activated when participants viewed videos of contingent interactions compared to a pair composed of an agent and its mirrored version (Georgescu et al., 2014). Furthermore, it seems for example that activity in the EBA, a region usually involved in processing of bodies, is increased when the bodies are facing compared to when they're not facing, similar to what happens in LOC when functionally related objects are presented (Abassi & Papeo, 2020), suggesting that in the human visual structures two bodies facing each other might be processed as a whole unit, similarly to how bodies are perceived as more than the sum of different body parts. Similarly, when participants viewed different pairs of full-body interactions, not only did the EBA show, similarly to pSTS, an ability to distinguish between different kinds of social interactions, but the activity in this region showed a dyadic information effect (Walbrin & Koldewyn, 2019). This research would suggest that the EBA might contain unique dyadic information which could result from processing socially interacting dyads as a whole, and not simply as just the sum of the agents involved in the interaction. Therefore, it seems like social brain structures involved in social interaction perception, person perception, and action perception, are not only sensitive to a series of physical cues between interactants, but are able to pick up on more complex social information and meaning from the dyads.

To summarise, research on the brain representations of social relations between humans has shown a wider network than the simple processing of single faces or bodies, such as dmPFC, IFG, aSTS, EBA, TPJ, pSTS depending on the complexity of the task or the stimulus used. This research suggests that social interactions from third person perspective have a special status and role for human vision and in the human brain, therefore it is natural to suspect that some attentional privileges might derive from this special status.

4.3 People watching across development

Despite the above discussed research showing a special place for social interactions in the adult social brain and vision, there is relatively little research explicitly investigating change in attention and visual perception of social interactions or more complex social information across development, especially during middle to late childhood and adolescence. What little evidence there is suggests that the ability to discriminate between interacting and non-interacting humans develops as early as 4 months old, with infants becoming capable of shifting their attention between speakers that face each other compared to non-facing speakers, and this ability seems to increase across the first year of life (Augusti et al., 2010; Handl et al., 2013). What's more, at 6 months, infants can discriminate between different types of interactions, by showing a preference for a pair of agents exhibiting helping behaviour, compared to those that hindering each other (Hamlin et al., 2007). This suggests an early ability to evaluate and understand complex interactive behaviour. At 4 years of age, children are able to explicitly distinguish a social interaction from a pair of agents acting independently side by side from a set of impoverished stimuli like point-light actions, although this ability doesn't seem to be adult-like until after 7 years of age (Centelles et al., 2013).

Although this research suggests that the ability to infer complex social information from impoverished stimuli seems to be present in children as young as 4 years old, as explained earlier in the chapter, the aim of the work in this dissertation is to investigate attentional orienting to more complex information, such as social interactions embedded in naturalistic scenes. Therefore, even though the scarce literature would suggest that the ability to make higher order social inferences might be already 'adult like' by 9 years of age, it is unclear how spontaneous attentional capture might be influenced by the complexity of the social scene and the presence or absence of social interactions. To our knowledge, there is only one developmental study that has investigated attention to interacting pairs compared to non-interacting pairs. The results show that typically developing children of approximately 9 years of age looked for longer at 2 agents facing each other when presented together with another pair of agents that did not face each other (Stagg et al., 2014). Crucially, even though this study suggests an attentional bias to social interactions also in pre-adolescent childhood, in this study age related differences in this pattern of attention were not investigated, and the figures were presented isolated from any background context, which might provide important extra information that could either facilitate processing of social interaction or, on contrary, hinder such mechanisms because of the increased complexity and competition for attentional resources.

When investigating preferential looking between a social scene containing social interactions and a non-social scene (all scenarios depicted moving geometric shapes), Shaffer et al. (2017) show a preference for the social scene in 5-17 years old children, but they don't find any change in looking behaviour with age (Shaffer et al., 2017). However, in this research, the presence or the absence of interactive information has not been explicitly manipulated and the paradigm is more a measure of preference for social information compared to non-social information and not necessarily the attentional capture of and engagement to social interactions compared to non-interactions when both are portrayed by social targets. Similarly, although they have not explicitly manipulated the interactive content in the scene, Hanley et al. (2012) show that differences in social processing between TD and Asperger groups are evident only when a social scene becomes more complex, for example when it depicts naturally occurring dyadic interactions, compared to only isolated faces (Hanley et al., 2013). This research increases the suspicion that if adding interactive content to a scene can increase differences seen between typically developed groups and clinical populations, similar manipulations could reveal developmental differences between different age-groups in social attention to interactive scenarios.

Although behavioural developmental research on the topic of social interaction perception is limited, neuroimaging research would suggest that social brain structures specialised for social interaction perception are still undergoing tuning for fine perception of social interactions in pre-adolescent childhood (Sapey-Triomphe et al., 2017; Walbrin et al., 2020). Indeed, when 6-year-old children were showed videos of point-light displays depicting social interactions and independent actions, although the right pSTS was activated by the interaction condition in both groups, the selectivity of this region for interacting dyads did not become similar to the adults until around 9 years old in the right hemisphere, and was not adult-like in the left hemisphere (Walbrin et al., 2020). Furthermore, the selectivity to bodies in EBA and to faces of the face area of the STS did not show this developmental change, suggesting that although the neural structures supporting social interaction perception might be already sensitive in childhood, they are not yet to fully tuned for this purpose. Similarly, Kirby et al. (2018) show also that although the pSTS responds also to stimuli of single biological motion agents (e.g. walking or painting), compared to scrambled stimuli, there is no clear developmental change in response to simple biomotion walkers in 7-13 years old children (Kirby et al., 2018), suggesting that documented age-related differences in the pSTS might be specific to information needed to specifically decode and perceive social interactions.

Interestingly, brain research shows that not only does the number of neural structures involved in the perception of interacting dyads seem to increase from childhood to adulthood, but it appears that in childhood the areas recruited were more visual perception areas, while adults recruit a wider network, including more areas involved in mentalizing and higher social inferences (Sapey-Triomphe et al., 2017; Walbrin et al, 2020). This research suggests that from childhood to adulthood there is a process of refinement of the social brain functionality towards perhaps a deeper understanding and meaning extraction from the social world. Not only does the sensitivity to social interactions change across childhood in the pSTS, but also mentalizing networks (including TPJ, precuneus, dorsal, medial and ventromedial PFC) which are recruited when making higher order social inferences, undergo functional specialization between 3 and 12 years (Richardson et al., 2018).

4.4 Neurodevelopmental research

In this final section I would like to mention some of the neurodevelopmental research on social attention, suggesting other reasons why attention to social interactions deserves further investigation.

It is widely established that social attention skills are disrupted in some neurodevelopmental disorders (Williams syndrome, ASD) in quite predictable ways. Indeed, it is generally known that ASD children orient their attention less towards social information compared to their typically developing peers. In particular, thanks to a wide variety of tasks and methods it is generally agreed that autistic individuals show decreased spontaneous attention towards the eyes of faces/humans in eye-tracking paradigms (Chita-Tegmark, 2016; Elison et al., 2012; Frazier et al., 2017; Papagiannopoulou et al., 2014), and importantly, in naturalistic scenes (Klin et al., 2002; Riby & Hancock, 2008; Williams et al., 2013). However, the nature and extent of these behavioural or attentional differences appears to vary substantially depending on the complexity and type of information being presented and on the task that individuals are asked to perform. Indeed, as shown in a recent meta-analysis, it would seem that even though autistic participants show overall reduced social attention through a variety of experimental stimuli and tasks used, compared to TD groups, the only factor that could reliably predict the effect size of this reduction was the amount of social content in the scene (e.g., the number of people depicted in the stimuli; Chita-Tegmark, 2016). Another meta-analysis suggests that the largest differences between TD and ASD groups in attention to social areas of interest like faces and eyes were driven by the presence of a social interaction in the scene (Frazier et al., 2017). Additionally, when Hanley et al. (2012) investigated attention to isolated faces and naturally occurring dyadic interaction scenes, differences in orienting to social information between the control group and the group with Asperger's syndrome occurred during only when viewing more complex stimuli, such as naturalistic social interactions; differences between groups were not evidence for isolated faces. Similarly, when showing arrays of still people and objects, moving faces and objects, and videos of children involved in a playful social interaction, the differences in social attention patterns between the TD and the ASD group became evident only when social interactions arrays were shown, suggesting that it is not only the motion or dynamic information in video stimuli, but the complexity of social information/cues conveyed by social interactions that actually distinguishes accurately between these two groups (Chevallier et al., 2015). Together, the evidence suggests that autistic individuals may attend to social interactions differently, raising the possibility that interactive content may be able to reveal developmental change in social attention as well as differences between typically developing and neurodiverse groups.

4.5 Summary of the section

Social interactions have been shown to be important theatres of human social exchange, and observing them from a third party perspective can provide an important source of information for social learning. The human brain seems to be highly tuned to picking up the unique cues typical of social interactions, but it is unknown how attention supports these processes. Some research suggests a preference to attend social interactions compared to non-interacting individuals but more research is needed. Additionally, even less is known about development. Children's brains are not yet fully specialized for the perception of interacting dyads, but somehow children are able to promptly learn social norms and rules from observation of interactions, suggesting attention might help selection of important interactive information from the environment. Additionally, social attention in demonstrably different in autism, though only reliably when scenes are socially complex, including when they contain social interactions, providing another reason for further investigation in this field.

5. Aims of the work and overview of thesis chapters

The aims of this work are, in the first place, to investigate how visual attention to social targets is altered in the presence of a social interaction, and to investigate whether attention to interactions changes during pre-adolescent childhood. Additionally, we aimed to investigate *spontaneous* attention to interactions by using free-viewing paradigms and naturalistic scenes. Therefore, in the first experimental chapter, we investigate attentional patterns – as measured by eye movements - in naturalistic scenes containing either an interacting dyad or a pair of humans who are not interacting, in both adults and in children aged between 6 and 12. A second aim in the current work is to investigate the strength of any social interaction bias by manipulating the social content – i.e. number of people – in scenes as well as pitting social interactions against other social targets in the same scene. In the second experimental chapter therefore, dyads and other social agents in the scene are competing for attentional resources. Again, we also investigate developmental change by collecting data in both typical adults and children (6-12 years old) and adults. Lastly, it is unclear how top-down knowledge about social interactions changes across development, and such knowledge may influence attentional patterns in naturalistic scenarios. Therefore, in the third experimental chapter, we investigate social attention and its development in both adults and 6-12 year-old children in ambiguous social scenarios, where a dyad might be considered interactive or not depending on the viewer's internal construct of social interactions. We investigate whether children and adults perceive these scenes differently, and whether the way in which a scene is categorised influences how participants attend to social information in the scene.

Therefore, the novelty of the work here presented is first of all in the clear investigation of attention to social interactions in naturalistic scenarios in pre-adolescent childhood, in the use of a wide variety of real life settings, and in the investigation of the role of social interactions for human attention when there is competition between different social targets and when the information presented is ambiguous. Finally, outside specific questions around the processing of social interactions, this work should provide additional information around social orienting and social scene processing in childhood.

Chapter 2. General methods

1. Overview of the chapter

In this chapter I will outline and justify the main methodology, design details and statistical approach used in the three empirical chapters. I will start by describing the samples used for this research, then explain the research design and its suitability for the research questions and provide a brief theoretical introduction to eye-tracking methodology. I will continue with a detailed description of the stimulus choice procedure for each of the three experiments and finally, I will describe the multilevel modelling analysis approach used in this work.

2. Participants

To address the outlined research questions, the initial plan was the recruitment of an adult and a developmental sample, including children 6-12 years of age and teenagers 13-17 (see Appendix A for pre-registrations of each experiment). Unfortunately, the COVID-19 pandemic interrupted data collection from both the child and adolescent sample. In fact, we were scheduled to start testing in the first secondary school the week of the first lock-down. Although we waited and hoped, we were not able to begin testing in schools again, nor recruit young participants to be tested in the University setting. Thus, the final sample included in the dissertation was not the one we planned for. We recruited, therefore, a total of 54 children 6-12 years old (M = 8.76, SD = 1.72; 28 females) and 101 adults (M = 21.75, SD = 5.29; 70 females, 1 other) but do not include any adolescent data. All participants took part in all three experiments. Indeed, although the experiments were designed separately, we designed the paradigm to facilitate data collection so that we could collect data for all three experiments in the same session. Data from all child participants are included in the analyses reported in the three empirical chapters, while 3 adult participants were removed from all analysis, 2 because they were older than our targeted 18-35 age-range, and 1 because of sleepiness-induced poor eye-tracking data quality. The final pool of adult participants was composed of 98 adults (M = 21.15, SD = 2.97; 70 females, 1 other).

Adults were mostly recruited through the university participant pool (SONA), but we also recruited through advertising on social media in the community. For the most part, children were recruited through a local primary school, but three were also recruited through word of mouth.

3. Research design

All the research questions in this work verted around the investigation of spontaneous gaze behaviour in response to complex scenarios depicting social interactions and investigated what, if any, developmental changes in such processes might take place during childhood. Therefore, the aim was to measure attentional orienting mechanisms in such experimental conditions.

All three experiments involved collection of eye-tracking data during a free exploration paradigm where no other task was required. As shown in past work, eye movements are very sensitive to the task participants are performing (Fletcher-Watson et al., 2008; Yarbus, 1967) so although there are a multitude of experimental paradigms that could allow more specific probing of attention to social interactions, in this work we first hoped to probe how social interactions are naturally attended to in the absence of any overt goal.

There are any number of potential measures that can be extracted from eye-tracking data but in trying to establish spontaneous gaze-behaviour to social interactions in complex scenes, we used a constrained set of well-established measures: dwelling time to measure engagement and time to first fixation to measure attentional capture. These two measures are the most commonly used measures in eye-tracking research of social attention.

The experimental manipulation in all experiments was through the choice of stimuli, in particular through changing the number of people in the scene as well as the ambiguity of the social content in the scene. This type of manipulation allowed us to measure 'spontaneous' attention while also testing specific hypotheses and is also easy to implement with a developmental sample.

The choice to use naturalistic stimuli was made for two different reasons. First, using scenes that are more similar to 'real life' takes the current work one step away from most of the research on attention to social interactions (Papeo et al., 2019; Stagg et al., 2014; Vestner et al.,

2019) which has used the facing direction of dyads to imply social interactions isolated from any background or context. Second, as our aim was to probe spontaneous attention to social interactions as might occur in a real life scenario, we chose to use complex, cluttered and varied scenes. Furthermore, a great deal of research has shown the benefits of using naturalistic stimuli for attentional research, as they provide context to the social information being viewed (Birmingham et al., 2009b; Smilek et al., 2006). Additionally, some research on groups with neurodevelopmental disorders shows that patterns of attention to social information change when the scenes become more cluttered and complex, rather than when social information is represented through isolated figures (Chevallier et al., 2015, 2016; Chita-Tegmark, 2016; Hanley et al., 2013).

It is important, however, to acknowledge the costs of using such varied and complex stimuli. Using such stimuli greatly increases the difficulty of controlling low-level visual features and saliency and being certain that the conditions are matched on non-social factors that could potentially drive attention in scene exploration (Itti, 2005; Itti & Koch, 2000). One way we addressed this was to colour match all stimuli with one sample stimulus (see below, section "5.1 Common procedures"). We also chose scenes specifically to represent similar locations and items across interactive and non-interactive scenes, and that scenes across conditions were roughly matched on the size of human AOIs and that the distance between interactors was similar to that between non-interactors. Additionally, much prior work has shown that attention to social information overrides visual saliency driven attentional mechanisms (Birmingham et al., 2009a; End & Gamer, 2017; Langton et al., 2000; Rösler et al., 2017) and that saliency can't predict early attention in free-viewing paradigms (Holmqvist et al., 2011, pp. 459). As our hypotheses and research questions are focused on the *social* information in scenes, we feel (relatively) certain that our analyses are not much impacted by low-level visual features of these scenes used across the three experiments. Finally, we relied on the sheer diversity of visual content across the scenes we used. All stimuli are extremely different from one another (see Appendix for details), so much so that it would be very hard to find any one single factor that might drive attention other than our experimental manipulations.

4. Eye-tracking methodology

4.1 Introduction to eye-tracking

The surrounding world is very rich in information but the human eye can only access relatively small portions of information in any detail at any one instant. Indeed, the fovea - the region of the retina with the highest resolution - encompasses only approximately 2 degrees of visual angle, and the further from the fovea information falls on the eye, the lower the resolution and acuity of the retina. Eyes therefore have to move constantly to shift the fovea to different positions in the visual scene in order to take in the details of a complex scene. In this way, a sequence of images and information is constructed on the retina. Such movements are one way in which visual information is selected by the human eye for further processing (Liversedge & Findlay, 2000; Rayner & Castelhano, 2008).

Eye-tracking is an experimental tool used to record eye movement and position over time, and is typically used to assess distribution and designation of visual attention on a visual display.

The type of eye-tracking used in this work is a pupil-and-corneal reflection method, where data is collected at a 1000 Hz sampling frequency – meaning that the device is recording eye position 1000 times per second. Infrared light is shone towards the eyes, and reflected by the cornea and collected by a camera. This reflective information, together with the relative position of the pupil, gives the software the needed information to calculate the position of the eye over time (Holmqvist et al., 2011, pp. 39; Figure 1 for schematic representation of the process). The quality of the data with this method is very sensitive to dry eyes, make-up (especially mascara), and glasses (Holmqvist et al., 2011, pp. 151-152). The remote set-up used in this work, furthermore, uses a sticker marker to track the head, which increases the ability to accurately track eye movements without the use of a chin rest, which makes the device very suitable for developmental research.

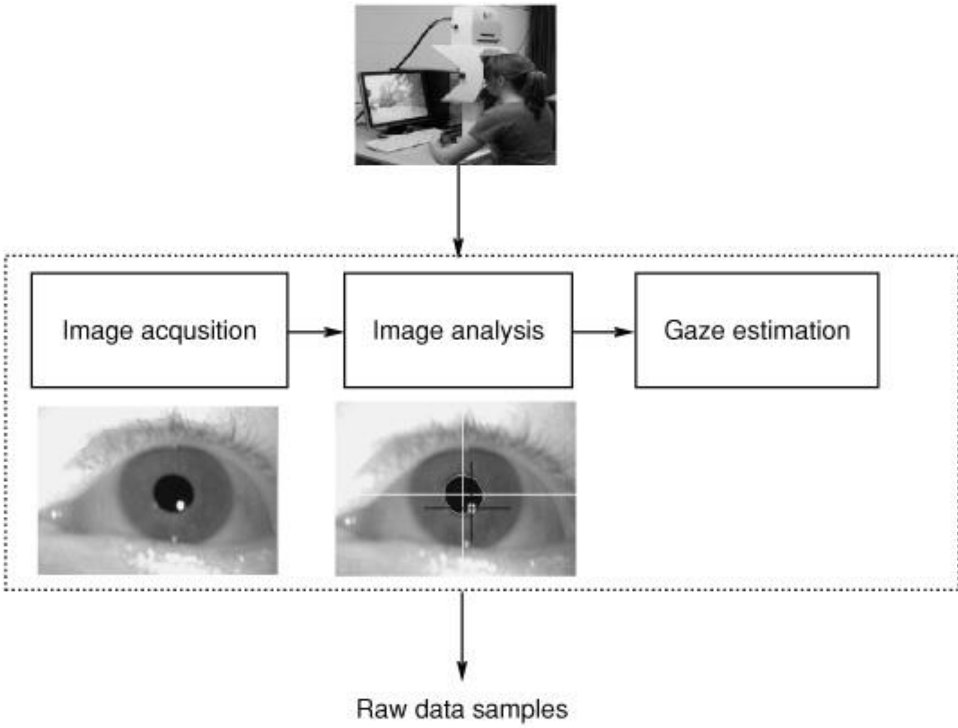

Figure 1. Schematic overview of a video-based tracking process. Figure from Holmqvist et al. 2011, pp. 44.

One of the most common gaze events that most eye-tracking measures rely on is the fixation, which represents the state of the eye staying “still” for a certain amount of time, which can last from a few tens of milliseconds to 200-300 ms. In reality, even during ‘fixation’, the eye is not fully still, but continues to make small movements such as tremor, drift, and microsaccades, typically considered noise in eye-tracking analyses (Duchowski, 2007, pp. 46). During fixation, the eye collects visual input from one section of visual space, with detail information available only from where the fovea rests. When repositioning the fovea, the eye moves rapidly from one position to the another, a movement that is called a saccade. These quick movements can last anywhere from 30 to 80 ms (Holmqvist et al., 2011, pp. 41), can be voluntary or reflexive, and represent a time during which the eye cannot receive new visual input. They are pre-programmed and once the saccade has started it cannot be interrupted (Duchowski, 2007, pp. 42).

4.2 Visual attention, cognitive processing, and eye-movements

The eye-mind hypothesis states that what is fixated is what is being processed (Just & Carpenter, 1980). Indeed, while visual attention anticipates eye movements by approximately 250 ms (Deubel, 2008), and attention and eye location can be separated (Michael I Posner et al., 1980), in a complicated scene it takes an executive effort to do so (Rayner & Castelhano, 2008). Therefore, attention and eye movements are tightly linked (Deubel & Schneider, 1996; Rayner, 2009), suggesting that when a viewer moves the eyes to a certain point in space, attention has just been allocated to that same position in space. This makes eye-tracking methodology a very useful technique for the investigation of attentional processes and even more so in development, since it is a non-invasive measure (Hessels & Hooge, 2019). This is particularly true when there are no task demands that encourage participants to attend to aspects of the scene covertly, for instance performing a task at a central fixation while pictures are shown in the periphery. In a free-viewing paradigm, like the ones used in this dissertation, participants have no reason to separate visual attention from eye-movements and eye-movements are likely a good indication of what participants are attending to and processing in a scene.

In this work two measures were used: dwelling time and time to first fixation to each area of interest (AOI). Dwelling time was defined in the Data Viewer Software (SR Research, 2013) as the time spent gazing within an AOI, including both fixations and saccades, in milliseconds (ms). We used dwelling time as a measure of attentional engagement with either social or non-social information in the scene. This measure can also indicate interest in a piece of visual information (Holmqvist et al., 2011, pp. 406).

Time to first fixation or “entry time” into an AOI is measured in milliseconds, relative to the start of the trial, and is the first moment the eye enters the area of interest. This was used as a measure of attentional capture by each AOI, indicating that the shorter the time to first fixation the faster the gaze entered that region of the scene. Additionally, generally is thought that shorter entry time indicates a higher ability to locate the piece of information of interest (Holmqvist et al., 2011).

5. Materials

In all three experiments we used naturalistic pictures, depicting humans in a variety of settings and actions. For experiments in chapters 3 and 5, the pictures were selected and rated in the same session by the same participants, while stimuli used in chapter 4 were selected and rated in a separate session, by a different set of raters. In this section I will first outline the common stimulus selection procedures, and then proceed with details on each set of pictures (see Appendix B for full set of final pictures).

5.1 Common procedures

The rating procedure aimed to select the pictures to be used for each separate research question and each separate experiment. Therefore, after selecting a larger set of pictures from several on-line databases, pictures underwent rating for the level of interactiveness of the agents depicted, and to obtain an agreed judgment over which people in the scene were involved in interactions, if any. This rating was necessary to accurately assign pictures to the various conditions (interacting, non-interacting or ambiguous) and to have a clear consensus around the interpretation of the depicted social situations, except in the case of the ambiguous pictures (chapter 5).

For all three experiments we selected pictures that met the following criteria:

- Emotionally neutral;
- Depicting a variety of ordinary life contexts (e.g., schools, library, market, shops, cafes, public transport, bus stops);
- Depicting 2, 3 or 4 people (or easy to modify to depict that number of people (e.g., cutting the picture to eliminate an 'extra' person in the periphery of the scene));
- Not depicting people who directly face the camera or appear to be interacting with "off-screen" agents;
- Depicting a variety of races, sexes and age-groups;
- Containing, as much as possible, people's entire body (e.g., not partially cropped from the scene).

After the initial choice of the scenes to undergo rating, each picture was subjected to a Photoshop (version CC 2019) routine containing the following steps:

- Neutralize to remove colour cast: “Adjustments – Match color” and selected the “neutralize colour” option;
- Automatic colour match based on a picture sample using the “Adjustments – Match colour” toll in photoshop where the image below (Figure 1) was used to match colour schemes across all pictures in the stimulus set;
- Sharpen: “Filter – Sharpen – Sharpen”.

Figure 1. Sample picture used for the photoshop routine. From SUN database (Xiao et al., 2010).

After pre-processing with Photoshop, the selected pictures were resized to 400 x 400 px to allow easy presentation in the PsyToolkit platform (Stoet, 2010, 2016), which was the software we used for the rating survey.

5.2 Picture selection for chapter 3 and chapter 5

For these two chapters the rating survey had the aim to find pictures depicting two people either having a clear social interaction, acting independently, or in an ambiguous social scenario (where there was no clear consensus on whether the people were having a social interaction or not).

Pictures

The stimulus pictures were selected from the on-line database SUN (Xiao et al., 2010) with the criteria mentioned above. The initial set of pictures was composed of 353 pictures, which was then narrowed down by quality of the pictures, size of the human figures, diversity of setting, and the ability to resize them to the desired configuration (i.e., depicting 2 people). The final set of pictures which underwent rating was composed of 190 pictures.

Procedure

The rating survey was performed on-line through the Psytoolkit platform (Stoet, 2010, 2016). Participants were presented with one picture at a time and asked to answer the question "How interactive is this picture?" on a slider Likert scale from 1 ("not interactive at all") to 7 ("very interactive"), with the mark initially set on 1. Pictures were presented in a randomized order and in a three-block structure to allow breaks.

Participants

Twenty-six university students ($M = 24.77$, $SD = 4.03$; range: 21 – 33; 16 females) took part in this rating survey. Participants were recruited through social media and the university's participant pool system, and were given credits for their participation. Every participant gave informed consent before proceeding to the experiment, and all procedures received ethical approval at Bangor University (ethics protocol number: 2018–16360).

Results

The average interactiveness score for all pictures across all participants was $M = 3.57$, $SD = 1.40$.

Pictures were then ordered in ascending order and we created 3 categories based on the ratings: the top 33% ($M = 5.17$, $SD = 0.58$, range: 4.35 – 6.46, $N = 62$) were assigned to the interactive condition, the lowest 33% to the non-interactive condition ($M = 1.94$, $SD = 0.46$, range: 1.12 – 2.72, $N = 62$) and the middle 33% to the ambiguous condition ($M = 3.59$, $SD = 0.45$, range: 2.8 – 4.32, $N = 66$). Pictures then underwent a final selection step to take into account the inter-rater agreement, which was calculated by assessing the percentage of scores for any particular photo that agreed with the decided category.

For the final set we selected 30 pictures for the interacting condition (*M* = 5.40, *SD* = 0.54, range: 4.77 – 6.16, agreement cut-off: *M* = 77.28 %, *SD* = 12.21 %), 30 pictures for the non-interacting condition (*M* = 1.86, *SD* = 0.40, range: 1.24 – 2.26, agreement on cut-off: *M* = 78.55 %, *SD* = 12.01 %) and 30 for the ambiguous social scenes experiment (*M* = 3.61, *SD* = 0.32, range: 2.96 – 3.61, agreement on cut-off: *M* = 35.74 %, *SD* = 9.16 %) (see Appendix XX to view picture set for each experiment). Importantly, the final set of pictures was very heterogeneous in terms of settings (e.g., schools, public transport, public markets, shops), and depicted ages and races.

5.3 Picture selection for chapter 4

Here, we selected pictures depicting three and four people to allow investigation of attentional competition between interacting dyads and other social attention targets in naturalistic scenes. Therefore, the chosen pictures underwent rating for the level of interactiveness of the scene and to inform decision which two people were part of an interacting dyad in the scene, if any.

Pictures

The stimulus pictures were selected from 4 on-line databases: "Recognizing Indoor Scenes" (Quattoni & Torralba, 2009), "Discovering Groups of People in Images" (Choi et al., 2014), "EMOTIC – Emotion Recognition in Context" (Kosti et al., 2017) and SUN (Xiao et al., 2010). We used the same criteria listed above to guide our initial selection of photographs with the additional criteria that ~50% of photos included a clear social interaction (by our estimate, at least) while the other ~50% did not.

Procedure

A total of 114 pictures were selected, 64 of which depicted 3 people and 50 of which depicted 4 people. Each picture was subjected to the Photoshop routine outlined above to obtain a uniform colour scheme across all the pictures in the database, and pictures were again resized to 400 x 400 px. We used Psytoolkit (Stoet, 2010, 2016) to present the rating survey to participants.

The pictures were judged for the level of interactiveness of the scene first, and then participants were asked to indicate the interacting dyad in the scene, if any.

To allow participants to indicate which people were involved in an interaction, for each picture, we used GNU Image Manipulation Program (The GIMP Development Team, 2019) to apply numbers next to each human present in the scene, from left to right, to allow easy choice of the interactive pair, if any (Figure 2 for an example).

a. b.

Figure 2. Example of original picture (a) and numbered picture (b) presented in the survey.

Pictures were presented in a randomized order and in a three-block structure to allow breaks and make sure everyone would complete the experiment. For each picture, the participant viewed first the original picture, without numbers, and was asked "How interactive is this picture?" where the response was on a 1 ("not interactive at all") to 7 ("very interactive") slider Likert Scale, with the marker initially placed at 1. Immediately after they registered a response, participants viewed the numbered version of the same picture and were asked "If there are people having a social interaction in this scene, which pair is it?" where the response was a choice among options containing each possible two-person combination of the people in the scene, together with the option of "no interaction at all", "everyone is interacting" and "3 people are interacting" (for the four-people scenes).

Participants

A total of 27 adults (M = 28.48, SD = 6.38; range = 21-50; 17 females) took part in this survey. Participants were recruited through social media and the university's SONA system, and given university credits for their participation. Every participant gave informed consent before proceeding to the experiment, and all procedures were approved by the ethical committee at Bangor University (ethics protocol number: 2018–16360).

Data analysis and results

The data from each question across all pictures were analysed separately and both taken into consideration later to inform the final choice of stimuli for the eye-tracking experiment.

a. *Interactiveness judgment*

For each picture, the average rating was calculated across all participants. The average interactiveness across all pictures and participants was M = 3.21, SD = 1.54. The pictures were then ordered in ascending order, and the top 33% of the pictures were temporarily assigned as "interactive", while the bottom 33% were assigned as "non-interactive" (see Table 1 for details). For each picture, inter-rater agreement was calculated by assessing the percentage of raters that rated any particular photograph as being in the "agreed" category (i.e., below the cut-off interactiveness score for non-interactive pictures and above the cut-off interactiveness score for interactive pictures).

Table 1. Descriptive statistics, cut-off and distribution of the pictures based on interactiveness ratings, with N as the number of pictures in the specific category.

Condition	People	Average rating	N	Cut-off	Agreement on cut-off
Non- Interactive	3	M = 1.39 (SD = 0.28)	25	≤ 2.08	M = 91.41 % (SD = 6.74 %)
	4	M = 1.63 (SD = 0.33)	13		M = 83.44 % (SD = 9.28 %)
Interactive	3	M = 5.16 (SD = 0.57)	19	≥ 4.15	M = 69.80 % (SD = 15.50 %)

	4	M = 4.82 (SD = 0.51)	20		M = 61.62 % (SD = 14.63 %)

b. Dyad selection

For each picture, a percentage of agreement among the participants was calculated for each possible answer to the dyad question (e.g. the percentage of people choosing that answer over the total number of participants that answered that question). Following this, the correct answer to the question was selected by taking the option with the highest agreement among raters. The picture was considered categorized as "interactive" if the highest agreement among participants was on a pair in the scene having a social interaction, as "non interactive" if the highest agreement among participants was on the answer "no one is having an interaction" and as "ambiguous" if the highest agreement was on the answer "everyone is interacting" or "three people are interacting" (for the four-people scenes) (see Table 2 for details on the answer categorization).

Table 2. Agreement categorization for the question regarding the interacting dyad in the scene.

Answer	People	Average agreement	N
Non-Interactive	3	M = 81.25 % (SD = 18.87 %)	31
	4	M = 72.49 % (SD = 21.57 %)	17
Interactive	3	M = 80.76 % (SD = 17.52 %)	25
	4	M = 78.12 % (SD = 17.66 %)	26

c. Final picture selection

Surprisingly, there were pictures with a low interactiveness rating, or a low agreement over whether a picture belonged in the interactive vs. non-interactive category, but with a very high agreement on which dyad was involved in a social interaction, if any. This suggests that the interactiveness question assessed the global interactiveness of the scene, which could lead to ambiguity when more than two people are present in the scene. For the aims of this rating study -

to select pictures with a high consensus about whether individual people in the scene were interacting or not - the criterion of the agreement over the interacting dyad, if any, in the scene, was the most important in providing information about whether a picture was ‘interactive’ or ‘non-interactive’.

Therefore, in choosing the pictures we considered first the pictures with at least 65% agreement on the dyad question and, of these, only the pictures with a clear “interactive” or “non interactive” answer (i.e., no pictures with an ambiguous answer like “all people are interacting” or “three people are interacting”). Among these pictures, at least one of the following criteria were used to select the final pictures:

- Belonging to the interactive/non-interactive range as defined by the rating cut-offs (Table 2);
- Inter-rater agreement above 70%;

Given these criteria, a total of 30 pictures containing 3 people and 22 pictures containing 4 people were selected, half of which were assigned to the “interaction” condition and half to the “non-interaction” condition (Table 3 for details on the agreement and pictures distribution across conditions).

Table 3. Rating statistics of the final set of pictures for each condition.

Condition	People	Agreement on dyad	Interactiveness rating	Agreement on rating	N
Non-Interaction	3	*M* = 94.25% (*SD* = 7.22%)	*M* = 1.27 (*SD* = 0.19)	*M* = 93.41% (*SD* = 6.28%)	15
	4	*M* = 85.30% (*SD* = 13.59%)	*M* = 1.60 (*SD* = 0.38)	*M* = 80.56% (*SD* = 22.16%)	11
Interaction	3	*M* = 87.57% (*SD* = 10.71%)	*M* = 4.63 (*SD* = 0.72)	*M* = 59.83% (*SD* = 16.21%)	15
	4	*M* = 90.62% (*SD* = 9.16%)	*M* = 4.57 (*SD* = 0.77)	*M* = 56.65% (*SD* = 16.35%)	11

The final database of pictures depicted a wide range of settings (e.g., airport, caffes, shops, public transport, playground, school), ages, races and genders across conditions (see Appendix B for pictures).

6. Procedures

In this section I outline details of the experimental procedures not mentioned in the individual empirical chapters, for other information see each empirical chapter individually (chapters 3-5).

The majority of the adult participants received university credits for their participation, but some received also monetary compensation. They performed the experiment in the eye-tracking laboratory at Bangor University. Children recruited through school performed the experiment during class time, in a quiet and isolated space allocated for the purpose. A few children, recruited through word of mouth, were tested in the eye-tracking laboratory at Bangor University. Despite the diversity of environments, each participant wore noise-cancelling headphones to mitigate any potential disruptions (e.g., people passing by an open window). A similar set-up was used in each environment. No other person was present in the room aside from experimenter(s) and the participant. For the developmental work, two experimenters were present, while only one experimenter was present for adult participants. Experimenters were always situated out of the participant's sight (behind them, running the host computer), and testing took place in a dimly lit room to reduce light artifacts. Sometimes, because of individual differences in eye sensitivity to light, the lights in the room were turned off, though this was rare.

Data for all three experiments were collected in the same session, with stimuli across the three experiments presented in a randomized order (see each empirical chapter for details on procedures), followed by the categorization task that was used only for pictures relevant to the third empirical chapter. Collecting the data across experiments, in random order, reduces the probability that participants would intuit the research question of any given experiment.

For each participant the procedures were as it follows:

- The participant sat comfortably on a stable, still chair 80 cm from the screen;

- The table was adjusted for the height of the participant to optimize their ability to see the pictures clear, and to obtain the cleanest eye-tracking data;
- Calibration procedure and task were explained;
- Sticker for head tracking was put on participant's forehead (participants had the option of placing it themselves, with guidance and a mirror);
- The camera was focused in order to increase the precision of image acquisition;
- 13 points calibration was completed: given individual differences in pupil size, glasses wearing, and eye movement characteristics, calibration not only allows accurate calculation of gaze direction and gaze placement on the screen, but helps to identify any need to correct angles, distances, or focus. Additionally, this procedure helps with drift correction (see below). The calibration 'chart' is constituted by a set of stimuli (i.e. dots) presented (Duchowski, 2007, pp. 87);
- Participants were asked to put on the noise-cancelling headphones;
- Free-exploration task, with breaks every 35 trials;
- Ambiguous pictures categorization task;
- Debrief.

During the experiment, drift correction procedures were put in place to allow the software to correct for the fact that the measured samples tend to move away from the true gaze position during the experiment (Holmqvist et al., 2011, pp. 61). For the adults, this procedure was in place before each trial – they had one central calibration point they had to fixate and then press space bar to proceed. For children, this procedure was replaced by a 2.5 seconds long animated gif at the centre of the screen in order to bring the participants' gaze back to the centre of the screen before each stimulus, and minimize drift error. For both groups, a drift correction procedure was in place before the start of each of the four blocks.

7. Data analysis and statistical approach

In this section I will outline eye-tracking data extraction aspects for the three experiments in this work, then proceed with a brief introduction to multilevel modelling and its benefits for the current work, to finish with an outline of common statistical decisions used in this work.

7.1 Data extraction

Data were collected from both eyes but we then chose and used data from the best eye based on manual checking of the data quality (i.e. calibration accuracy) or on notes taken about which eye was likely best during the experimental session. We extracted dwelling time data and time to first fixation (as defined in the "Eye-tracking methodology" section above). In line with best practice (S. Fletcher-Watson et al., 2009), we only considered trials where dwelling time on the picture was more than 33%. Additionally, as a saccade can last anywhere from 30 to 80ms (Holmqvist et al., 2011, pp. 41), we took a conservative approach and considered only time-to-first fixation times longer than 80 ms. When the time to first fixation was smaller than 80ms, we considered the time to second fixation to that area of interest, and then the time to third fixation if the value was still smaller than 80 ms. This was to ensure that time-to-first-fixation data were actually indicative of attentional orienting processes, rather than sometimes reflecting pre-emptive eye-movements.

7.2 Introduction to multilevel modelling

Hierarchical or multilevel modelling is a statistical approach that takes into consideration the hierarchical (nested) structure of the data, and accounts for the lack of independence of errors encountered in repeated measures or within subjects designs. Indeed, data are often hierarchical, meaning that variables are often clustered or nested within each other (A. P. Field & Wright, 2011). Examples of such data come from educational research where students are grouped in classes (taught by different teachers), in different schools (with different educational approaches and policies), in different towns (different social demographics). Therefore, the variables of interest can be influenced not only by the experimental conditions explicitly manipulated by researchers, but by the different "contexts" within which data are gathered, for example different teaching styles or different educational approaches within the school (A. Field et al., 2012; A. P. Field & Wright, 2011). Typical statistical approaches that use measures of means differences, such as ANOVAs, miss this intra-level variability, while hierarchical structures such as multilevel modelling take this into account. These different "contexts" (e.g. teacher, class, school) can be considered as different levels where confounding variance can happen. In the educational example, for instance, the different levels would be the town at the highest level, and within each town nested are the schools, followed by the classrooms within each school and finally the student at the first level (A. Field et al., 2012).

Furthermore, multilevel models are especially suited for repeated measures or within subjects designs (A. Field et al., 2012), where stimuli or trials can be considered as nested within individuals (L. Hoffman & Rovine, 2007). Indeed, in these cases, datapoints within each participant are not independent, therefore residuals (errors) will be correlated, violating typical assumptions of independence for ANOVAs or regressions (A. P. Field & Wright, 2011). The multilevel structure accounts for this lack of independence by calculating the intraclass correlation (ICC), as a measure of the proportion of the total variability in the dependent variable that can be attributed to the "context" (the class, the school, the town in the example above) (A. Field et al., 2012; Twisk, 2006). Therefore, in these models, the participant can be considered at the highest level, then within the participant each experimental condition is nested, at lower levels, until the outcome variable is considered at the lowest level. The hierarchical structure allows us to consider the "context" by modelling this interdependence, through inclusion of random effects in different levels (see below) (A. P. Field & Wright, 2011).

Multilevel modelling also has the benefit of dealing very well with missing data (A. Field et al., 2012) and with the lack of homogeneity of variance, as this variability can be modelled at the level of the regression slopes (see below) (A. Field et al., 2012; A. P. Field & Wright, 2011). Multilevel models can be considered as a more complex version of multiple regression models (Twisk, 2006), with parameters being estimated at separate levels nested within each other. Indeed, in traditional regression it is assumed that the parameters (slope and intercept) are fixed. In multilevel models, these parameters are not fixed, but can instead be allowed to vary, and they can vary at the different levels of the hierarchical structure (A. Field et al., 2012). In particular, in a random intercept model, intercepts vary across the contexts defined by the structure of the model (the classroom, the school or the town in the above example): this means that the relationship between independent variables/predictors and outcome is constant across contexts (has same slope) but it is located in a different position (different intercepts). In a random slope model, slopes vary across levels in the structure, meaning that the relationship between the variables is different at the different levels in the structure, but they are fixed at the same position (same intercept). Finally, a model could allow both intercepts and slopes to vary across levels.

Therefore, in a multilevel model, the fixed coefficients are the coefficients of the variables that are assumed to be constant across the different levels of the model (more simply said, the independent variables or the predictors), while the random coefficients indicate the variables that are allowed to vary across levels, and either their slope or intercept, or both, may vary.

Once a model is built, its fit is assessed through a chi-square likelihood ratio test. The R software used for the analyses in this work produces two adjusted versions of this value, the AIC (Akaike's information criterion) and the BIC (Schwarz's Bayesian criterion). Both these criteria are interpreted in a similar way as the log-likelihood, wherein the smaller its value the better the model fit (A. Field et al., 2012).

Multilevel modelling assumptions

Being a complex extension of multiple regression, multilevel modelling shares similar assumptions. There should not be zero variance in any of the fixed effects and no multicollinearity (in other words, there shouldn't be a perfect linear relationship between predictors). Homoscedasticity implies that for each level of each predictor, the variance of the residuals is constant, although hierarchical models can overcome this through modelling this variability in the slopes (A. Field et al., 2012). Independence of errors is one assumption that is "solved" by using multilevel models, as the hierarchical structure accounts for the lack of independence (Twisk, 2006). Finally random coefficients are assumed to be normally distributed around the overall model, and the errors in the model should be normally distributed (A. Field et al., 2012).

7.3 Statistical approach and decisions

In this work we used multilevel models as an alternative to ANOVA (L. Hoffman & Rovine, 2007) where each participant, picture, condition and AOI act as "contexts" for each datapoint. Indeed, modelling the data as a hierarchical structure allows us to take into account that each participant's gaze behaviour is different from everyone else's, and helps to account for the naturalistic and complex nature of the scenes. Given the diversity of the stimuli we are using, each picture will likely draw different gaze patterns from participants than other pictures. Multi-level modelling allow us to capture and explain much more of this variability in the data compared to traditional ANOVAs, even though this method necessitates the use of more sophisticated techniques and softwares.

Model building procedure

After assessing the need for a hierarchical model evaluating the variance at different levels – i.e. random intercepts only models – we proceeded with a nesting procedure by adding one factor at a time as fixed effects along with their respective interactions. We did not always choose the *best* model as often interaction terms of interest to answer our research questions did not necessarily improve the model fit. We used maximum likelihood estimation (ML) as a method for model estimation and to allow comparison between models (A. Field et al., 2012). Evidence suggests ML is better at estimating the parameters of fixed effects instead of the REML approach (restricted maximum likelihood estimation; Twisk, 2006).
In R software there are various tool-box options, but here we used the nlme package (Pinheiro et al., 2016), with the "lme" function for model building, "anova" for model comparisons or assessment of main effects and interactions within one specific model, and "summary" to assess the degree of variance explained by the model and to visualize multicollinearity. Where necessary, post-hoc comparisons were performed with the emmeans package (Lenth et al., 2018) and we additionally used Tukey HSD correction for pairwise comparisons to avoid type I error.

Common statistical decisions

Multilevel modelling assumes that residuals (error) at each level are normally distributed. In this work the time to first fixation data was always skewed, because of its very nature – i.e in our data screening approach, this measure could never be smaller than 80 ms, no zeros were included, and given the meaning of the measure, there is a natural limit to how late the gaze would enter the AOI. Therefore the data was positively skewed, with a higher frequency of lower values compared to higher values. As a consequence, the multilevel assumptions of normality of residuals were often not met by time to first fixation models. Additionally, with multilevel modelling the analysis considers raw data rather than means, therefore the size of the sample is much bigger than what it would be if we used an analysis of variance approach. Normality tests like Shapiro-Wilk or Anderson-Darling are very sensitive to large samples and, in such large samples, extreme values are often not enough to influence the results. Therefore, whenever the distortion was substantial (graphical normality was clearly skewed) we proceeded with data transformation (Tabachnick & Fidell, 2013). As indicated by Tabachnik et al. (Tabachnick & Fidell, 2013) (although see below for outliers decisions), we proceeded by trying different transformations that could help improve the graphical representation of residuals, but usually were not able to fully 'fix' the skew in the time-to-first-fixation data.

In this work outliers in the data as defined by strong deviations from the group mean were not eliminated. Indeed, while such procedures can be useful for central tendency based approaches like ANOVAs, in the case of hierarchical modelling it is not necessary. For instance, while data were manually checked for quality and outliers were dropped when there was too much missing gaze data, we kept extreme values as we considered such values to be indicative of the specific gaze behaviour of that participant to that picture. Additionally, when outliers could actually skew the distribution, which was rare, the distribution would be transformed rather than removing data (see above).

Another data decision applied across the three experiments was to centre (by group mean) the age in the developmental sample when assessing developmental changes in social attention. This procedure is useful when data with a value of 0 is meaningless (as age is in this case), therefore assigning a meaningful zero helps interpretation of the intercept (A. Field et al., 2012). Finally, this procedure can help with multicollinearity between predictors (A. Field et al., 2012).

Chapter 3. Development of attention to social interactions in naturalistic scenes

Note: Part of the work here presented, namely "Experiment 1" was submitted for publication and is currently under (revision) review:
https://www.biorxiv.org/content/10.1101/2021.02.26.433078v1

Abstract

Human attention is easily captured by social information in naturalistic scenes, a "social bias" present since infancy. Additionally, recent findings suggest people might also preferentially attend to and more quickly detect interacting dyads compared to non-interactive individuals. However, little work has investigated how interactive mechanisms influence attention in naturalistic scenes, nor how these effects may change across development.
Here we recorded the eye-movements of 73 adults and 54 children in a free viewing experiment. Naturalistic scenes contained dyads who were either interacting or not. We explored the influence of the presence (vs. absence) of a social interaction on attentional orienting to social vs. non-social information. Areas of interest (AOIs) were divided between "social" (entire human figures in the scene) and "non-social" (all other elements). Results confirm a "social bias" in both age groups, indicated by increased attentional engagement and faster capture by AOIs than other scene elements. Crucially, this bias is increased by the presence of a social interaction, in both groups in a similar way. Implications for social attention and its development are discussed.

Introduction

Successful navigation in the social world requires the ability to attend to and understand a wide range of social cues in a cluttered and complex environment, skills that develop starting very early in life. Human adults are indeed experts at extracting social information from the surrounding world, and under most circumstances their attention is captured by and preferentially held by human information, including both faces and bodies (e.g. Doherty et al., 2017; Fletcher-Watson et al., 2008). This "social attentional bias" seems to be automatic (Rösler et al., 2017), relatively independent of low-level features of the scene, including saliency, and resistant to top-down task demands (Flechsenhar & Gamer, 2017). Crucially, while this bias is already present in infancy, it is considerably stronger in adulthood (Frank et al., 2012; Soto-Icaza et al., 2015). Research investigating developmental changes in this social preference during pre-adolescent childhood is, however, not conclusive (e.g., Doherty et al., 2019; Van Der Geest et al., 2002).

Most prior research into this "social attentional bias" has focused on attention to isolated individuals. In the social world, however, we often observe multiple people in a scene, and have the opportunity to glean more complex and richer social information from observed social interactions than we can from isolated individuals. Indeed, observed social interactions represent a unique source of social information (e.g., about *relationships* between people) and social learning (Papeo, 2020; Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017). Recent research suggests social interactions may be attended to and processed differently than are two non-interactive individuals. For example, several studies suggest interacting dyads are processed as a meaningful gestalt, similarly to a single individual rather than as two individual actors (Papeo et al., 2019; Papeo & Abassi, 2019; Vestner et al., 2019; Walbrin & Koldewyn, 2019). Similarly, in visual search tasks, two agents facing each other show a search advantage compared to non-facing pairs (Papeo et al., 2017) or pairs facing the same direction (Vestner et al., 2020), although it isn't yet clear if, or under what circumstances, this effect is specific to human information (e.g., Vestner, Over, Gray, & Cook, 2021; Vestner, Over, Gray, Tipper, et al., 2021). However, very few studies have investigated how attention to social information is influenced by the presence of social interactions in naturalistic scenes, and conclusions from those that have are contrasting. Birmingham et al. (2008) used scenes depicting one or 3 people either being active (doing something independently), inactive (independently doing nothing) or interacting (3 people doing something together), to investigate attention to the eyes of the agents in the scene. The level of activity and the number of people in the scene increased the attention to the eyes of the agents, but they found no difference between interactive and independent active scenes in fixation proportions towards human information. In contrast, Kujala and colleagues (2012) demonstrated that adults attend more to heads when freely viewing pairs of people facing each other engaged in friendly interactions than when the same figures faced away from each other (Kujala et al., 2012), and Villani et al. (2015) show that attention to faces and arms is increased in interactive paintings compared to non-interactive ones. Importantly all this research has focused on *social* attention but has not directly investigated how both social and *non-social* attention may change in the presence of social interactions.

Like the adult literature, work investigating the development of social attention has focused primarily on attention to isolated individuals. Work looking at the development of social attention in mid-to-late childhood is also surprisingly sparse. Indeed, while infant research suggests that the attentional bias towards human information and sensitivity to interactive social information are both present in the first year of life, research about how social attention develops across childhood is limited and shows contrasting results. Quite a number of studies have investigated the attentional bias to human information in the light of comparisons between typically developing and autistic individuals, showing that there is a preference towards social information in typical mid-childhood that is reduced in Autism Spectrum Disorder (e.g., Riby & Hancock, 2008; Sasson & Touchstone, 2014). The focus of this work has been on between-group differences, however, so developmental change within this age group, or between age-groups, is not discussed. Work that has looked at developmental change suggests that attentional capture by social information is actually higher in 6-10 year-old children compared to adults (Doherty et al., 2019) and that there may be a steep increase in social attention in early childhood followed by a milder increase from age 5 into adulthood, particularly in the proportion of time participants spend attending to social information in scenes (Amso et al., 2014).

Similarly, there is very little developmental research on how attention is allocated in complex social scenes that contain social interactions, especially during childhood. We do know that infants are sensitive to the presence of interactions. Indeed, the ability to distinguish interacting from non-interacting dyads and shift attention appropriately between two conversing adults compared to two non-facing adults is present already by 4 months of age (Augusti et al., 2010; Handl et al., 2013). Additionally, at 6 months, infants also show a preference for a pair of agents who help each other compared to those that compete with each other, suggesting an early sensitivity to higher-order social interaction processing (Hamlin et al., 2007). Infants also appear to process a facing dyad as a unique chunk, similar to adults, while this is not true for non-facing agents (Papeo et al., 2020). Research investigating attention specifically to social interactions has rarely been carried out in later childhood, however. School-aged children show a greater attention shift from faces to bodies when multiple people are in a scene compared to adults (Stoesz & Jakobson, 2014), suggesting that they struggle a bit more than adults in processing fine-grained social signals when scenes contain multiple social targets. To our knowledge, however, there is only one study that has specifically investigated attention to social interactions during childhood, showing that 9 year-old children looked longer at agents who were facing each other when they were presented on screen at the same time as another pair of agents who were not facing each other (Stagg et al., 2014). Although this research suggests that pre-adolescent children show a bias to attend to interactions over other human targets, the focus of the research was on autism rather than development, and children's performance was not compared with that of either younger or older age groups. In addition, similarly to the adult visual search studies discussed above, the only interactive cue was the facing direction of the agents and stimuli were presented isolated from a noisy real-life background. In real life, scenarios are much richer in both the diversity of social cues and the amount of distracting clutter. Importantly, there is also good reason to think that there may be developmental change in how attention is allocated to social interactions in complex naturalistic scenes, as the brain structures supporting perception and processing of social interactions are not yet fully developed in children 6-11 years old, both structurally (Mills et al., 2014) and functionally (Sapey-Triomphe et al., 2017; Walbrin et al., 2020).

In the current study we investigate the influence of social interactions on social attention in cluttered naturalistic scenes in adults (exp. 1), and assess whether the influence of social interactions on social attention changes during pre-adolescent childhood (exp. 2). To do so, we evaluated looking behaviour during a free viewing paradigm where participants viewed naturalistic scenes depicting either two people interacting or not. We expect to re-confirm the social attentional bias to human information in both age groups, including both faster capture by and more attentional engagement to humans compared to non-social scene elements. We also expect this social bias to be increased in pictures containing social interactions. Additionally, in line with the developmental imaging data showing an increase with age in neural sensitivity to social interactions, while we expect that children will show a social bias to social interactions, we expect the difference between interactive and non-interactive scenes to be smaller than in adults.

General methods

Both studies were pre-registered with an a priori sample size calculation in G Power (Erdfelder et al., 2009; Faul et al., 2007) to obtain a large effect size (Cohen's $f = .40$), and reach 80% power ($\alpha \leq .05$).

All participants had normal or corrected-to-normal vision.

Stimuli and apparatus

The stimuli were selected from an online database (SUN, Xiao et al., 2010) to be emotionally neutral, to depict a variety of ordinary life contexts (e.g. schools, shops, markets), and to not contain any agent looking directly at the camera or at a (presumed) off-screen individual.

The initial stimulus set contained 127 pictures, depicting either 2 people having a social interaction or two people acting independently (see Appendix B1 for the full set of pictures). Twenty-six independent judges rated the pictures for 'level of interactiveness' on a Likert scale from 1 ("Not interactive at all") to 7 ("Very interactive). The 30 pictures receiving the lowest and the 30 receiving the highest average score were selected to be part of the "non-interactive" and "interactive" conditions, respectively (see "Materials" section in the General Methods chapter).

Each picture in the final set was pre-processed in Photoshop (version CC 2019) by neutralizing the colour cast ("match color – neutralize color" option), matching it for colour scheme with one sample picture ("match color" option) and sharpening ("sharpen" option) (see Chapter 2: "General methods - Materials" section for details on the picture selection process). The pictures were presented in PsychoPy 2 (Peirce et al., 2019) on a 380 x 215 mm (1920 x 1080 px) monitor on a grey background and each picture had a size of 860 x 860 pixels (13.6° x 13.6° visual angle). Each stimulus was presented with the screen centre-most margin 60 pixels either to the left or right from the fixation cross, (0.85° visual angle). The data was collected with an EyeLink Portable Duo Tracker with remote binocular system at a 1000hz sampling rate. Data from both eyes were collected, but we used monocular data for the analysis, choosing which eye's data to use individually for each participant, based on calibration accuracy.

Procedure

Participants sat comfortably on a stable, non-swivelling chair approximately 80 cm from the screen and freely viewed a total of 142 pictures: 60 pictures belonged to this experiment, while the rest belonged to two other experiments that will not be discussed here. All stimuli were fully randomized so that each participant saw the pictures in a different order.

Before the task began, a 13-point calibration procedure was carried out for each participant. Every stimulus was presented for 5s and presentation side for each picture was counterbalanced across participants.

Data preparation

For each picture two areas of interest (AOIs) were defined with the "freehand" option in Eyelink Data Viewer (SR Research, 2013). The social AOIs contained the whole of both human figures in the scene, while the non-social AOIs contained everything else in the scene. We extracted dwelling time –including fixations and saccades, spent inside each AOI – as a measure of attentional engagement with social and non-social information in the scene. Time to first fixation – time the eye took to enter a specific AOI for the first time for each picture – was used as a measure of attentional capture by social and non-social information.

Experiment 1

Aim

In the first part of this work the goal was to re-confirm the presence of a social attentional bias, i.e., more attention given to social information compared to non-social information, in adult participants in cluttered naturalistic scenes, and investigate the influence of social interactions on this bias.

Participants

The sample size analysis established a sample of 70 adults (pre-registered on AsPredicted; https://aspredicted.org/blind.php?x=se7k7b) (see Appendix A1 for pre-registration). Seventy-three adult participants were recruited, but data from two participants outside our age range and one who was inattentive due to sleepiness during the experiment were removed. The final adult sample was composed of 70 participants (*M* = 21.07, *SD* = 2.63, range = 18-35; 47 female and 1 other). Participants provided informed consent and received either money or university credits as compensation. All procedures were approved by the ethical committee at Bangor University (ethics protocol number: 2018-16360).

Procedure

Participants were asked to freely observe the pictures through both an oral explanation of the experiment and an on-screen visual prompt at the start of the session. A drift correction procedure was carried out before each picture was presented, where participants had to fixate a calibration point at the centre of the screen and then press the space bar to proceed to the next trial. These between-trial procedures served to draw participants' gaze back to the centre of the screen before the beginning of the next trial.

The procedure consisted of 4 blocks of 35 trials and lasted around 20 minutes. Participants could take short breaks to rest their eyes between blocks if needed.

Data analysis

Trials that had less than 33% of total engagement time with the target picture were treated as missing (e.g., S. Fletcher-Watson et al., 2009), and this included both off-screen looking time and missing data due to poor signal or blinks. This procedure led to the loss of 0.24 % of trials, with a range of 0-2 trials per participant.

For each of the two measures – dwell time and time to first fixation – we used a separate model. After assessing the variance in the dataset, for each of the two measures, we analysed the data using multilevel modelling with a 2x2 design (nlme package (Pinheiro et al., 2016)) using a four-level hierarchical model. At the highest level we modelled participant information, and, nested within each participant, the social content of the scene (i.e., interacting or non-interacting) was modelled as a third level predictor, while AOI type (i.e., human or background) was modelled at the second level. The measure – time to first fixation or dwell time – for each AOI was modelled at the first level, nested within trial and participant. Finally, pairwise comparisons were performed using post-hoc Tukey's HSD using emmeans package in R (Lenth et al., 2018).

All supplementary materials for this chapter are in Appendix C.

Results

Attentional Engagement

We analysed dwelling time to each area of interest (AOI) in each picture to investigate overall attention to human information compared to the background – our measure of the social attentional bias – and whether the presence of a social interaction in the scene influenced the amount of attention given to social or non-social regions in a cluttered scene. The relationship between the different conditions and dwell time showed significant variance in intercepts across participants, conditions, and area of interest (SD = 261.97, $\chi^2(3) = 342.26$, $p < .001$). While social and non-social AOIs were not significantly different in size from each other, the area of AOIs can influence both capture and engagement. Thus, we added area of the AOIs in pixels as a random effect to the model. This procedure helps to mitigate any effect of difference in the size of the humans across pictures, as well as the naturally larger size of the background compared to the humans (see Supplementary materials S1 for details on AOI sizes). The addition of the AOI size to the model did not create a significantly different model from the above mentioned model (SD = 886.18, $\chi^2(4) = 0.00$, $p = .99$). Therefore, in the final model, we set the fixed effects in the model as the type of scene and the type of AOI, while our random effects were at the participant, condition, AOI, and AOI size levels.

The model showed a non-significant main effect of type of scene (interactive or not), $F(1,69) = 0.42$, $p = .52$, $\eta^2_p = 0.01$ with attention to interactive scenes (M = 1804.73, SD = 920.09) and non-interactive scenes (M = 1782.92, SD = 928.70) being similar. There was a significant main effect of AOI type (social vs non-social), $F(1,138) = 18.84$, $p < .001$, $\eta^2_p = 0.12$, with overall more attention to the social AOIs (M = 1866.17, SD = 932.92) compared to the non-social regions (M = 1721.49, SD = 910.20). Additionally, the analysis revealed a significant interaction between type of scene and type of AOI, $F(1,138) = 43.35$, $p < .001$, $\eta^2_p = 0.24$ (Figure 1). In particular, dwell time was greater for social information (M = 1986.53, SD =

910.75) compared to the background (*M* = 1622.93, *SD* = 893.27) in the interactive scenes , $t(138) = 7.73$, $p < .001$, $d = 0.66$, while within the non-interactive pictures, visit time was similar for social information (*M* = 1745.58*, SD* = 939.49) and non-social AOIs (*M* = 1820.25*, SD* = 916.49), $t(138) = -1.59$, $p = .31$, $d = -0.14$. This suggests that the presence of a social interaction increases the amount of attention given to social information. Indeed, significantly more attention was given to the human AOIs in the interactive scenes compared to the corresponding AOIs in the non-interactive scenes, $t(69) = 5.12$, $p < .001$, $d = 0.62$.

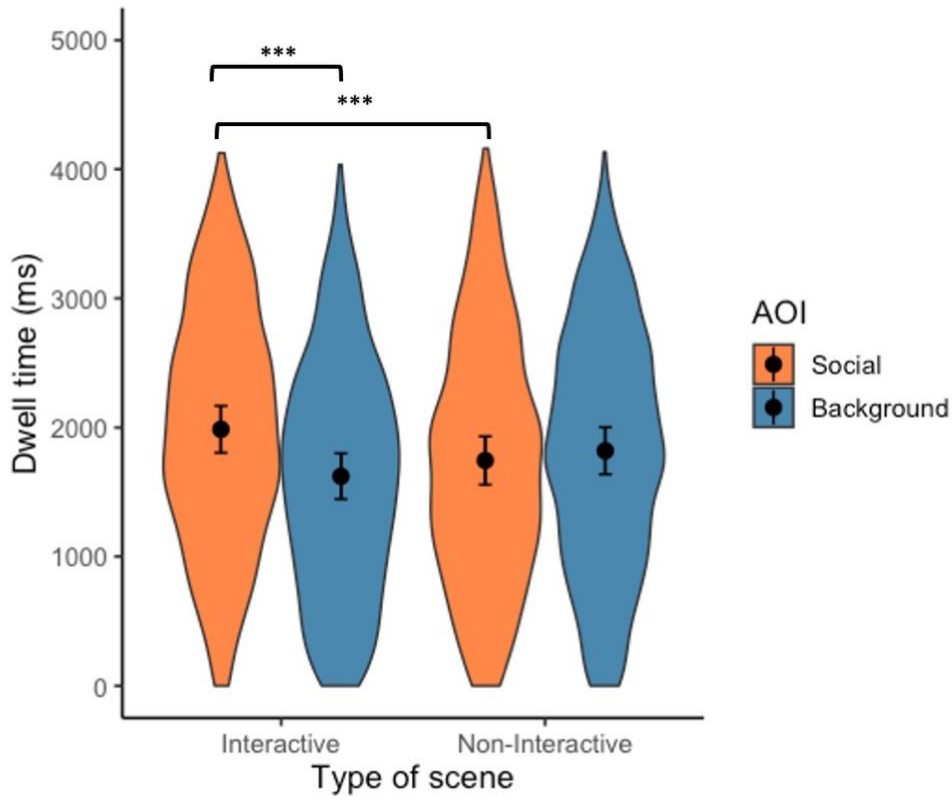

Figure 1. Violin plot for mean dwell time for AOI and scene type. Error bars represent 95% confidence intervals.

Attentional Capture

We analysed time to first fixation to each AOI in each picture to explore capture by social information in cluttered scenes and determine whether the presence of a social interaction altered this time. Missing data on capture measurement due to never gazing in the AOI were 2.14 % of data for all areas of interest and participants.

Time to first fixation was transformed (logarithm in base 10) to meet multilevel modelling assumptions, However, after transforming the data, the structure of the model and the results did not change, therefore we present the untransformed data here to facilitate understanding and interpretation of the results. See Supplementary materials n. S2a for details on the transformation and the full analysis and results using transformed data.

The model assessment showed significant variance in intercepts between participants, condition and area of interest (*SD* = 217.84, $\chi^2(3) = 395.81$, $p < .001$). We added size as a further random effect, and this model did not differ significantly from the originally planned one (*SD* = 685.55, $\chi^2(4) = 0.00$, $p = .99$). Therefore, similar to the prior analysis, we set the fixed effects in the model as the type of scene and the type of AOI, while our random effects were at the participant, condition, AOI and size levels.

The model showed a non-significant main effect of type of scene (interactive or not), $F(1,69) = 0.19$, $p = .89$, $\eta^2{}_p < 0.001$) with capture to interactive scenes (*M* = 673.33, *SD* = 713.73) and non-interactive scenes (*M* = 676.40, *SD* = 724.94) being similar. The main effect of AOI type – i.e. social and non-social – did reach significance, $F(1,138) = 423.33$, $p < .001$, $\eta^2{}_p = 0.75$, with overall faster orienting to social AOIs (*M* = 483.98, *SD* = 535.80) compared to non-social regions (*M* = 867.52, *SD* = 822.22). Additionally, the analysis revealed an interaction between type of scene and type of AOI, $F(1,138) = 21.56$, $p < .001$, $\eta^2{}_p = 0.14$ (Figure 2). Indeed, while participants were faster to orient to social information compared to non-social in both interactive [social (*M* = 440.54, *SD* = 462.51); non-social (*M* = 910.34, *SD* = 836.14), $t(138) = 17.85$, $p < .001$, $d = 1.52$] and non-interactive scenes [social (*M* = 528.11*, SD* = 598.08); non-social (*M* = 824.77*, SD* = 806.03), $t(138) = 11.24$, $p < .001$, $d = 0.96$], participants looked at social information earlier in interactive than non-interactive scenes, $t(69) = 3.33$, $p = .01$, $d = 0.40$.

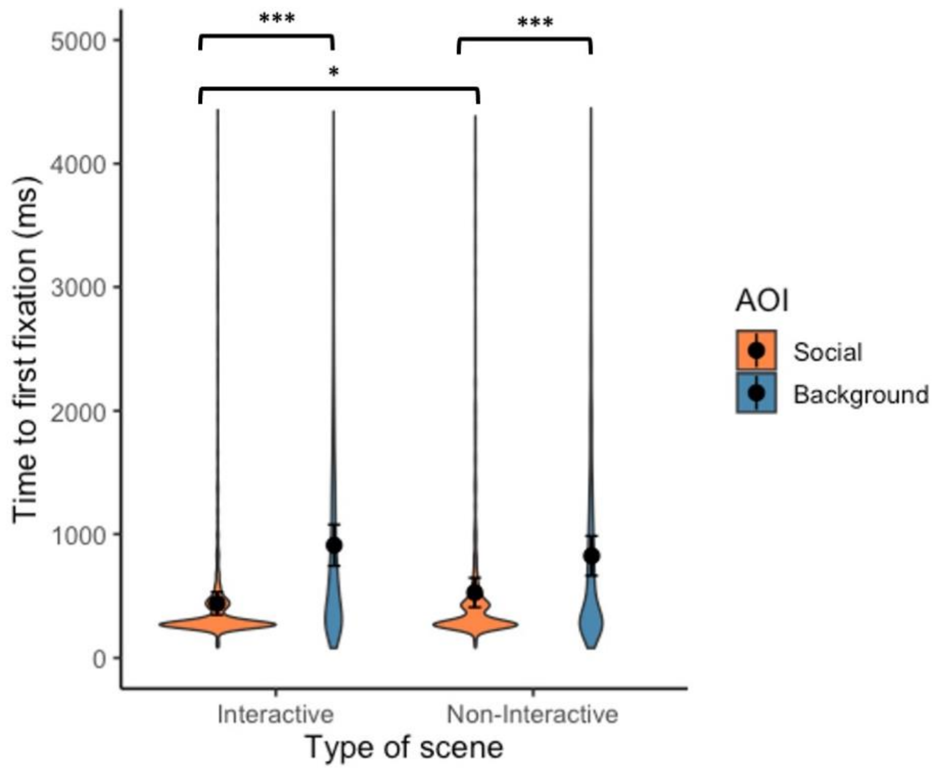

Figure 2. Violin plot for mean time to first fixation for AOI and scene type. Error bars represent 95% confidence intervals.

Discussion

Here, using complex naturalistic scenes depicting pairs of people either interacting or not, we re-confirmed a previously demonstrated (Doherty et al., 2017; Flechsenhar & Gamer, 2017; Sue Fletcher-Watson et al., 2008; Rösler et al., 2017) overall social attentional bias to human information both in engagement and capture, irrespective of scene type. Additionally, when scenes contained a social interaction, participants were faster to look at and spent more time attending human information in the scene. These results are in line with previous research that suggest that interactive dyads are attended to and processed differently than non-facing agents (Papeo et al., 2017; Papeo & Abassi, 2019; Stagg et al., 2014; Vestner et al., 2019, 2020), at least in adulthood.

Interestingly, our results also suggest that the presence of a social interaction moderates attention more weakly in attentional capture than in attentional engagement. If true, this could suggest that interactive information primarily influences the later stages of scene exploration, after the basic social relevance of the information has already been processed. Indeed, it seems likely that social information first captures attention in the early stages of scene exploration regardless of scene type, and then holds it for longer when the agents are engaged in social interaction. This would suggest a hierarchy of social information in scenes, where all human information automatically orients attention and is then further prioritised when an interaction is present.

After investigating the basic mechanisms of social attention to naturalistic scenes and showing that in adults the presence of a social interaction will increase the attention to human information in the scene, in the next section we investigate these processes in pre-adolescent childhood and then explore eventual developmental differences between children and adults in these processes. Indeed, while brain imaging research has shown that brain structures supporting social interaction perception are not yet tuned in this period of development (Sapey-Triomphe et al., 2017; Walbrin et al., 2020), behavioural research has shown that the ability to parse social interactions and learning observationally form them is already in place by school years (e.g., Skinner et al., 2017). Therefore experiment 2 aims to clarify how the social attentional bias changes across childhood, and whether it is influenced by the presence of a social interaction.

Experiment 2

Participants

The power analysis (pre-registered on AsPredicted; https://aspredicted.org/blind.php?x=sk8st6) (Appendix A2) for the developmental sample established a sample of 90 youths between 6 and 18 years old for adequate power to detect a possible three-way interaction. Unfortunately, we were forced to end data collection in March 2020 due to the COVID-19 pandemic, having collected data only from younger participants aged 6 - 12. Thus, the final developmental sample was composed of 54 children (*M* = 8.76, *SD* = 1.72; range = 6-12; 28 female). Children gave assent and each child's guardian(s) gave consent for them to participate. Each child received their choice of small toys as compensation. All procedures were approved by the School of Psychology's Ethics committee at Bangor University (ethics protocol number: 2019-16586)

Procedure

The procedure was broadly similar to the one followed with adult participants. Children were verbally instructed to freely observe the pictures but did not receive text instructions. We also removed the drift correction procedure before every trial. Instead, as the task was split in 4 blocks of 35 trials, a drift correction procedure was carried before each block. Between trials, children were presented with an animated gif at the centre of the screen for 2.5 seconds as a fixation point to draw their gaze back to the centre of the screen before the start of the next trials. Unlike adults, however, they did not need to hit a key to proceed. The experiment lasted around 20 minutes and participants could take short breaks to rest their eyes between blocks if needed. Children filled in a "sticker chart" as they completed different steps in the task to further encourage task engagement and motivation.

Data analysis

As we did for the adult group, we treated trials with less than 33% of total engagement time with the picture as missing (e.g., S. Fletcher-Watson et al., 2009). For the developmental group this led to the loss of 3.06% of trials, with a range of 0 – 17 trials per participant. Considering the range of data loss in this group (0% - 28% per participant), during the model building procedure we explored whether the number of missing trials influenced experimental effects. We found that although the missing trials unsurprisingly produced a significant effect on the total dwelling time $F(1,51) = 18.28$, $p < .001$, ${\eta^2}_p = 0.26$), the presence of the missing trials as a predictor in the model did not eliminate any of the effects of interest (see Supplementary material n. S3a for details). We thus feel confident that missing trials did not drive any effects of our experimental variables (i.e., condition, AOI) in the data. Our final model does not, therefore, include the number of missing trials as a fixed effect.

As in the adult analysis, for each measure (i.e., dwell time, time to first fixation) we used a separate multilevel model with a 2x2 design (nlme package; (Pinheiro et al., 2016)) using a four-level hierarchical model. At the highest level we modelled participant information, and, nested within each participant, the social content of the scene (i.e., interacting or non-interacting) was modelled as a third level predictor, while AOI type (i.e., human or background) was modelled at the second level. The measure – time to first fixation or dwell time – for each AOI was modelled at the first level, nested within trial and participant. To look at developmental change across our age-range, the age of each participant was modelled as a continuous predictor.

Results

Attentional Engagement

We analysed dwelling time to each AOI in each picture to investigate the presence of a social attentional bias in cluttered scenes in pre-adolescent children, and to explore whether the presence of a social interaction in the scene had any effect on this bias. The relationship between type of information in the scene and dwell time showed significant variance in intercepts across participants, condition, and area of interest (*SD* = 297.38, $\chi^2(3) = 220.21$, $p < .001$). As in the adults group model, we added the area of the AOIs in pixels as a random effect. This model did not differ significantly from the originally planned one – *SD* = 1060.08, $\chi^2(4) = 0.00$, $p = .99$. Therefore, the model had age, type of scene and AOI as fixed effects, and participant, type of scene, AOI and AOI size as random effects.

The model showed a non-significant main effect of age ($F(1,52) = 1.86$, $p = .18$, $\eta^2{}_p = 0.03$), suggesting that there was very little change in *overall* attentional engagement across our age range. Similar to the adult analysis, there was also a non-significant main effect of type of scene, $F(1,52) = 0.04$, $p = 0.84$, $\eta^2{}_p < 0.001$) with similar attention given to interactive scenes (*M* = 1830.14, *SD* = 1092.89) and non-interactive scenes (*M* = 1823.34, *SD* = 1110.45). Unsurprisingly, the main effect of information – i.e. social and non-social – did reach significance, $F(1,104) = 68.34$, $p < .001$, $\eta^2{}_p = 0.40$, with overall more attention given to social AOIs (*M* = 1989.44, *SD* = 1098.06) compared to non-social regions (*M* = 1664.09, *SD* = 1081.01). Additionally, as in the adult group, the analysis revealed a significant interaction between type of scene and type of AOI, $F(1,104) = 30.93$, $p < .001$, $\eta^2{}_p = 0.23$ where children spent more time gazing within social AOIs (*M* = 2101.30, *SD* = 1058.53) than at the background (*M* = 1558.98, *SD* = 1059.22), $t(104) = 9.77$, $p < .001$, $d = 0.93$ in the interactive scenes. In non-interactive scenes, however, dwell time was similar for social (*M* = 1875.93*, SD* = 1125.79) and non-social AOIs (*M* = 1770.75*, SD* = 1092.72), $t(104) = 1.87$, $p = 0.19$, $d = 0.18$. Additionally, and again similar to the adult findings, children gave significantly more attention to the human

AOIs in the interactive scenes compared to the corresponding AOIs in the non-interactive scenes, $t(52) = 4.08$, $p < .001$, $d = 0.52$ (Figure 3).

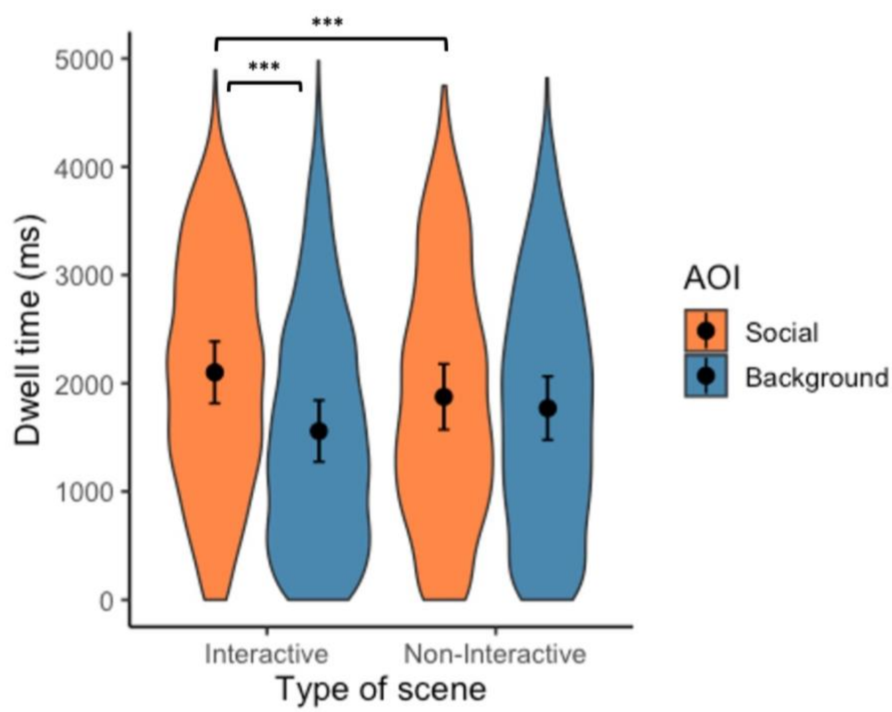

Figure 3. Violin plot for mean dwell time for AOI and scene type. Error bars represent 95% confidence intervals.

There was no significant interaction between age and scene type ($F(1,52) = 0.02$, $p = .89$, $\eta^2_p < 0.001$), but there was a significant interaction between age and type of information in the scene, collapsing across scene-type ($F(1,104) = 12.80$, $p < .001$, $\eta^2_p = 0.10$), where the slope of the linear relationship between age and dwell time differed between AOI type ($t(104) = -3.57$, $p < .001$, $d = -0.34$). Interestingly, this difference is maximal at the youngest ages, and we see a mild decrease in attention to social AOIs and a steeper increase in attention to non-social AOIs across the age-range (Figure 4). The three-way interaction between age, scene and AOI did not improve the model fit significantly, and when tested, was not significant ($F(1,104) = 0.15$, $p = .70$, $\eta^2_p < 0.001$ (Figure 5).

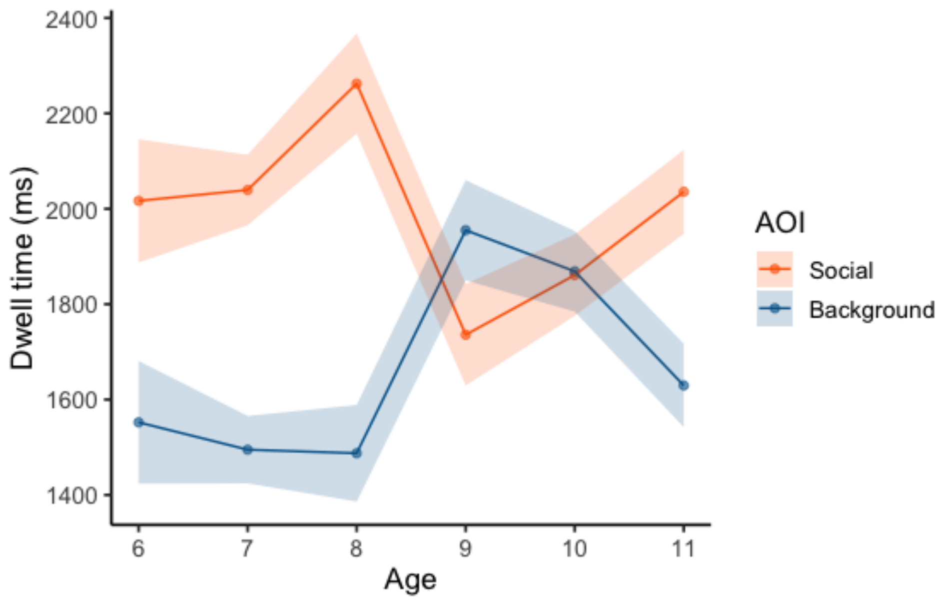

Figure 4. Average dwell time to social and background AOIs across scenes in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

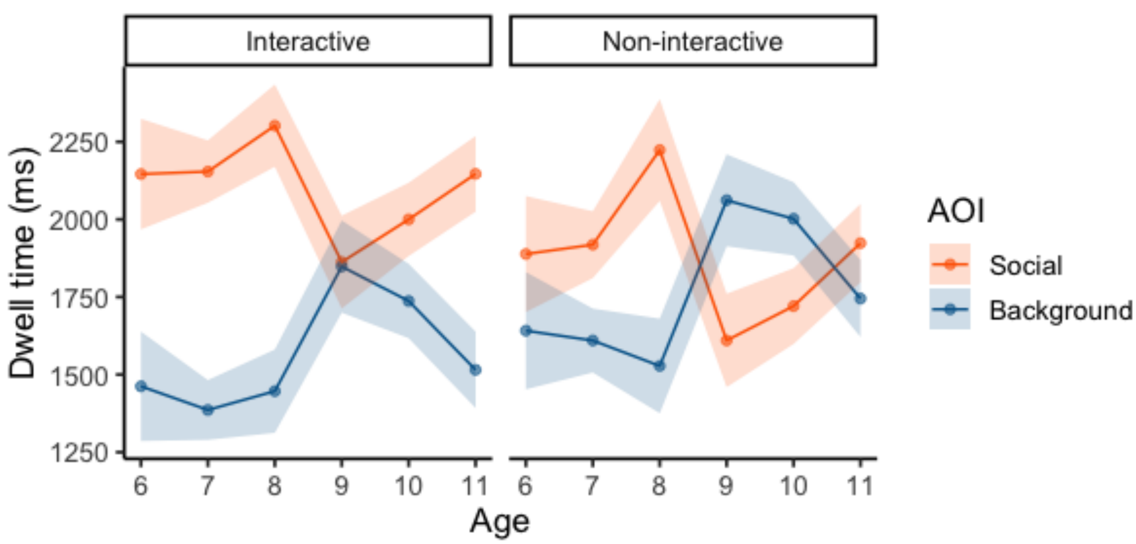

Figure 5. Average dwell time to social and background AOIs in interactive and non-interactive scenes in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

Attentional Capture

We analysed time to first fixation to each AOI in each stimulus to investigate the presence of a social bias in attentional capture in pre-adolescent children, and to explore whether a social interaction in the scene had any effect on how quickly either social or non-social information captured attention. Missing data for our capture measure due to never gazing in the AOI were 3.82 % of data for all areas of interest and participants.

Time to first fixation was transformed (logarithm in base 10) to meet multilevel modelling assumptions. However, after transforming the data, the structure of the model and the results did not change, therefore we present here the untransformed data to facilitate understanding and interpretation of the results. See Supplementary materials n. S2b for details on the transformation and the full analysis and results from transformed data.

The model assessment showed significant variance in intercepts between participants, condition and area of interest (SD = 217.13, $\chi^2(3)$ = 253.92, p<.001). After controlling for the correlated error, participants' centred age within this group was modelled as a continuous predictor. When we added area of the AOIs in pixels as a random effect, the model did not change significantly from the originally planned one – SD = 713.53, $\chi^2(4)$ = 0.00, p = .99. Therefore, the model here shown had age, type of scene and AOI as fixed effects, and participant, type of scene, AOI and AOI size as random effects.

There was a main effect of age, $F(1,52)$ = 3.99, p = 0.05, $\eta^2{}_p$= 0.07, with younger children tending to be slightly slower to look at AOIs, collapsed across AOI and scene type. This trend, however, seems to be primarily driven by an interaction between age and AOI type ($F(1,104)$ = 7.29, p = .01, $\eta^2{}_p$ = 0.07). The slopes of the linear relationship between time to first fixation and age in the two different AOIs were significantly different ($t(104)$ = 2.69, p = .01, d = 0.26), although this difference is best characterised by a decrease in first-fixation time to non-social information across age, while attentional capture time for social AOIs remains relatively steady (Figure 6). Despite this, the social bias in attentional capture is maintained across development and the interaction between age and type of scene is not significant ($F(1,52)$ = 0.04, p = .84, $\eta^2{}_p$< .001. As in adults, the three-way interaction between age, type of scene, and type of AOI neither improved the model fit nor was significant ($F(1,104)$ = 0.02, p = .90, $\eta^2{}_p$< .001) on the attentional capture, meaning that there was no developmental change in the way attention was captured by social and non-social information in interactive and non-interactive scenes (Figure 7).

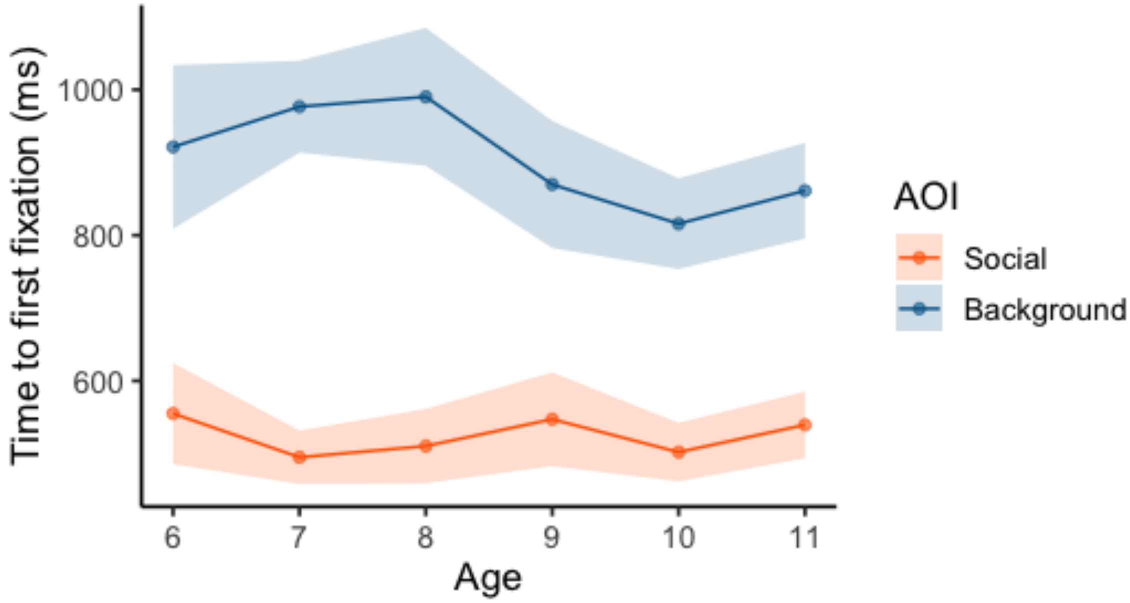

Figure 6. Average time to first fixation to social and background AOIs across scenes in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

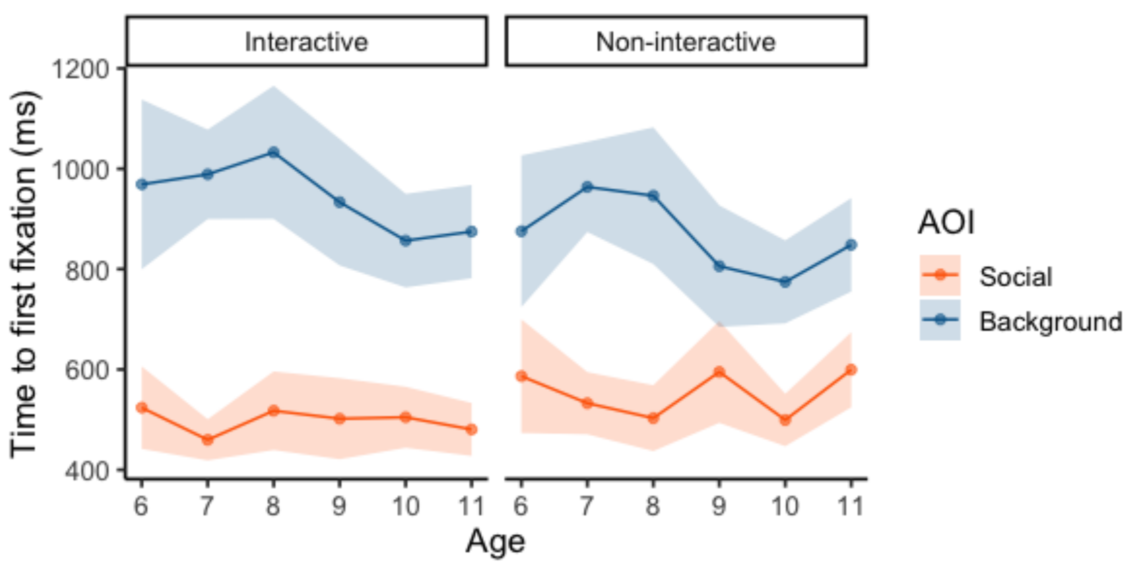

Figure 7. Average time to first fixation to social and background AOIs in interactive and non-interactive scenes in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

Attentional capture did not differ by type of scene (F(1,52) = 0.001, p = 0.97, $\eta^2_p < .001$), with non-interactive scenes (M = 708.97, SD = 748.58) capturing the first fixation as quickly as did interactive scenes (M = 709.47, SD = 743.65). Instead, as predicted, there was a significant main effect of AOI type, F(1,104) = 309.83, p < .001, $\eta^2_p = 0.75$, with faster orienting of attention to the social AOIs (M = 518.71, SD = 549.03) compared to the non-social regions (M = 903.17, SD = 861.33). Additionally, as in the adult data, there was a significant interaction between type of scene and type of AOI, F(1,104) = 7.61, p = .01, $\eta^2_p = 0.07$ (Figure 8). Indeed, like adults, children were faster to orient to social information (M = 490.84, SD = 502.07) compared to the background (M = 935.83, SD = 874.19), t(104) = 14.43, p < .001, d = 1.42 in both interactive and, to a lesser extent, non-interactive scenes (social: M = 547.66, SD = 592.68; non-social: M = 870.38, SD = 847.26; t(104) = 10.43, p < 0.001, d = 1.02). Unlike in adults, however, the difference between capture by interactive humans and non-interactive humans did not reach significance in post-hoc analyses, t(52) = -1.85, p = 0.20, d = -0.26.

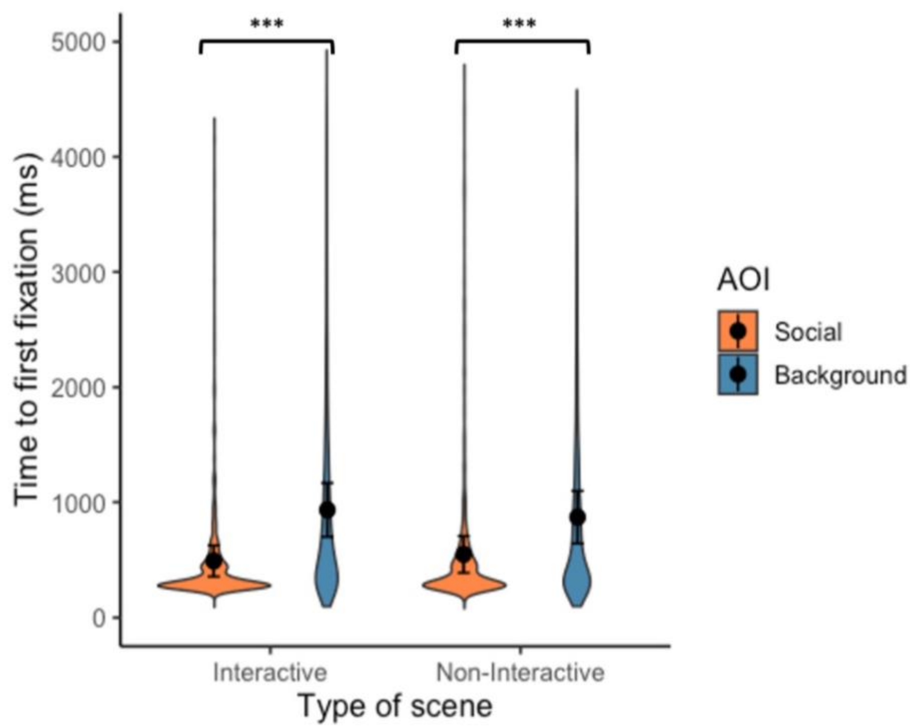

Figure 8. Violin plot for mean time to first fixation for AOI and scene type. Error bars represent 95% confidence intervals.

Developmental changes in social attention between children and adults

Attentional engagement

So far, we have assessed children and adults separately and looked at developmental change only *within* our child group. To assess developmental change *between* children and adults, age was modelled as a categorical predictor (children vs. adults), and we followed the same model building procedures and logic that we have used previously. Additionally, since children and adults differed in the amount of missing data, we added this as a fixed effect in the model to explore the possibility that the amount of available data explain some of the variance, and therefore have an influence on our effects of interest. Missing trials produced a significant effect on the total dwelling time $F(1,121) = 19.49$, $p < .001$, $\eta^2_p = 0.14$), but this factor did not change any of the effects of interest (see Supplementary material n. S3b for details). We therefore continued the model building without including number of missing trials as a fixed effect. The final model included age-group, type of scene, and AOI as fixed effects, allowing intercepts to vary at participant, condition, AOI and AOI size level.

Reassuringly, there was no main effect of age-group on overall attention to the scene, $F(1,122) = 1.37$, $p = .25$, $\eta^2_p = 0.01$. Unsurprisingly, given the results of both separate adult and child group analyses, there was no main effect of scene-type ($F(1,122) = 0.37$, $p = .54$, $\eta^2_p <$ 0.01) and the main effect of area of interest remained highly significant ($F(1,244) = 74.47$, $p <$ .001, $\eta^2_p = 0.23$) with overall more attention given to social information ($M = 1918.99$, $SD =$ 1008.77) compared to non-social information ($M = 1696.90$, $SD = 987.35$), irrespective of age group and scene type – i.e. interactive or not. Likewise, the interaction between condition and area of interest was maintained, $F(1, 244) = 72.21$, $p < .001$, $\eta^2_p = 0.23$, with more attention given, across groups, to the social information ($M = 2035.88$, $SD = 978.55$) than the background ($M = 1595.43$, $SD = 968.50$, $t(2445) = 12.33$, $p < 0.001$, $d = 0.79$) in interactive scenes, while there was essentially no difference in attention to the two types of AOIs (social: $M = 1801.23$, $SD = 1025.06$; non-social: $M = 1799.12$, $SD = 995.70$) in the non-interactive scenes ($t(244) =$ 5.50.41, $p = .99$, $d = 0.03$) (Figure 10).

The interaction between age group and type of scene was not significant ($F(1,122) = 0.07$, $p = .79$, $\eta^2_p < 0.001$ but there was a significant interaction between age group and type of AOI ($F(1, 244) = 12.01$, $p < .001$, $\eta^2_p = 0.05$ (Figure 9). In particular, while both children ($t(244) = 8.27$, $p < 0.001$, $d = 0.53$; social: $M = 1989.44$, $SD = 1098.06$; non-social: $M = 1664.09$, $SD = 1081.01$) and adults ($t(244) = 4.22$, $p < 0.001$, $d = 0.27$; social: $M = 1866.17$, $SD = 932.92$; non-social: $M = 1721.49$, $SD = 910.20$) looked at social information for longer compared to the non-social information, this difference was larger for children. Indeed, children looked for longer at the social information than did adults, $t(122) = 3.27$, $p = .01$, $d = 0.30$, while there was no between-group difference in looking time to non-social information, $t(122) = -1.62$, $p = .29$, $d = 0.15$.

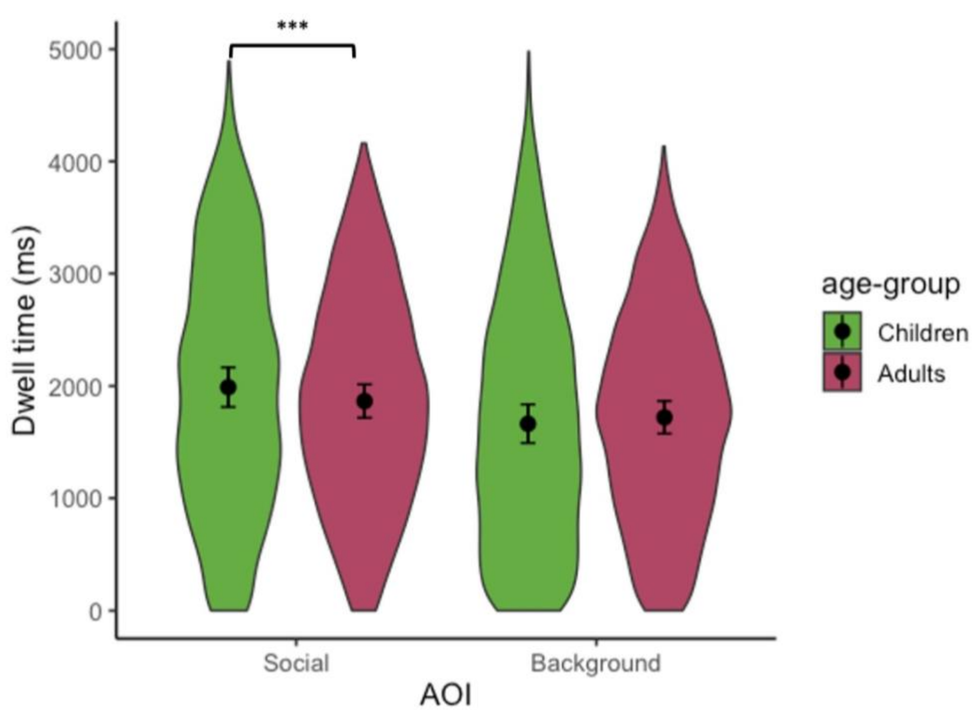

Figure 9. Violin plot for mean dwelling time for AOI and age-group. Error bars represent 95% confidence intervals.

Finally, the three-way interaction between age-group, type of scene and type of AOI not only did it not improve the model fit, but also produced a non-significant effect ($F(1,244) = 0.00$, $p = .99$, $\eta^2_p = 0.00$) (Figure 10; see Supplementary materials n. S4 for descriptive statistics).

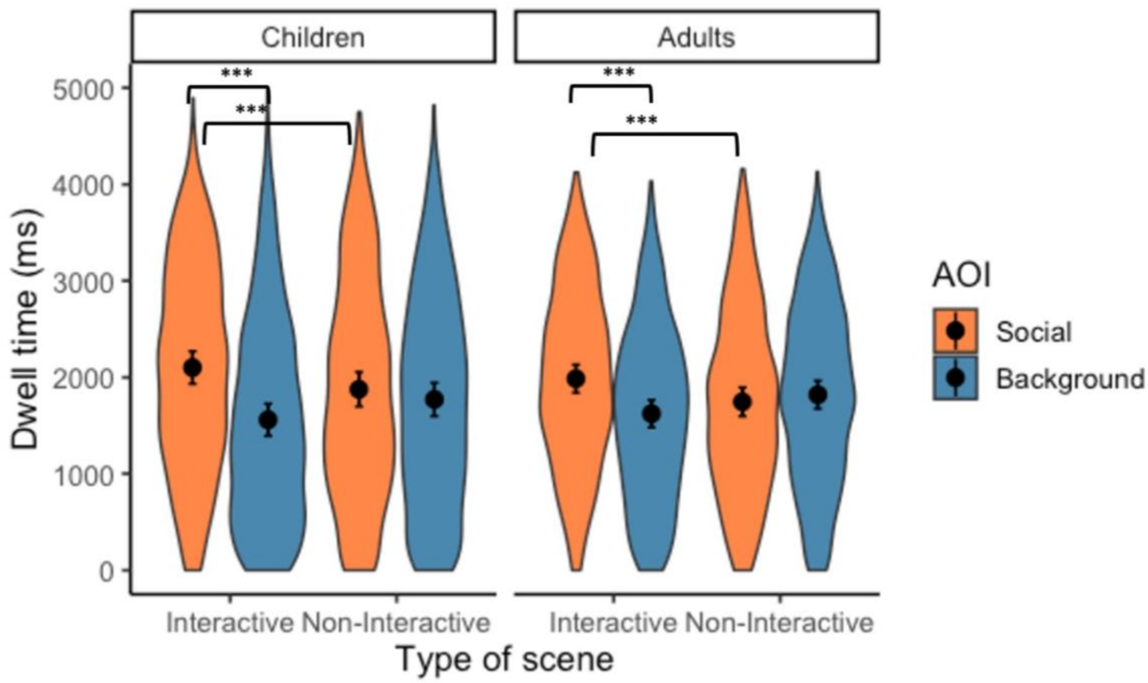

Figure 10. Violin plot for mean dwelling time for type of scene and AOI, in the two age-groups. Error bars represent 95% confidence intervals.

Attentional capture

Time to first fixation was transformed (logarithm in base 10) to meet multilevel modelling assumptions. However, after transforming the data, the structure of the model and the results did not change, therefore we present here the untransformed data to facilitate understanding and interpretation of the results. See Supplementary materials n. S2c for details on the transformation and the full analysis and results from transformed data.

We used the same hierarchical structure with age modelled as a categorical predictor as above to assess possible differences between children and adults in attentional capture. Here, we see a significant main effect of age-group, $F(1,122) = 5.75$, $p = .02$, $\eta^2_p = 0.04$ with children ($M =$

709.22 , *SD* = 746.03) being slower to orient attention to the scene compared to the adults (*M* = 674.86 , *SD* = 719.30).

The main effects of scene ($F(1,122) = 0.01$, $p = .94$, $\eta^2_p < 0.001$) and AOI remained unchanged ($F(1,244) = 713.61$, $p < 0.001$, $\eta^2_p = 0.75$) with faster orienting towards social information (*M* = 498.74, *SD* = 541.70) compared to non-social information (*M* = 882.59, *SD* = 839.11), irrespective of age group and scene type – i.e. interactive or not. Similarly, the interaction between scene type and AOI was maintained across groups, $F(1, 244) = 27.58$, $p < .001$, $\eta^2_p = 0.10$ with slightly faster orienting to social information in interactive scenes and faster orienting to non-social information in the non-interactive scenes (Figure 11).

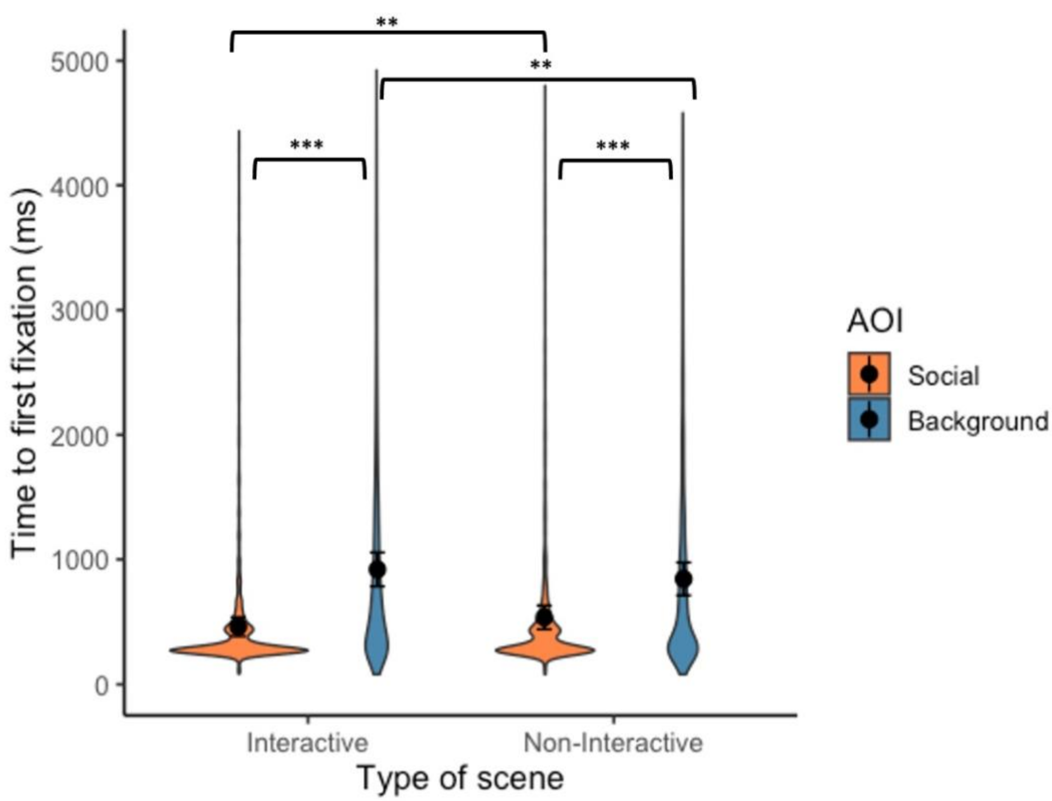

Figure 11. Violin plot for mean time to first fixation for type of scene and AOI. Error bars represent 95% confidence intervals.

As in the dwell-time analyses, there was no interaction between age group and type of scene ($F(1,122) = 0.03$, $p = 0.87$, $\eta^2_p < 0.001$. In contrast to the dwell-time analyses, there was also no interaction between age group and type of AOI ($F(1,244) = 0.00$, $p = .99$, $\eta^2_p = 0.00$), with children and adults being similar in orienting to different types of information in the scenes – i.e. social or non-social. Finally, as in the dwell time analysis, the three-way interaction between age-group, type of scene and AOI did not improve the model fit, and when included in the model, did not produce a significant effect ($F(1, 244) = 0.80$, $p = .37$, $\eta^2_p = 0.00$ (Figure 12; see Supplemental materials S4 for descriptive statistics).

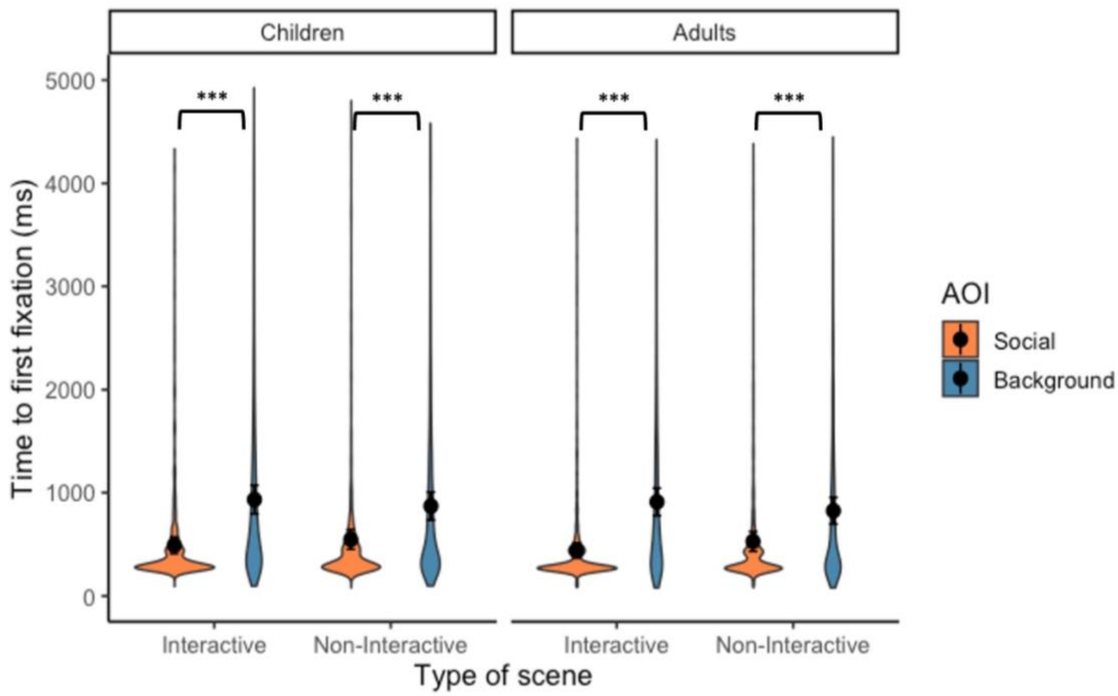

Figure 12. Violin plot for mean time to first fixation for type of scene and AOI, in the two age-groups. Error bars represent 95% confidence intervals.

General discussion

Here, we demonstrate both a general attentional bias towards social information in complex scenes, and that this bias is increased in the presence of a social interaction in *both* adults and in pre-adolescent children. Indeed, although children spend more time looking at social information across all scenes, regardless of interactive content, they show a similar pattern to adults in both how their attention is captured by social information and how they prioritise social information more in interactive, as compared to non-interactive, scenes.

The general social bias in both children and adults is consistent with much previous literature showing a general attentional preference towards social information (e.g., Bindemann et al., 2010; Doherty et al., 2017, 2019; Sue Fletcher-Watson et al., 2008; Rösler et al., 2017; Sasson & Touchstone, 2014; Van Der Geest et al., 2002). Our results in this regard contrast with prior research in two ways. First, we find that children engage with social information more than adults do across scene type, a result potentially driven by the youngest children. This contrasts with some prior findings that suggest a slow increase in social attentional engagement across childhood and adolescence (Amso et al., 2014). This finding may suggest either than children need more processing time to understand social content in complex scenes or that social elements play a more important role in scene perception in children than in adults. Secondly, our findings suggest that there is no developmental change between middle-childhood and young adults in how much more quickly attention is captured by social than non-social information. This finding contrasts somewhat with Doherty et al. (2019) who, during a non-social visual search task, show greater attentional capture from social information in pre-adolescent children compared to adults. Importantly, however, the two studies measure different aspects of social attention. Indeed, while we use a free exploration task, Doherty et al. use a visual search task in which attention to social distractors must be inhibited to correctly perform the task. Thus, in contrast to our data, Doherty et al. may be measuring age differences in the ability to suppress interference from social information while doing a non-social task rather than how attention is spontaneously captured by social information in the absence of another task. Interestingly, the possible influence of general developmental mechanisms is echoed in our current finding of children being generally slower to orient to the scenes compared to the adults, likely due to development differences in processing speed which continues to increase into adolescence (Luna et al., 2008), and ongoing development of the orienting attentional network in childhood (Pozuelos et al., 2014).

This work also contributes to a small but growing body of work suggesting that observed social interactions may be processed differently than the same number of individuals engaged in independent tasks (Papeo & Abassi, 2019; Vestner et al., 2019; Walbrin & Koldewyn, 2019). The current work extends previous research by looking at the moderating effect of social interactions on social attention in childhood. Given that infants are able to distinguish between two people facing compared to non-facing, and prefer the facing pair (e.g., Augusti et al., 2010; Handl et al., 2013), it is perhaps unsurprising that children also attend preferentially to social interactions. We were, however, surprised that the presence of a social interaction increases social attention to human information in scenes just as much in children as it does in adults. That this ‘adultlike’ moderation effect of social interactions is present as early as 6 years of age suggests that prioritized attention to social interactions may be of special importance in social development and social learning. In line with this idea, children as young as 5 are able to learn social norms from observation of others’ behaviors and interactions (e.g. Skinner et al., 2017), suggesting not only perceptual sensitivity to interactions but “higher-order” social cognitive abilities that rely on interactive information from relative early in development.

The lack of developmental change in attention to social interactions both within our child group and between children and adults is somewhat surprising considering the substantial structural (Mills et al., 2014) and functional (Walbrin et al., 2020) differences between children and adults in social brain structures that support social interaction perception. This apparent disconnect suggests that developmental changes in brain regions that support social interaction perception do not reflect changes in spontaneous attentional mechanisms. Instead, they may reflect changes in higher order social-cognitive processing. Indeed, future studies could investigate the extent to which attentional prioritization of interactions in both children and adults reflects automatic processing or a deeper understanding of social information, by investigating how social knowledge relates to social attention patterns across childhood and adolescence. Similarly, it will be important to investigate the extent to which these attentional patterns, and their similarity between children and adults, are maintained in scenes where social interaction information competes with other social information and when participants are performing tasks, rather than simply observing the scene, as in the current study. Understanding how mechanisms of attention to social interactions develops across stimuli and paradigms has the potential to offer important insights into how social brain regions interact and develop to support the detection, discrimination, and selection of important social information in the environment (Amso & Scerif, 2015).

Finally, the fact that the presence of a social interaction strongly moderates engagement with social information but alters attentional capture less strongly, might reflect the possibility that social information captures attention regardless of the clutter in the scene, but engages it for longer only when it is relevant and interesting. Indeed, it would seem that social information first captures attention in the early stages of scene exploration regardless of scene type, and then holds it for longer when the agents are engaged in social interaction. This suggests a hierarchy of social information in scenes, where all human information automatically orients attention and is then further prioritised when an interaction is present. Such hierarchical mechanisms could be further studied by putting different types of social information – e.g interacting and non-interacting individuals – in direct competition for attentional resources within the same scene.

There are, however a few limitations to the current study. In the first place, the developmental analyses would be more strongly supported if we had been able to obtain our intended sample size – as planned from the pre-registered power analyses – which in the current study was not possible due to the interruption of data collection by the ongoing COVID-19 pandemic. While our results show little evidence of developmental differences between groups, we do not have the power to be certain that smaller developmental changes are not present. Additionally, our stimuli were purposefully chosen to reflect a wide range of scenarios. As such, many aspects of the pictures are not well controlled. While this very heterogeneity is also a strength of the study, it remains possible that some unintended visual aspect of the scenes differed systematically between interactive and non-interactive scenes. This heterogeneity also means that it is difficult to pinpoint the visual information and social cues that identify a scene as interactive, or not. Unlike many other studies, these cues are not singular (i.e., facing direction) but instead differ from scene to scene. Similarly, we deliberately used scenes without strong emotional content, yet few 'real-life' interactions are truly 'neutral'. Future studies could investigate the influence of emotional content on attention to social interactions, both to better reflect real life scenarios and attentional patterns and to understand how emotional and interactive content uniquely influence attentional patterns. First, however, future work will need to investigate how the influence of interactive content on attentional engagement is altered when interactions are present in the *same scene* as other social targets. Indeed, in the work presented in Chapter 4 we investigate these mechanisms in multiple people pictures, where social interactions have to directly compete for attention with one or two other non-interacting individuals.

Chapter 4. Development of the attentional priority of social interactions in naturalistic scenes

Abstract

Human visual attention is specialized for capture and engagement of social information in naturalistic scenes, and this pattern of “social bias” is present since infancy. Recent research, additionally suggests that people preferentially attend to and more quickly detect interacting dyads compared to non-interactors. However, very little work has examined interactive mechanisms in complex scenes depicting multiple people, nor how such mechanisms arise and change across development. How does social attention change across development when there is competition between targets? In this work, we recorded eye-movements in 98 adults and 54 children in a free viewing experiment while participants viewed naturalistic scenes where pairs of people were either interacting or not, and were depicted with either one or two additional non-interacting individuals. We find an attentional bias to social information in both engagement and capture, but this bias was not moderated by the presence of an interaction in the scene nor by the age of the participants. However, when interacting and non-interacting humans compete for attention in the same scene, the interacting dyad captures attention more quickly and engaged it for longer when there is one other person in the scene across both age groups. When the competitor is another (non-interacting) pair, only adults show a significant effect of interaction. These results suggest that in complex social scenes with multiple social targets, interactions take attentional priority, but this priority is weaker for children and when social content in the scene increases. These findings and their implication for both scene perception and social development are discussed.

Introduction

As deeply social beings, humans are drawn to and highly skilled in understanding visual information that gives clues about other people’s relationships, intentions and mental states (Quadflieg & Westmoreland, 2019). As a result, attentional processes that support social understanding are honed from an early age, including those that support preferential looking to ‘human information’ like faces, bodies, and actions. Specifically, interacting people are of special interest, as they contain key interpersonal cues and may uniquely support both social understanding and social learning, especially across development (e.g., Quadflieg & Penton-Voak, 2017; Skinner et al., 2017). Additionally, real world scenarios are typically cluttered and involve multiple people, often interacting with each other. Adults are very good at selecting important social information, even when the scenes involve many different physical, social and emotional contexts (e.g., Birmingham et al., 2008; Quadflieg & Westmoreland, 2019). But how do our attentional systems develop to support the detection, selection, and understanding of interpersonal cues in cluttered environments?

The primary purpose in the current study is to investigate how social attention in complex scenarios develops in childhood and how social interactions might influence these patterns of attention. Additionally, we explore how social interactions compete with other social information for human attention. To answer these questions, we use a free exploration paradigm to examine attention to social information (e.g. humans) and especially social interactions, in a variety of naturalistic scenes containing multiple people, in 6-12 years old children and adults.

Most social attention research supports the existence of a strong attentional bias to social information in scenes, as indicated by the fact that we easily and spontaneously orient to and preferentially process bodies, faces and eyes as compared to non-social information (e.g., Bindemann et al., 2010; Birmingham et al., 2009; Doherty et al., 2017; Fletcher-Watson et al., 2008; Mayer et al., 2015). Additionally, the orienting of attention to social information in scenes appears to be automatic and unintentional (Rösler et al., 2017) and can certainly occur orthogonally to the task participants are actually performing (Flechsenhar & Gamer, 2017). Interestingly, much of the developmental research on this social bias has been done with infants, showing that they develop a broad preference for social stimuli during the first year of age (for reviews see Bertenthal & Boyer, 2015; Soto-Icaza et al., 2015). It is very likely, however, that these processes may undergo changes across childhood and adolescence that have not been fully explored (Bertenthal & Boyer, 2015; Soto-Icaza et al., 2015). Indeed, research exploring this social attentional bias during childhood is inconclusive. For example, when performing a non-social visual search task that included both social and non-social distractors, children and adults were both more sensitive to the social distractors, but children’s attention was more often captured by social information compared to the adults (Doherty et al., 2019). This would suggest that children's attention is *more* sensitive to social information in naturalistic pictures yet Amso et al. (2014) show a mild increase from 6 to 12 years in the proportion of social information that children attend to in a free viewing paradigm, suggesting the social bias *increases* across development. These contrasting results, albeit from quite different paradigms, suggest further investigation of the social bias in development is needed, especially regarding cluttered naturalistic scenarios.

What's more, much of our understanding of the 'human bias' in attention comes from work on isolated single individuals, separated from other social agents. This literature has provided invaluable insights into how we select social information and is an excellent starting place for understanding the mechanisms that support our perception of more complex social scenarios, such as interactions between two or more people (Papeo, 2020). Indeed, observed social interactions provide social cues that are rarely if ever encountered when observing individuals, including cues that provide clear understanding of social roles, relationships between people, and the social intentions of future interaction partners (Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017). Interestingly, recent research shows that we might process interacting pairs as more than just the sum of two agents, both in the brain (Isik et al., 2017; Walbrin & Koldewyn, 2019) and in behavioural tasks (e.g., Ding et al., 2017; Papeo et al., 2017). Such results suggest that human perception includes specialisation for detecting and processing *interactive* information. Additionally, tightly controlled behavioural studies show a search advantage for dyads that face each other compared to non-facing pairs (Papeo et al., 2019) or pairs that face in the same direction (Vestner et al., 2020), although this advantage may be driven by gaze-based spatial attentional cuing, rather than specifically social or interactive cues (Vestner et al., 2020). Further, when participants were presented with pairs of facing or non-facing dyads at the same time, they spent more time looking at the facing pairs compared to the non-facing pairs (Stagg et al., 2014), suggesting an attentional preference for interacting dyads compared to non-facing individuals. Although this evidence suggests an attentional bias towards social interactions, so far this research has relied almost entirely on tightly controlled stimuli that show figures and dyads taken out of context. To our knowledge, only one study has investigated attention to dyadic interactions in naturalistic scenes, showing that attention to human information is increased when viewing a scene that includes a dyadic social interaction, compared to scenes with two non-interacting people (Skripkauskaite et al., under review). These findings suggest that interactive information can drive attention in complex scenes, effects that may differ when scenes contain multiple social targets and when interactive information is in direct competition with other social targets.

There is also relatively little research explicitly investigating how attention towards social interactions may change across development, especially during middle to late childhood. What little evidence there is suggests that the ability to discriminate between interacting and non-interacting humans develops as early as 4 months old, with infants being able to shift their attention between speakers that face each other compared to non-facing speakers (Augusti et al., 2010; Handl et al., 2013) and that they can even discriminate between different kinds of interactions (Hamlin et al., 2007). Although these findings suggest that infants and children have the ability to detect social interactions, we currently have only limited knowledge about the attentional mechanisms underlying this skill in childhood. In chapter 3 we've shown that just like the adults, when two people are interacting in a scene, attention to social information is increased compared to the scenes in which they are not interacting. Additionally, to our knowledge, only one study specifically investigated attention to interactions in development. Stagg et al. (2014) investigated attention to interacting pairs in childhood, demonstrating that 9-years-olds looked for longer at two agents facing each other than pairs of agents who were not facing when the two pair-types were presented together. Although this suggests that there is an attentional bias to social interactions also in pre-adolescent childhood, age related changes in this pattern of attention in much more complex scenes have not been investigated. Additionally, investigating attention to social interaction in naturalistic complex scenes may reveal important extra information about developmental changes in social attention as contextual information could either facilitate processing of social interaction or, on the contrary, hinder such mechanisms through competition for limited attentional resources.

There is good reason to think that there may be a developmental change in the way children process and attend to social interactions, as the brain systems that support this processing (e.g. Walbrin et al., 2018) are not yet adult-like in pre-adolescent children both structurally (Mills et al., 2014) and functionally (Sapey-Triomphe et al., 2017; Walbrin et al., 2020). If the neural system underlying social interaction perception is not yet fully "tuned" at 6, this could have 'carry-over' consequences for how children and adolescents detect, attend to, and process social interactions, especially when faced with complex naturalistic scenes.

Work investigating social attention in neurodevelopmental disorders also supports the idea that social interactions in a scene may change how children attend to it. Although social attention processes are altered in quite predictable ways in neurodevelopmental disorders such as ASD and Williams syndrome (Chita-Tegmark, 2016; Frazier et al., 2017; Klin et al., 2002; Riby & Hancock, 2008; Williams et al., 2013), the nature and extent of these behavioural or attentional differences appears to vary substantially depending on the complexity and type of information being presented. Indeed, a recent meta-analysis suggests that although autistic participants show overall reduced social attention across a variety of experimental stimuli and tasks, the only factor that predicted the effect sizes of such reductions was the amount of 'social content' in the scene (e.g., the number of people depicted in the stimuli) (Chita-Tegmark, 2016). Additionally, another meta-analysis suggests that the largest differences between ASD and typically developing (TD) groups in attention to social areas of interest like faces and eyes were driven by the presence of a social interaction in the scene (Frazier et al., 2017). This evidence suggests that the presence of a social interaction in a socially cluttered scene, depicting multiple people, might influence attention orienting in unexpected ways, and unveil developmental change in social attentional orienting.

Interestingly, very little research has manipulated either social content or the presence of social interactions explicitly. Research that has tackled such questions most often has used tightly controlled stimuli with isolated figures taken out of context. However, although experimental control over stimuli is important, the social world is complex and rich and using naturalistic scenes has the advantage of more closely mimicking the competition for attention – between social and non-social clutter in the first place, and between different social agents in the second place – that is usual in everyday experience. Studies have rarely investigated how increasing the social content of a scene by adding *additional people* nor how manipulating the interactive content through including *social interactions* within the scene might affect social attention. To our knowledge, there is only one study that has done so. During a free viewing paradigm, Birmingham et al., (2008) used scenes containing either one or 3 people who were either active (doing something independently), inactive (doing nothing independently) or interacting (the 3 people did something together). Interestingly, the authors showed that increasing the number of people in the scene, especially if the 3 people were active (independently or interacting), increased the attention to the eyes of the agents, indicating a change in the pattern of attention when multiple people are involved. However, they show no difference between "interactive" and "independent" active scenes in fixation proportions across ROIs (eyes, heads, bodies, foreground and background objects), suggesting no influence of "interactiveness" on attention within the scene.

Finally, developmental research on multiple people scenarios and interactions is very limited. When viewing static or dynamic scenes, children (6-8 years old) shift their attention from faces to bodies as the number of people in scene increases, and this shift is more marked in children than in adults (Stoesz & Jakobson, 2014), but it is unclear whether the type of social content matters and how this changes in development.

To summarize, on one side, while it is established that there is a human attentional bias in naturalistic scenes that is present from early infancy, there is little evidence about how that bias may change during childhood and adolescence. Similarly, neuroimaging and behavioural evidence support a perceptual bias towards interacting dyads, which again, may change during childhood. In particular, it is still unclear how, and if, social interactions might compete with other social information for human attention and, if so, to what extent this competition might be different in pre-adolescent childhood.

In the present study, we investigate spontaneous social attention in complex naturalistic scenes containing more than two people across development, and explore the strength of the bias to social interactions when other social targets (e.g., other human agents) are present in the same scene.

In the light of previous literature, we expect to find: i. a stronger bias to humans in adults compared to children, across all scenes; ii. that the bias to human figures in naturalistic scenes is moderated by the presence of a social interaction in the scene, with a stronger effect in adults compared to children; and iii. that both groups will give interacting dyads attentional priority, when presented together with another person or a pair of non-interacting agents, and that this effect will again be stronger in adults compared to children.

Methods

Participants

Two groups participated in this study, an adult group and a group of children aged 6 -12 years. Power analyses were performed in G Power (Erdfelder et al., 2009; Faul et al., 2007) and were pre-registered on AsPredicted.

For the adults group (pre-registered on AsPredicted: https://aspredicted.org/blind.php?x=hz9ay7) (Appendix A3 and A4) data collection was planned to proceed until we would reach the sample size indicated by the power analysis – 231 participants necessary to reach 80% ($\alpha \leq .05$) and a medium effect size (Cohen's f = 0.25) – or until 31st of July 2020, whichever would come first. We recruited 101 participants, of which 3 were eliminated, one due to sleepiness and two for being out of our desired age range. Our final adult group included therefore a total of 98 adults (*M* = 21.15, *SD* = 2.97; range = 18-35; 70 females, 1 other). This sample size is enough to reach a large effect size (Cohen's f = 0.40), with 80% power ($\alpha \leq .05$).

The developmental power analysis suggested 90 participants in order to detect a large effect size (f = 0.40) and reach 80% of power ($\alpha \leq .05$), or the pre-registered plan to cease data collection at the end of the academic year (21st of July 2020), whichever would come first (AsPredicted: https://aspredicted.org/blind.php?x=4ju8ug). We therefore collected data from 54 children (*M* = 8.76, *SD* = 1.72; range = 6-12; 28 females) who were all included in the final developmental group. All participants had normal to corrected vision. Adults gave informed consent, and received money or university credits as compensation for their participation. Children gave assent, and each child's guardian gave consent for them to participate, and they received toys as compensation. All procedures were approved by the ethical committee at Bangor university (ethics protocols: 2018-16360 and 2019-16586).

Stimuli and apparatus

The stimulus pictures were selected from 4 online databases (Choi et al., 2014; Kosti et al., 2017; Quattoni & Torralba, 2009; Xiao et al., 2010). They were chosen to be emotionally neutral, and to depict a variety of ordinary life contexts (e.g. schools, shops, markets). Scenes contained 3 or 4 people where either all were not engaged in social interactions, or where one dyad was interacting, while either 1 or 2 other people were not.

Twenty-six independent judges rated the pictures for interactiveness on a 1-7 Likert scale and were then asked to explicitly indicate the interacting dyad in each scene, if any (see

“Materials – Picture selection for chapter 4” section in the General methods chapter for details on the rating experiment). First, out of the initial 114 selected pictures (64 depicting three people and 50 depicting four people) the lowest-scoring 33% of the pictures (cut-off interactiveness score ≤ 2.08) were chosen for the non-interactive condition, and the highest-scoring 33% (cut-off ≥ 4.15) were chosen for the interactive condition. The final set of pictures, which included 22 pictures depicting 4 people and 30 pictures depicting 3 people, were selected from these pools to include equal numbers of interactive and non-interactive pictures that also had the highest level of agreement across judges regarding which was the interacting dyad in the scene, if any, (rater agreement ≥ 65%).

All pictures were matched for colour with one sample picture using Photoshop (version CC 2019), neutralised to remove colour cast (“adjustments – match colour – neutralize colour” option) and sharpened (“filter – sharpen” option).

Stimuli were presented on a 1920 x 1080 px screen, on grey background, using Psychopy 2 (Peirce et al., 2019). Each picture had a size of 860 x 860 pixels (13.6° x 13.6° visual angle) and was presented in a location shifted either right or left of center, with the centre-most edge 60 pixels left or right of the fixation cross (0.85° visual angle).

An EyeLink Portable Duo tracker (EyeLink x, SR Research, Ontario, Canada) with remote binocular system at 1000 hz sampling rate was used to collect data from both eyes, but monocular data were used for the analyses. Which eye’s data was used was based on each participant’s individual calibration accuracy.

Procedure

Participants sat comfortably on a stable chair approximately 80 cm distant from the screen. They viewed a total of 142 pictures in randomised order, in 4 blocks of 35 trials each. These blocks not only contained the pictures for this task, but also pictures belonging to other 2 experiments not discussed here. Breaks were allowed between blocks, after which another drift correction procedure was performed.

Adults were instructed through on-screen written instructions to freely observe the pictures for 5 seconds, with no other specific instructions, while children were verbally instructed to do so.

Before the task, a 13 point calibration procedure was carried out for participants in both groups. Instructions appeared on the screen for the adult group, while the calibration procedure was explained verbally for children.

For adults, a drift correction procedure was carried out before each picture was presented, where participants were asked to fixate a calibration point at the centre of the screen before pressing the space bar to proceed through the task. Children were presented with an animated gif at the centre of the screen for a duration of 2.5 seconds as a fixation point between each stimulus. In both cases, this served to draw participants' gaze back to the centre of the screen before the beginning of the next trial. The experiment took around 20 minutes to complete.

Data analysis

Two areas of interest (AOIs) were defined for each picture using the "freehand" option in Eyelink Data Viewer (SR Research, 2013). One AOI included all visible human information (faces, bodies) that was not occluded by other objects while the other AOI consisted of all other information in the scene (objects, background elements). We extracted a measure of dwelling time in each AOI (with fixations and saccades) as a measure of attentional engagement. We measured attention capture as the time to first fixation for each area of interest.

Each participant viewed every picture, which were chosen so we could assess the influence of 3 factors: number of people in the scene (3 or 4), whether a two-person interaction is taking places in the scene (interactive or not) and the AOI type (humans or background).

Trials with less than 33% of total engagement time with the stimulus were considered missing. These trials included off-screen looking time, poor signal, missing data, and blinks.

We used multilevel modelling (nlme package in R, Pinheiro et al., 2016) to perform two different levels of analyses to answer our two main research questions. The first level analysis assessed developmental change in the attentional bias to humans in complex scenes and the extent to which this bias was moderated by social interactions across all the pictures. The second level analysis investigated the attentional competition between social interactions and other social information, so only interactive pictures were entered into this analysis.

Unless stated otherwise, time to first fixation data in all the parts of the analysis were transformed to meet multilevel modelling assumptions. When the model with transformed data and the one with untransformed data did not differ in structure and outcome effects, we present the untransformed data for ease of understanding. Details about all transformations and analyses using the transformed data are in the Supplementary materials n. S5 (Appendix D).

Part 1: Development of the attentional social bias in cluttered scenes

Aim

The first goal of this work was to investigate developmental changes in the attention to social information compared to non-social information – i.e. the social attentional bias – and the role of social interactions in moderating this bias also in more complex scenes, independently of the number of the people in the scene – i.e., 3 and 4 people scenes were merged.

Data analysis

The data cleaning procedure described above led to the loss of 0.31% of trials within the adult group (range of 0 -1 trials per participant), and 3.24% of trials within the child group (range of 0 - 4 trials per participant).

We used a four-level hierarchical model for each of our two measures (dwell time; time-to-first-fixation). The structure of the model was as it follows: participant information modelled at the fourth and highest level, the social content of the scene – i.e., whether an interaction was present or absent - was nested within each participant at the third level, and the type of AOI (human, background) was nested at the second level; the first level included either the measure of dwelling time or time to first fixation for each AOI, trial and participant. Age-group (child, adult) was modelled as a categorical predictor. Additionally, pairwise comparisons were performed using post-hoc Tukey's HSD with the emmeans package in R (Lenth et al., 2018).

Results

Attentional engagement

The relationship between conditions and dwell time showed significant variance in intercepts across participants, type of scene and area of interest (SD = 323.86, $\chi^2(3)$ = 851.17, $p < .001$).

After setting participant, condition and AOI as random effects, we found that overall attention to the scenes was similar between children (*M* = 1831.97, *SD* = 1117.75) and adults (*M* = 1801.43, *SD* = 952.02), resulting in a non-significant main effect of age-group ($F(1,150)$ = 1.23, p = .27, $\eta^2{}_p$= .01). Similarly, equal time was spent looking at the interactive (*M* = 1818.51, *SD*= 1015.21,) and the non-interactive scenes (*M* = 1805.62, *SD* = 1010.66) ($F(1,150)$ = 0.29, p = .59, $\eta^2{}_p$= .002), and this was similar for children (interactive: *M* = 1842.43, *SD* = 1112.06; non-interactive: *M* = 1821.50, *SD* = 1123.53) and adults (interactive: *M* = 1805.72, *SD* = 959.27; non-interactive: *M* = 1797.13, *SD* = 944.79); $F(1,150)$ = .05, p = .82, $\eta^2{}_p$ < .001.

As predicted, all participants showed a social attentional bias, with more attention given to human areas of interest (*M* = 2033.90, *SD* = 998.66) compared to the background (*M* = 1590.23, *SD* = 977.97) – $F(1, 300)$ = 342.88, p < .001, $\eta^2{}_p$= .53, and this effect was moderated by age group – $F(1,300)$ = 10.87, p = .001, $\eta^2{}_p$= .03. Indeed, children looked at background information (*M* = 1664.75, *SD* = 1091.21) for longer compared to the adults (*M* = 1550.38, *SD* = 909.25), $t(150)$ = -3.11, p = .01, d = - 0.25, while attention to human information was similar between the two groups (children: *M* = 1999.19*, SD* = 1119.16; adults: *M* = 2052.47, *SD* = 927.40), $t(150)$ = 1.55, p = .41, d = 0.13 (Figure 1).

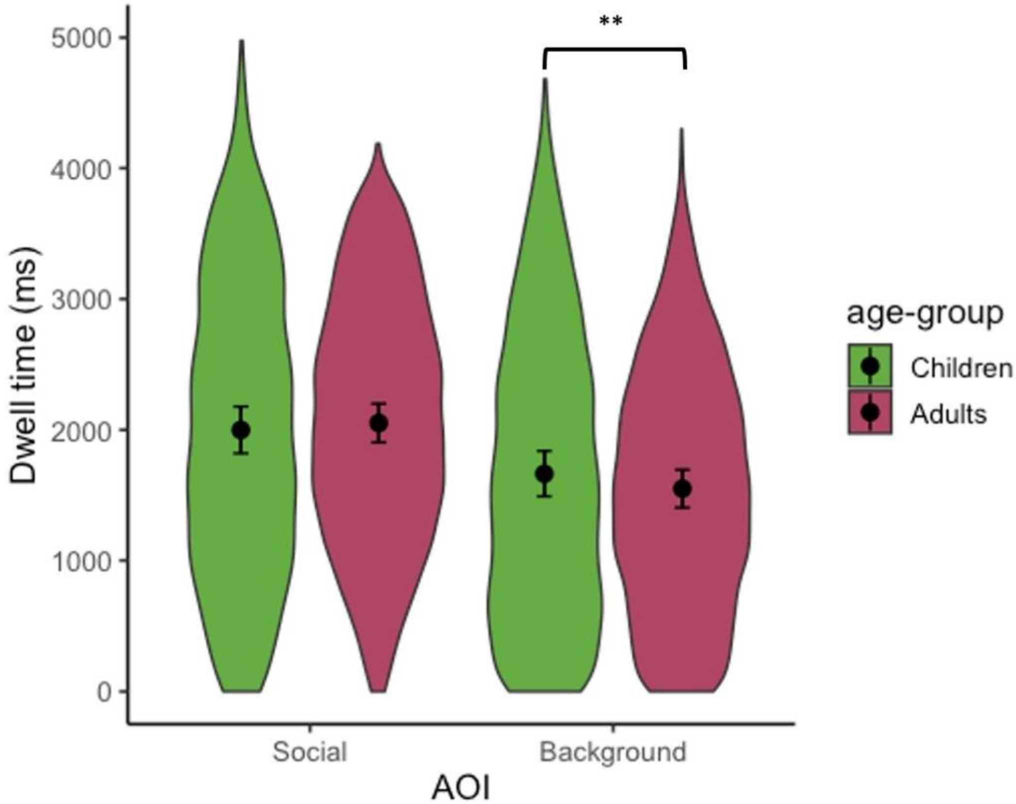

Figure 1. Violin plot for mean dwell time for AOI and age-group. Error bars represent 95% confidence intervals.

Finally, there was a significant interaction between type of scene and region (F(1,300) = 18.36, p < .001, η^2_p= .06). In particular, surprisingly, more attention was given to the social AOIs in the non-interactive scenes (M = 2078.55, SD = 980.87) compared to the interactive scenes (M = 1989.29, SD = 1014.29), t(150) = 2.44, p = 0.06, d = 0.20, and more attention was given to the background in the interactive (M = 1647.73, SD = 987.13) compared to the non-interactive (M = 1532.69, SD = 965.41) scenes, t(150) = 3.26, p = .01, d = 0.27, while still presenting a strong human bias in both types of scenes (interactive: t(300) = 8.97, p < .001, d = 0.52; non-interactive: t(300) = 14.67, p < .001, d = 0.85).

Additionally, this effect was similar in the two age-groups, as the three-way interaction between age-group, condition and AOI did not improve the fit of the model, and was not significant (F(1,300) = .05, p = .83, η^2_p< .001) (Figure 2; see Supplementary materials n. S1, Table S1a for descriptive statistics).

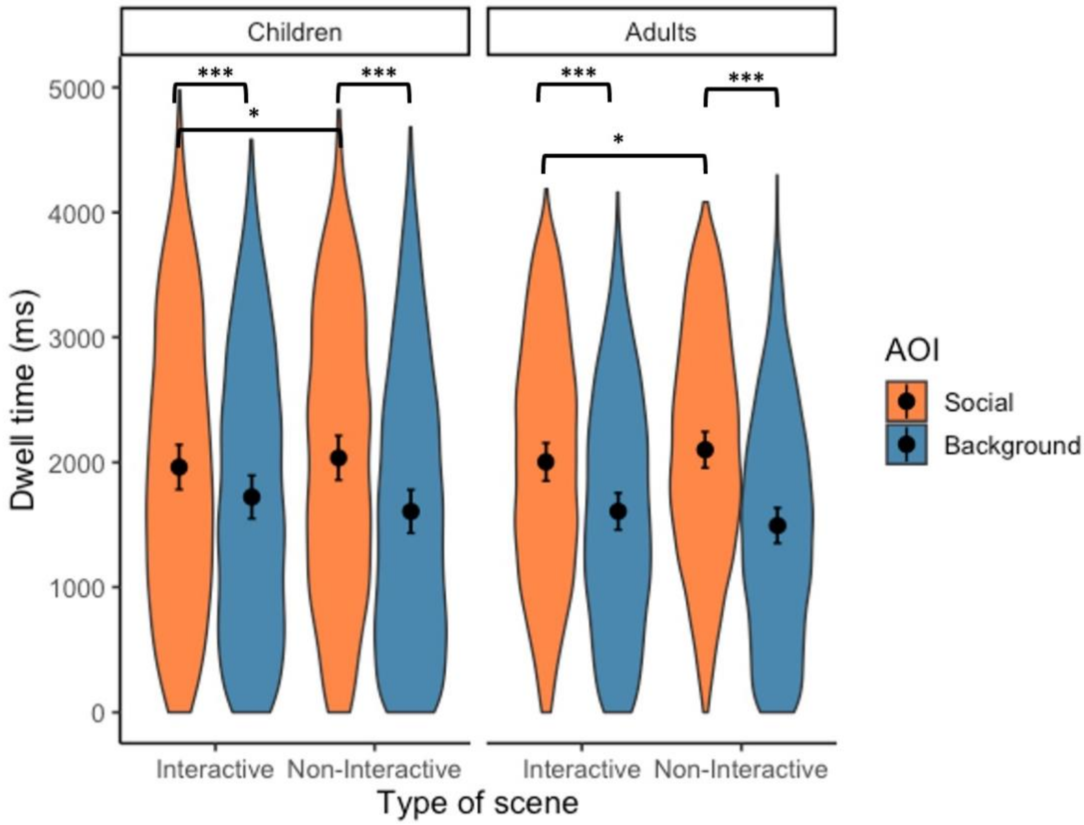

Figure 2. Violin plot for mean dwell time for children and adults across types of scenes and AOIs. Error bars represent 95% confidence intervals.

Exploratory analyses

Since the presence of a social interaction in the scene moderated attention to human information in the opposite direction compared to our prediction, we conducted exploratory analyses to investigate the potential cause for these results. First, we investigated which pictures, between the three and the four people scenes, are driving this effect. Therefore, we introduced the number of people in the scene as a further fixed effect to a similar hierarchical model to the one used above, with the two age-groups merged in one sample (see Supplementary Materials n. S2 – 2a for details). The analysis revealed that the "inverse" interaction moderation was driven by the three people pictures, while the same effect was not found in the four people pictures (Figure 3). In particular, while in the three people pictures significantly more attention was given to the non-interactive humans (M = 2169.10, SD = 1003.61) compared to the interactive (M =

1938.91, *SD* = 1010.47), $t(151) = 5.92$, $p < .001$, $d = 0.48$, the same is not true for the four people pictures (non-interactive: $M = 1955.49$, $SD = 935.45$; interactive: $M = 2057.78$, $SD = 1015.76$; $t(151) = -2.39$, $p = .20$, $d = -0.19$). Additionally, in the three people pictures, more attention was given to background information in the interacting scenes ($M = 1670.51$, $SD = 997.72$) compared to the non-interactive ($M = 1424.11$, $SD = 998.06$) - $t(151) = -6.33$, $p < .001$, $d = -0.52$ – while in the four people pictures there was no difference (interactive: $M = 1616.77$, $SD = 972.01$; non-interactive: $M = 1680.25$, $SD = 898.67$; $t(151) = 1.47$, $p = .84$, $d = 0.12$).

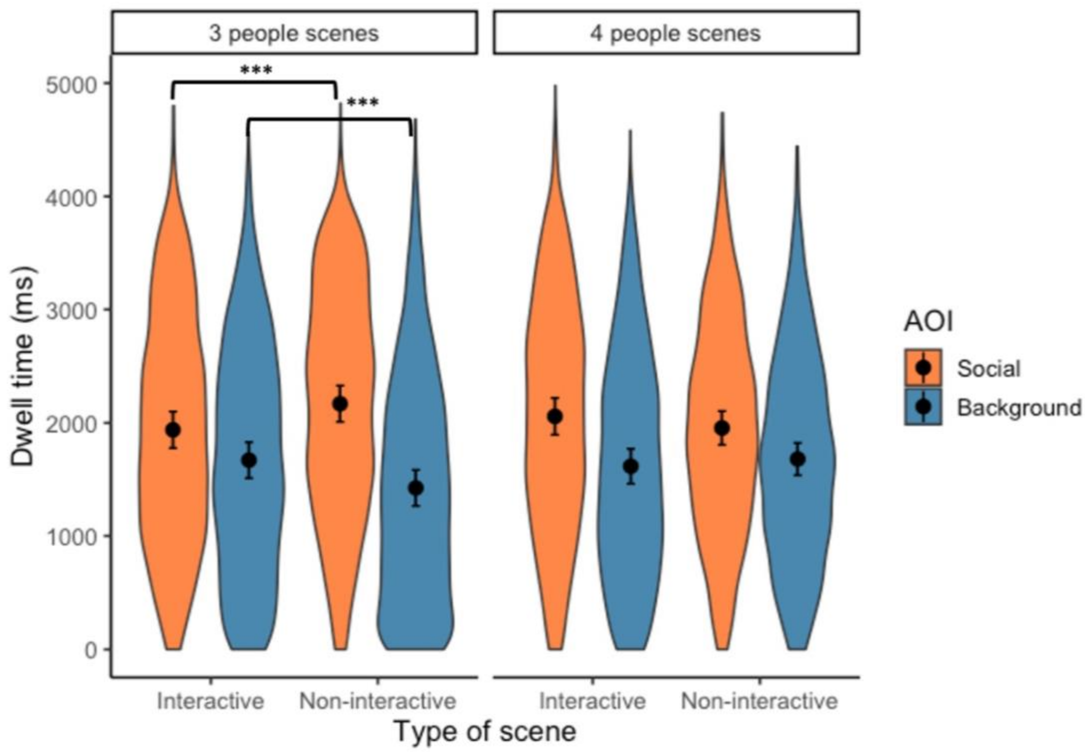

Figure 3. Violin plots for mean dwell time for AOIs across types of scenes. All contrasts social >background were significant at $p < .001$. Error bars represent 95% confidence intervals.

Considering the unexpected effect was driven by the pictures depicting three people, we explored further the effect of another factor on these patterns of attention: size of the AOIs in each scene (see Supplementary materials n. S3 – 3a for details on the size of the AOIs in the 3 and 4 people interactive and non-interactive scenes).

We performed a similar hierarchical model on the three people pictures, with participant, type of scene and AOI as random effects, and type of scene, AOI and size of the AOI as fixed effects, to investigate the role of the size of the AOI in moderating social orienting in interactive and non-interactive scenes (see Supplementary Materials n. S3 – 3b for details). Unsurprisingly, the AOI area had an effect on the dwelling time (F(1,8372) = 2969.39, p < .001, $\eta^2{}_p$ = .26) showing that the larger the AOI, the more attention it received. This was not, however, moderated by either type of scene (F(1,8372) = 0.25, p = .61, $\eta^2{}_p$ < .001), AOI (F(1,8372) = 0.41, p = .52, $\eta^2{}_p$ < .001) or both (F(1,8372) = 0.94, p = .33, $\eta^2{}_p$ < .001), suggesting that AOI size in the three people pictures cannot explain the inverted interaction effect found on social orienting in these scenes.

The second factor we considered as a potential contributor to this effect was the idea that interacting dyads may be processed differently compared to independent individuals. Indeed, a growing literature suggests that interacting dyads are processed as unique "gestalts" (e.g., Ding et al., 2017b; Papeo et al., 2019; Walbrin & Koldewyn, 2019). In other words, interacting dyads may be processed as a single 'unit' and treated as one object. If true, the area *between* them might attract more attention than other regions of the background, including the space between non-interacting people. In the way that AOIs were defined, we had originally included this space between interactants in the background AOI. To explore the possibility that the space between interactants might be attracting more attention other areas of the background, we performed an analysis on the attention given to the space between the two interacting agents in pictures depicting 3-people, the space between the interacting and the non-interacting individual (mixed space) in the interactive scenes, and the average space between the non-interacting individuals in the non-interactive pictures. We expected that holistic processing of the interacting agents compared to a piecemeal visual scan of the non-interacting individuals would result in the space between interactants receiving more attention when participants fixated the centre of the 'object', while mixed and non-interactive space would not receive such 'extra' attention because participants would be scanning individuals instead. Therefore, we expected the interactive space would show higher dwell times compared to either of the non-interactive spaces (see Supplementary Materials n. S4 for details on how the spaces were created). In a hierarchical model including participant information as a random effect, and type of AOI (interactive space, mixed space, non-interactive space) as a fixed effect, we find a main effect of AOI (F(1,4932) =

55.34, $p < .001$, $\eta^2{}_p = 0.02$), with attention to the interactive interpersonal space ($M = 382.88$, $SD = 498.89$) higher than the mixed space ($M = 226.93$, $SD = 417.69$), $t(4932) = 10.01$, $p < .001$, $d = 0.14$ and the non-interactive space ($M = 270.80$, $SD = 351.73$), $t(4932) = 8.09$, $p < .001$, $d = 0.12$, and more attention given to the non-interactive space compared to the mixed space, $t(4932) = 3.07$, $p = .01$, $d = 0.04$ (Figure 4).

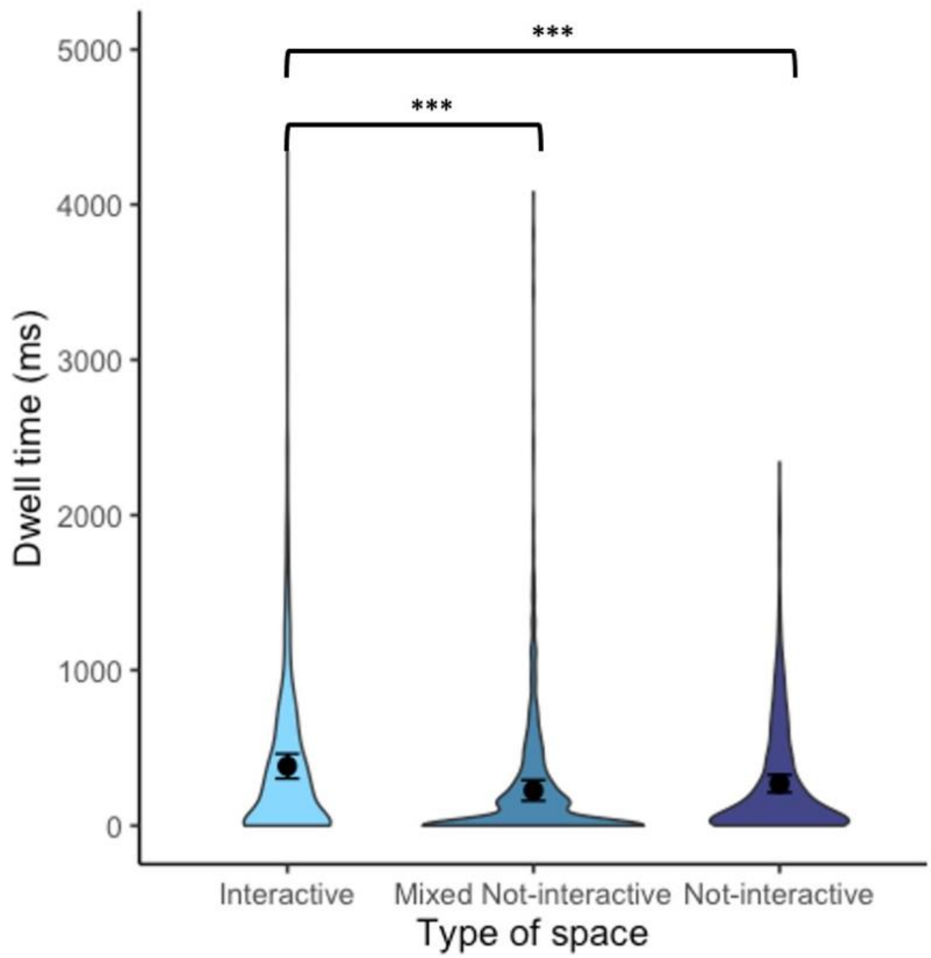

Figure 4. Violin plot for mean dwell time for interpersonal spaces. Error bars represent 95% confidence intervals.

Unfortunately, because of the naturalistic and heterogeneous nature of the scenes used in this study, the number of pictures depicting 4 people that showed a clear configuration of interpersonal space (e.g., non-interacting people were sometimes on opposite sides of the interacting dyad) was not large enough to perform the same analysis, though it would have been informative to contrast the interpersonal space between interacting dyads and non-interacting individuals.

As the space between interactors might be driving the unexpected results in the moderation of social orienting by social interactions, we then also investigated whether this pattern was similar between our age groups. We first performed an exploratory analysis on all pictures with a model with a 2 (age-group) x 2 (people in the scene) x 2 (type of scene) x 2 (AOI) structure of predictors to investigate whether the pattern found in the merged groups – i.e., more attention to humans in the non-interactive scenes compared to the interactive scenes in the 3 but not in the 4 people pictures, is similar between adults and children. The results suggest that these effects are different between children and adults, showing a significant four-way interaction ($F(1,14978) = 6.47$, $p = .002$, $\eta^2_p < .001$; details of the analysis are in the Supplementary materials n. S2 – 2b) between age-group, number of people in the scene, type of scene and AOI (Figure 5). In particular, the inverse interaction effect in the three-people scenes was present only in adults, with more attention to humans in the non-interactive scenes compared to the interactive scenes ($t(150) = 5.77$, $p < .001$, $d = 0.47$), and more attention to the interactive background compared to the non-interactive ($t(150) = 5.78$, $p < .001$, $d = 0.47$). However, this was not true in the four people pictures for adults (human: $t(150) = 2.77$, $p = .18$, $d = 0.23$; background: $t(150) = 2$, $p = .79$, $d = 0.16$) and also not true for children, as they spent the same amount of time looking at interactive and non-interactive human information ($t(150) = 2.22$, $p = .59$, $d = 0.18$) and at interactive and non-interactive background ($t(150) = 2.9$, $p = .13$, $d = 0.24$) in both three-people and four-people scenes (human: $t(150) = 0.28$, $p = 1$, $d = 0.02$; background: $t(150) = 0.22$, $p = 1$, $d = 0.02$; Table S2d in Supplementary materials n. S2 for descriptive statistics). It is important to note, however, that children numerically show the same pattern as adults and that this analysis was already exploratory (i.e., a post-hoc, unplanned analysis), and that we did not have adequate power to be confident about the results of this analysis, particularly considering that the key contrast is a four-way interaction. Thus, these results need to be interpreted with considerable caution and replicated in future work.

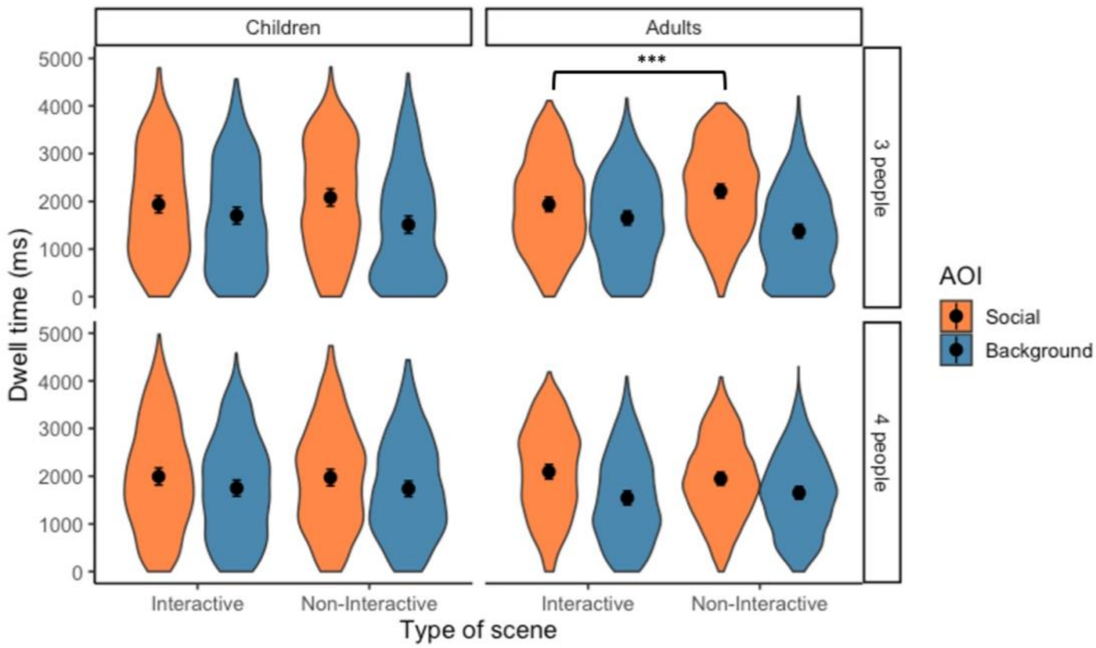

Figure 5. Violin plot for mean dwell time (ms) between age-groups, 3 and 4 people scenes, type of social content, and human and background information. All contrasts human – background are significant at $p < .001$. Error bars represent 95% confidence intervals.

These results suggest that children might show less holistic processing of interacting dyads than adults. To look at this possibility, we performed a final exploratory analysis which is a repeat of the analysis shown in Figure 4, but adding age-group as a factor. Contrary to our expectation, the main effect of AOI ($F(1,4930) = 55.37$, $p < .001$, $\eta^2_p = 0.02$ did not interact with age ($F(1,4930) = 1.71$, $p = .18$, $\eta^2_p < 0.001$; Figure 6), as a similar pattern was seen in both groups (Supplementary materials n. S4 for details on the analysis). Indeed, children actually paid more attention to the interpersonal space between interactors ($M = 311.95$, $SD = 483.14$) compared to the adults ($M = 281.58$, $SD = 384.21$), $F(1,150) = 4.52$, $p = .04$, $\eta^2_p = 0.03$.

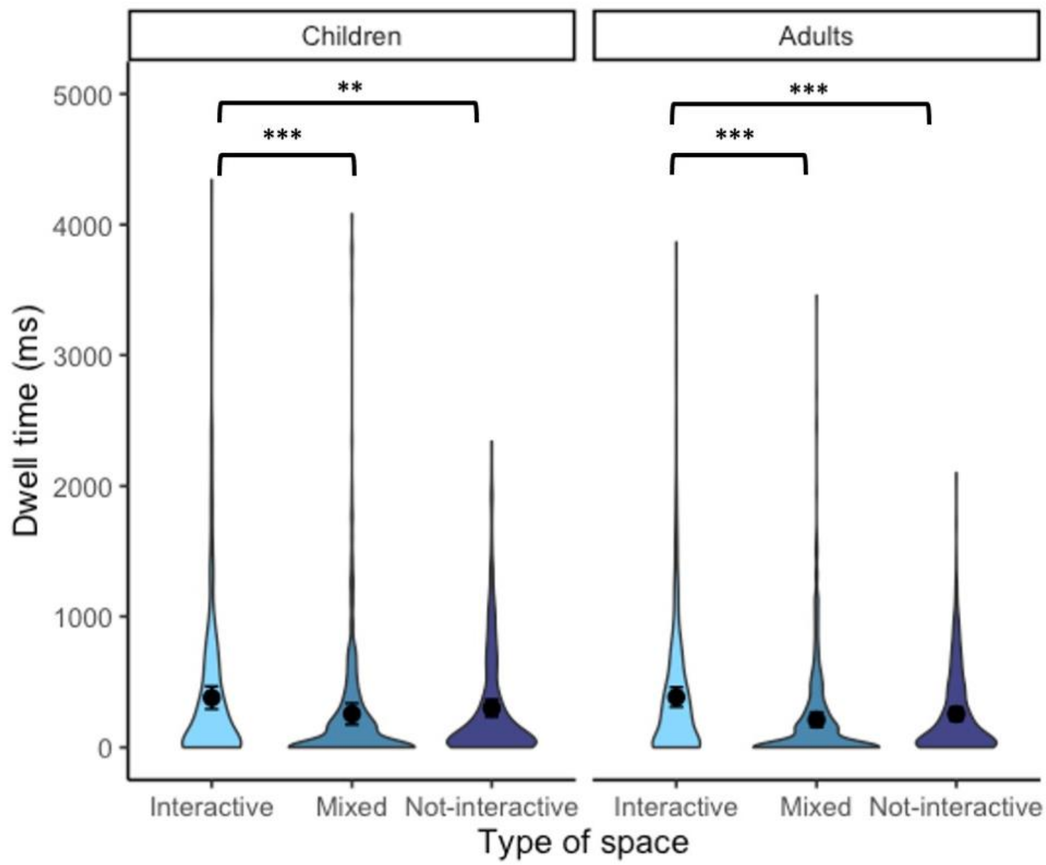

Figure 6. Violin plot for mean dwell time (ms) for interpersonal spaces in the two age-groups. Error bars represent 95% confidence intervals.

These results suggest that the developmental differences in attention to the three people pictures observed in the analysis in Figure 5 were most likely due to a lack of statistical power to support similar effects in children as were found in the adults.

Attentional capture

The time to first fixation data was transformed (logarithm in base 10) to meet multilevel modelling assumptions (see Supplementary materials n.S5 – 5a for details on transformation). Since the models after transformation showed different effects compared to the untransformed data, here we report the transformed data.

The relationship between conditions and the transformed time to first fixation showed significant variance in intercepts across participants, type of scene and area of interest (*SD* =

0.08, $\chi^2(3) = 288$, $p < .001$). After setting participant, condition, and AOI as random effects, the model included age-group as between subjects effect, and type of scene and AOI as within subjects fixed effects.

We found that children were generally slower to orient to the scenes (*M* = 2.70, *SD* = 0.31) compared to the adults (*M* = 2.66, *SD* = 0.32). In other words, there was a significant main effect of age-group ($F(1,150) = 35.4$, $p < .001$, $\eta^2_p = .19$) but the type of scene had no effect on the time to first fixation ($F(1,150) = 0.4$, $p = .54$, $\eta^2_p = .003$) with similar time taken to orient to the interactive (*M* = 2.67, *SD*= 0.31) and the non-interactive (*M* = 2.68, *SD* = 0.32) scenes. As predicted, participants showed a human bias also in attentional capture (main effect of AOI ($F(1, 300) = 326$, $p < .001$, $\eta^2_p = .52$), orienting faster to the human elements of the scene (*M* = 2.62, *SD* = 0.26) compared to the background (*M* = 2.73, *SD* = 0.36).

Age-group did not interact significantly with the type of scene ($F(1,150) = 1.6$, $p = .21$, $\eta^2_p = .01$), nor with AOI-type ($F(1,300) = 2.4$, $p = .12$, $\eta^2_p = .01$). This suggests that children were similar to adults in how they oriented to the two types of scenes and two types of AOIs (Children: interactive scenes: *M* = 2.69, *SD* = 0.31; non-interactive scenes: *M* = 2.71, *SD* = 0.32; social AOIs: *M* = 2.65, *SD* = 0.27; background AOIs: *M* = 2.75, *SD* = 0.34. Adults: interactive scenes: *M* = 2.66, *SD* = 0.32; non-interactive scenes: *M* =2.66, *SD* = 0.32; social AOIs: *M* = 2.60, *SD* = 0.25; background: *M* = 2.72, *SD* = 0.37).

There was, however, a significant interaction between type of scene and AOI, $F(1,300) = 29.7$, $p < .001$, $\eta^2_p = .09$. Indeed, again against our expectations, participants were faster to look at social information in the non-interactive scenes (*M* = 2.61, *SD* = 0.25) compared to the interactive scenes (*M* = 2.64, *SD* = 0.27), $t(150) = -2.83$, $p = .02$, $d = -0.23$, while orienting to background information was faster for interactive scenes (*M* = 2.71*,* *SD* = 0.35) compared to the non-interactive scenes (*M* = 2.75, *SD* = 0.37), $t(150) = -4.28$, $p < .001$, $d = -0.35$) although the social bias to attend to human information first was maintained in both types of scenes (interactive: $t(300) = -8.29$, $p < .001$, $d = -0.48$; non-interactive: $t(300) = -15.37$, $p < .001$, $d = -0.89$) (Figure 7; see Table S5b and Figure S5a in Supplementary materials n. S5 for untransformed data).

As with the dwell-time analysis, this effect was not moderated by age, as the three-way interaction between age-group, type of scene, and AOI did not improve the fit of the model, and was not significant ($F(1,300) = 0.3$, $p = .60$, $\eta^2_p < .001$) (Figure 7).

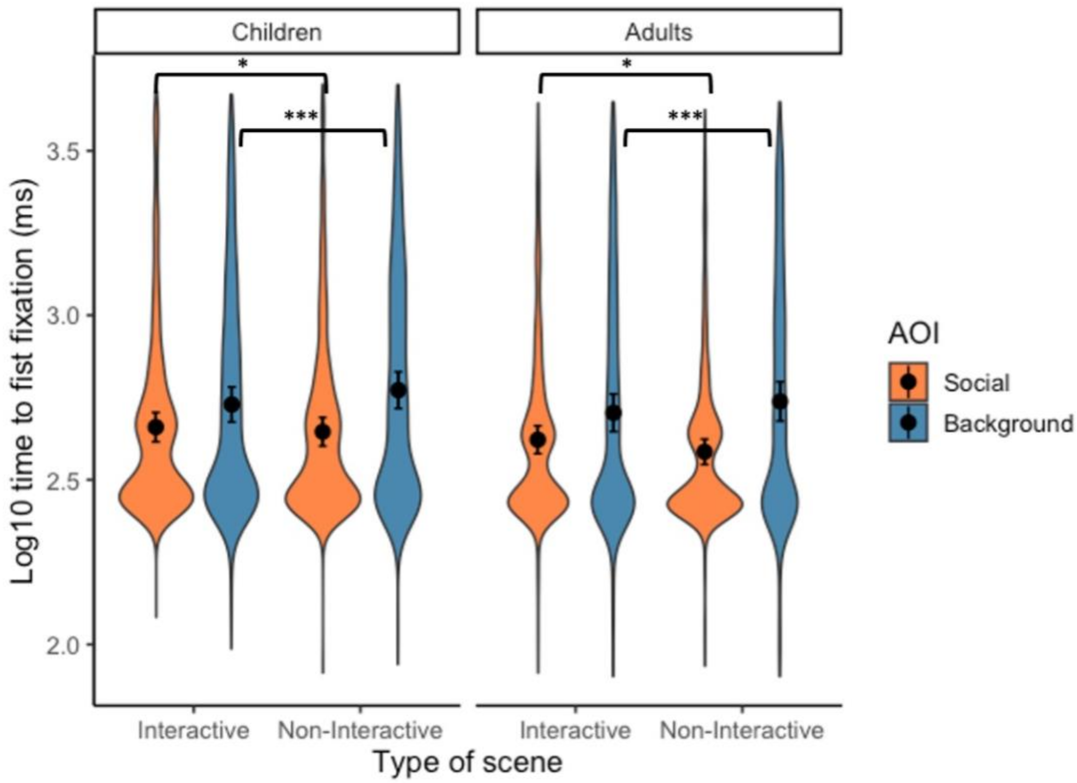

Figure 7. Violin plot for mean log transformed time to first fixation to social and non-social AOI across types of scenes in the two age-groups. All contrasts human – background are significant at p < .001. Error bars represent 95% confidence intervals.

Exploratory analyses

As for the dwelling time measure, the presence of a social interaction in the scene moderated capture by human information in the opposite direction compared to what we expected. Therefore, we conducted the same exploratory analyses that we used to look at dwell-time in order to investigate the potential cause for these results. We first performed an analysis where the number of people in the scene was added as a further fixed effect, while age-groups were merged, resulting in a model with a 2 (people in the scene) * 2 (type of scene) * 2 (AOI) structure of predictors (see Tables S2e and S2f in Supplementary Materials n. S2 – 2c for details) to investigate which group of pictures was driving this effect. The analysis revealed that the "inverse" interaction moderation was again driven by the 3-people pictures and was not seen in the 4-people pictures (Figure 8). In other words, there was a three-way interaction between number of people, type of scene and AOI ($F(1,14565) = 12.6$, $p < .001$, $\eta^2_p < .001$). In the three people pictures time taken to first fixate the non-interactive humans ($M = 2.59$, $SD = 0.25$) was significantly shorter compared to the interactive ($M = 2.63$, $SD = 0.27$), $t(151) = -3.97$, $p = .001$, $d = -0.32$, while the same was not true for the four people pictures (non-interactive: $M = 2.63$, $SD = 0.26$; interactive: $M = 2.64$, $SD = 0.27$; $t(151) = -0.99$, $p = .99$, $d = -0.08$). Additionally, in the three people pictures, participants were faster to orient to the background in the interacting scenes ($M = 2.72$, $SD = 0.35$) compared to the non-interactive scenes ($M = 2.77$, $SD = 0.38$) - $t(151) = -5.22$, $p < .001$, $d = -0.42$ – while in the four people pictures there was no difference (interactive: $M = 2.71$, $SD = 0.35$; non-interactive: $M = 2.72$, $SD = 0.35$; $t(151) = 1.17$, $p = .97$, $d = 0.10$) (Figure 8).

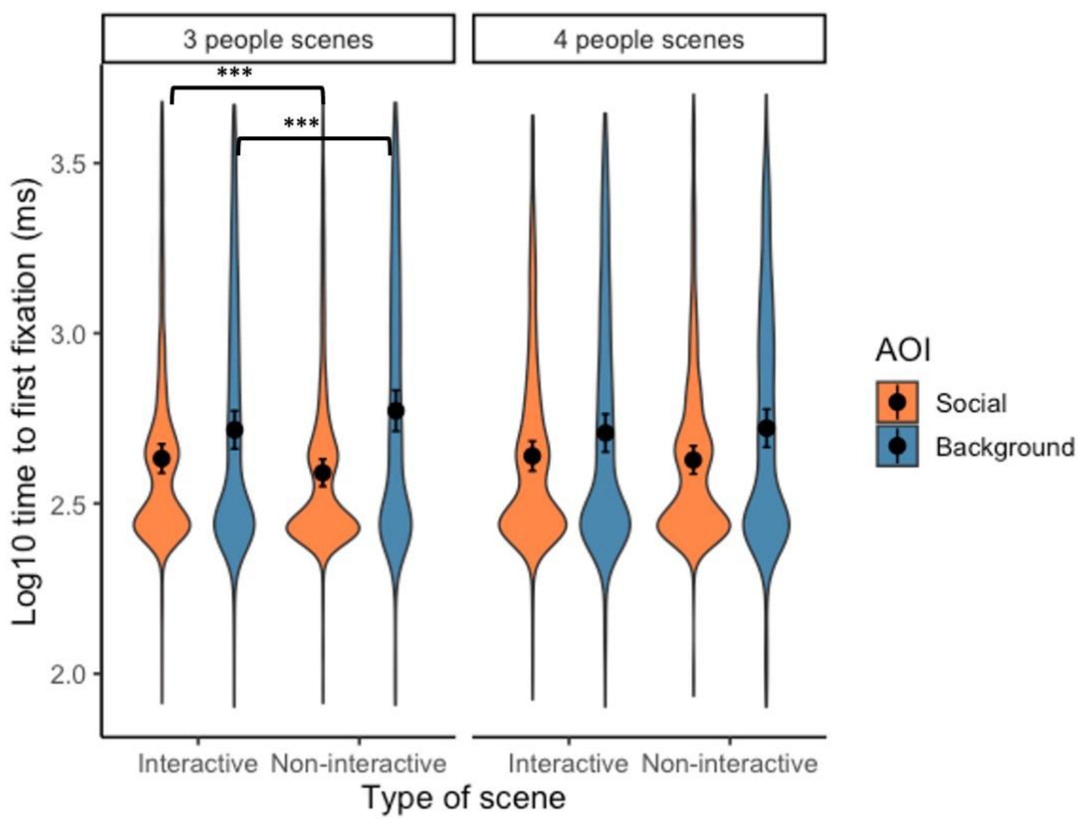

Figure 8. Violin plot for mean transformed time to first fixation for AOIs across types of scenes. All contrasts human – background are significant at $p < .001$. Error bars represent 95% confidence intervals.

To assess whether the size of the AOI might be important in driving these effects, we performed a similar hierarchical model on the three people pictures, with participant, type of scene and AOI as random effects, and type of scene, AOI and size of the AOI as fixed effects, to investigate the role of AOI size in moderating social orienting in interactive and non-interactive scenes (see Supplementary Materials n. S3 – 3c for details).

Unsurprisingly, the AOI area had an effect on the time to first fixation ($F(1,8085) = 640.3$, $p < .001$, $\eta^2{}_p = .07$) – meaning the larger the AOI, the stronger the attentional capture – although this was not moderated by the type of scene ($F(1,8085) = 3.5$ $p = .06$, $\eta^2{}_p < .001$). It was, however, moderated by AOI-type ($F(1,8085) = 21,3$, $p < 0.001$, $\eta^2{}_p = .003$) but there was no 3-way interaction between size, scene-type, and AOI-type ($F(1,8085) = 0.4$, $p = .50$, $\eta^2{}_p < .001$.

Interestingly, although the three-way interaction was not significant, ($F(1,8085) = 0.4$, $p = .50$, $\eta^2{}_p < .001$), post-hoc comparisons on the interaction between AOI-type and type of scene ($F(1,302) = 25.7$, $p < .001$, $\eta^2{}_p = .08$) showed that the time taken to fixate the human AOI in the non-interactive scenes compared to the interactive scenes was no longer significantly different ($t(151) = -1.61$, $p = .37$, $d = -0.13$), and capture by the interactive background was not different from the non-interactive ($t(151) = -1.67$, $p = .34$, $d = -0.14$). This suggests that adding the area of the AOIs to the model contributes to partially explaining the inverse effects found in the main analysis in the three-people pictures.

Interim discussion

In this first set of analyses we reconfirm a social attentional bias to human information also in more complex naturalistic displays, in both pre-adolescent children and adults. We find that children and adults are similar in the way they deploy attention to social information, although children engage more with non-social information compared to adults. Additionally, although a social interaction in the scene moderated attention to social information in the opposite direction to what we expected, exploratory analyses suggest this may be explained by holistic processing of the interacting dyads, in both age groups. Finally, we find that attentional capture in naturalistic displays can be greatly influenced by the size of regions of interest, with important future methodological implications.

Our strong social bias findings echo previous research showing a strong preference for and orienting to human information in scenes across a variety of attentional tasks using naturalistic scenarios (Birmingham et al., 2008; Doherty et al., 2017; Flechsenhar & Gamer, 2017; Sue Fletcher-Watson et al., 2008; Rösler et al., 2017). We extend this finding by showing that children aged 6 -12 years, show the same pattern of social capture and engagement as the adults, despite the clutter in the scenes. This is partially contrasting with previous research showing mild increases in social engagement across childhood (Amso et al., 2014) or even greater capture by social information in children (Doherty et al., 2019), although this difference in findings is likely the result of methodological differences between studies. Indeed, our study focuses on more global attention to social information compared to just face information as in Amso et al. (2014). As a result, we may therefore be losing some of the fine-grained attentional patterns across development specific to faces. Additionally, in our study participants were simply asked to freely view the scene. Thus our results capture spontaneous attention, while participants were performing non-social visual search task in Amso et al.'s work, which may also have measured components of executive functioning and children's ability to suppress interference from distractors (both social and not).

Interestingly, while no developmental differences were found in attention to social information, children were more engaged than adults by background information, perhaps suggesting either that they needed more contextual information to process these complex social scenes, or that they have not yet fully developed the ability to filter out irrelevant information when faced with so much visual information. Indeed, previous research shows that as the complexity of naturalistic scenes and/or social information increases (e.g., including multiple people or adding motion) children's looking behaviour changes as they use mechanisms to deal with high attentional demands, for example more off-screen time (Stoesz & Jakobson, 2014). Additionally, this difference in non-social attention could be further explained by differences in general attentional capacities between children and adults, which are not yet fully developed until late adolescence (Amso & Scerif, 2015).

Given our own prior results when investigating two-people scenes (Skripkauskaite et al., n.d.), the fact that the presence of a social interaction did not moderate social attention in these more complex scenes was surprising. Indeed, these findings contrast with prior literature suggesting the important role that social interactions play in visual attention in adults (Papeo, 2020; Papeo et al., 2017, 2019; Vestner et al., 2019) and in children (Stagg et al., 2014). Our findings do, however, match findings from the one prior study looking at interactive vs non-interactive scenes containing more than two people, though interactive content was not the focus of that particular study (Birmingham et al., 2008). One reason why we might not see a moderating effect of social interaction here is because of the complexity of the scenes and the diversity of the interactive cues that were present in our stimulus set. Although the naturalistic nature of these scenes is a strength in measuring spontaneous attention, any potential bias towards interactive scenes might obscured by the sheer variability in scene and cue-type. Additionally, we contrasted social attention in interactive and non-interactive scenes using a measure of attention to all human AOIs, regardless of whether they were interacting or not, which may dilute any potential attentional bias to interactive, but not non-interactive, individuals. In the following analyses, constrained to only interactive scenes, we investigate this possibility by looking at attentional differences between interactive and non-interactive people present *in the same scenes.*

Unexpectedly, more attention was given to non-interacting individuals compared to interacting agents, a pattern that was opposite to our hypothesis. Additional exploratory analyses suggested that this pattern of results may be the result of a holistic processing of interacting agents where more attention is dedicated to the interpersonal space between interactive, but not non-interactive, individuals. These results are in line with previous research showing a holistic processing of interacting dyads compared to two non-interactive individuals (Papeo et al., 2019; Papeo & Abassi, 2019; Walbrin & Koldewyn, 2019). However, these results must be treated with caution as they were exploratory and, especially when investigated developmentally, lack sufficient power. These exploratory results do, however, suggest the need for future studies to investigate these processes in more controlled but naturalistic scenarios. Crucially, although results in our developmental sample are underpowered, they suggest that interactive dyads may be processed holistically also in childhood, adding to a small but important literature looking at social interaction processing in childhood (Stagg et al., 2014; Walbrin et al., 2020).

Part 2: Social attentional competition of social interactions across development

Aim

In this second set of analyses, we investigate how social interactions compete for attentional resources when other social targets are in the scene, and explore developmental changes in this attentional competition. Therefore, for this second aim, only the pictures containing social interactions were considered. We first consider only data from the adult sample (n = 98), then we proceed by investigating whether there are any developmental changes across pre-adolescent childhood (n = 54) and finally consider whether and how these patterns differ between children and adults.

Data analysis

The total of 26 interactive pictures – 11 depicting 4 people and 15 depicting 3 people, were selected after the data screening procedures for part one were applied. Similarly to the first part, we used a separate hierarchical model for each of the two measures – dwell time and time to first fixation – and we analysed the data using multilevel modelling (nlme package (Pinheiro et al., 2016)). In each model, at the highest level we modelled participant information, and, within each participant the type of scene (depicting three or four people). The type of AOI (interacting or not) was then nested within each scene at the 3rd level, with the type of AOI (interacting or not-interacting human) modelled at the 2nd level, and the measure (dwell time or time-to-first-fixation) was modelled at the first level, nested within each trial and participant. Finally, pairwise comparisons were performed using post-hoc Tukey's HSD using emmeans package in R (Lenth et al., 2018).

a. Attentional competition of social interactions (adults)

Results

Attentional engagement

The relationship between conditions and dwell time showed significant variance in intercepts only across participants and number of people in the scene (*SD* = 66.90, $\chi^2(3) = 9.74$, $p = .002$).

Therefore, we set participant and number of people as random effects, and our fixed effects were the number of people in the scene and the type of human (interacting or not interacting).

We found that dwell time was significantly longer for the three people pictures (*M* = 637.67, *SD* = 474.57) compared to the four people pictures (*M* = 523.18, *SD* = 357.44); $F(1,97) = 91.08$, $p < .001$, $\eta^2{}_p = .48$. As expected, more time was spent looking at the interacting humans (*M* = 612.25, *SD* = 394.18) compared to the humans not involved in a social interaction (*M* = 566.30, *SD* = 466.93), $F(1,4884) = 15.02$, $p < .001$, $\eta^2{}_p = .003$. Finally this attentional advantage for interacting humans was not moderated by the number of people in the scene, $F(1,4884) = 0.28$, $p = .60$, $\eta^2{}_p < .001$ (Figure 9; see Supplementary materials n. S1, Table S1b for descriptive statistics), although post-hoc comparisons show that this advantage was only statistically significant in the three people pictures ($t(4884) = 3.29$, $p = .004$, $d = 0.05$) but not in the four people pictures ($t(4884) = 2.12$, $p = .11$, $d = 0.03$).

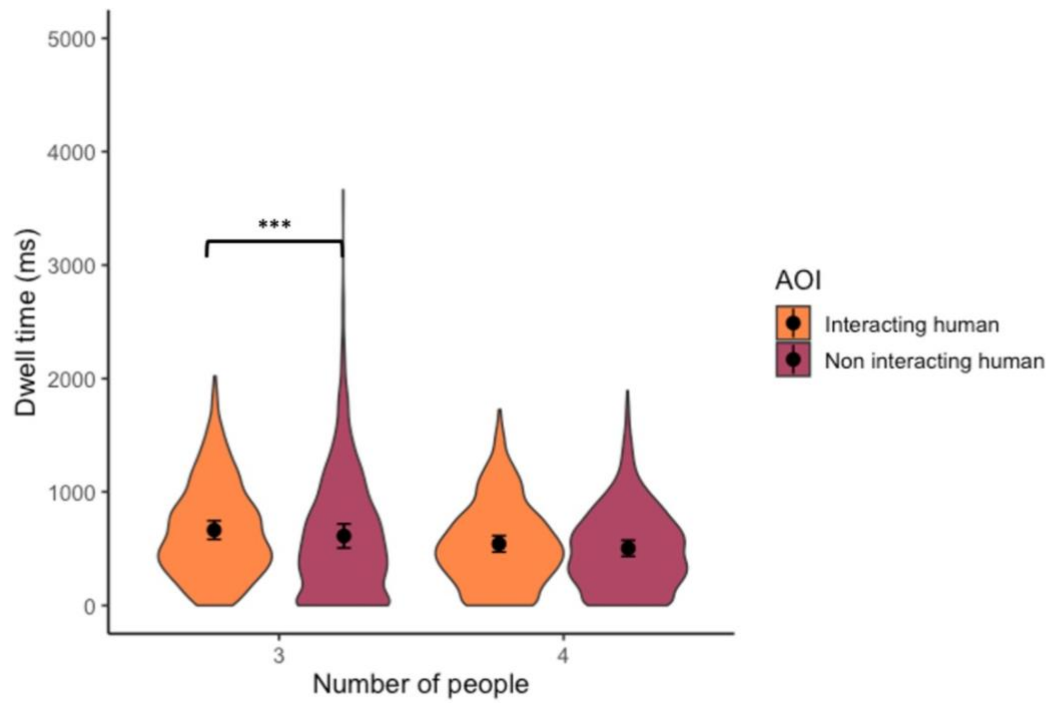

Figure 9. Violin plot for mean dwell time (ms) for interacting and non-interacting humans in three and four people pictures. Error bars represent 95% confidence intervals.

Attentional capture

Time to first fixation was transformed (square root) to meet multilevel modelling assumptions, but as the structure of model and the results did not change as a consequence of transformation, we present the untransformed data for ease of understanding (see Supplementary materials n. S5 – 5b for transformed data).

The measure of attentional capture showed variance at participant, scene and AOI level, (SD = 209.07, $\chi^2(3) = 143.96$, $p < .001$). Our fixed effects were therefore, as for the dwelling time, number of people in the scene and type of human AOI – i.e., interacting or not-interacting. We found that speed of orienting to the scenes containing 3 (M = 1012.86, SD = 973.27) or 4 people (M = 1014.61, SD = 906.44) was not significantly different ($F(1,97) = 0.004$, $p = .95$, $\eta^2_p < .001$). However, as predicted, participants were significantly faster to orient to the humans involved in a social interaction (M = 819.75, SD = 799.51) compared to not interacting humans (M = 1229.43, SD = 1043.18), $F(1,194) = 226.39$, $p < .001$, $\eta^2_p = .54$. Additionally, this effect was different for the two types of pictures, $F(1, 194) = 35.45$, $p < .001$, $\eta^2_p = .15$. In particular, the difference in capture by interacting (M = 761.44, SD = 772.39) versus non-interacting humans (M = 1312.54, SD = 1095.81), was stronger for the three-people pictures, $t(194) = -15.27$, $p < .001$, $d = -1.10$, while it was milder for the four-people pictures (interacting human: M = 903.94, SD = 830.32; non-interacting human: M = 1127.11, SD = 965.27), $t(194) = -5.36$, $p < .001$, $d = -0.39$) (Figure 10).

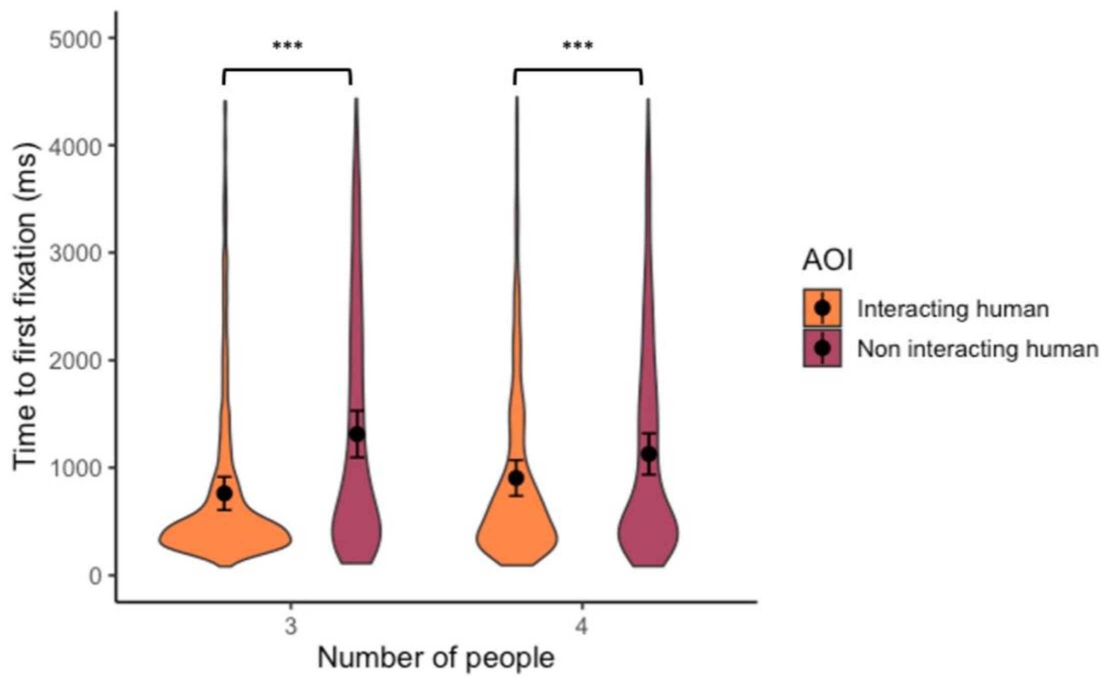

Figure 10. Violin plot for mean time to first fixation (ms) for interacting and non-interacting humans in three and four people pictures. Error bars represent 95% confidence intervals.

b. Attentional competition of social interactions in pre-adolescent childhood

In this section, for both measures, age was modelled as a continuous predictor, while number of people in the scene and type of human were the between-subjects predictors.

Results

Attentional engagement

Dwelling time data was transformed (square root) to meet multilevel modelling assumptions (see Supplementary materials n. S5 – 5c for details), but since the results did not change after the transformation, we present here the untransformed data.

The relationship between conditions and dwell time showed significant variance in intercepts across participants, number of people in the scene and type of human AOI (*SD* = 82.37, $\chi^2(3) = 4.36$, $p = .04$).

In the model that included age, number of people in the scene and type of human (interacting or not interacting) we found that age had no effect on the dwelling time to these scenes, $F(1,52) = 0.00$, $p = .99$, ${\eta^2}_p = .00$, and did not interact significantly with the type of scene ($F(1,52) = 0.05$, $p = .83$, ${\eta^2}_p < .001$) or with the type of AOI ($F(1,104) = 0.01$, $p = .94$, ${\eta^2}_p <$ .001.

As was true for adults, children's dwell time was significantly longer for the three people pictures (*M* = 616.63, *SD* = 603.21) compared to the four people pictures (*M* = 498.24, *SD* = 453.05), a main effect of number of people $F(1,52) = 32.37$, $p < .001$, ${\eta^2}_p = .38$. Similarly, more time was spent looking at the interacting humans (*M* = 630.45, *SD* = 513.73) compared to the humans not involved in a social interaction (*M* = 501.93, *SD* = 572.08), main effect of AOI, $F(1,104) = 39.07$, $p < .001$, ${\eta^2}_p = .27$. Unlike for adults, this effect was moderated by the number of people in the scene $F(1,104) = 8.12$, $p = .01$, ${\eta^2}_p = .07$ (Figure 11). Indeed, only the three people pictures showed a significant difference in dwell time between interacting humans (*M* = 706.13, *SD* = 532.87) and non-interactors (*M* = 527.12, *SD* = 654.38), $t(104) = 6.60$, $p < .001$, $d = 0.65$. While children attended more to interactors than non-interactors *numerically* in the 4-people scenes, this difference was not significant (interacting: *M* = 528.50, *SD* = 468.23; non-interacting: *M* = 467.98, *SD* = 435.65; $t(104) = 1.93$, $p = .21$, $d = 0.19$).

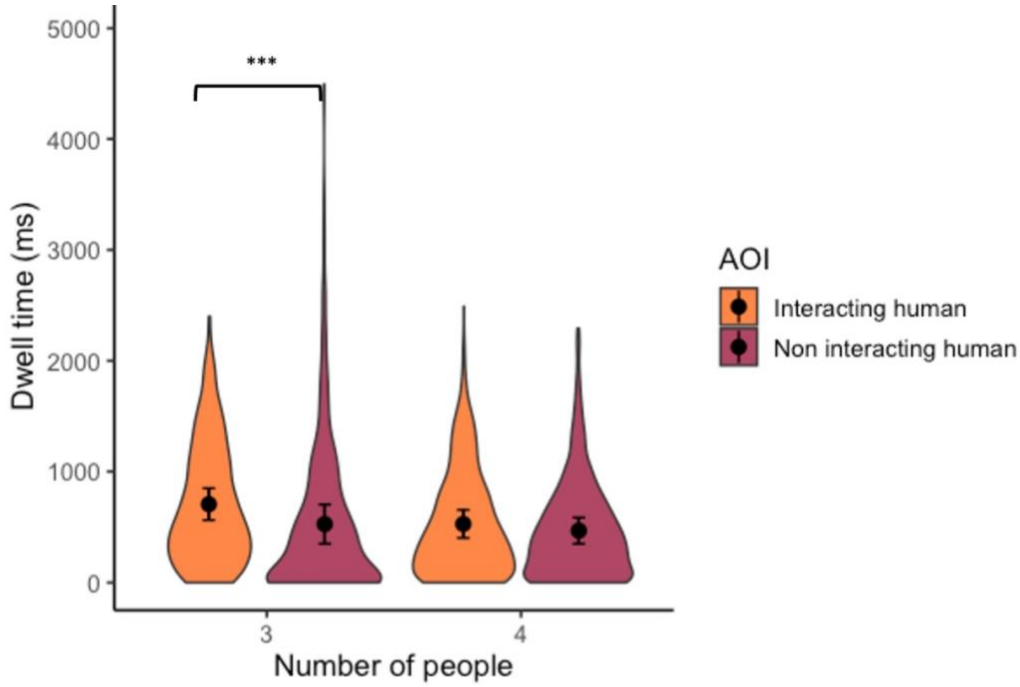

Figure 11. Violin plot for mean dwell time (ms) for interacting and non-interacting humans in three and four people pictures. Error bars represent 95% confidence intervals.

Interestingly, age also further moderated this relationship, $F(1,104) = 7.64$, $p = .01$, $\eta^2_p = .07$ (Figure 12) with a constant advantage to the interacting humans in both types of pictures until 11 years of age.

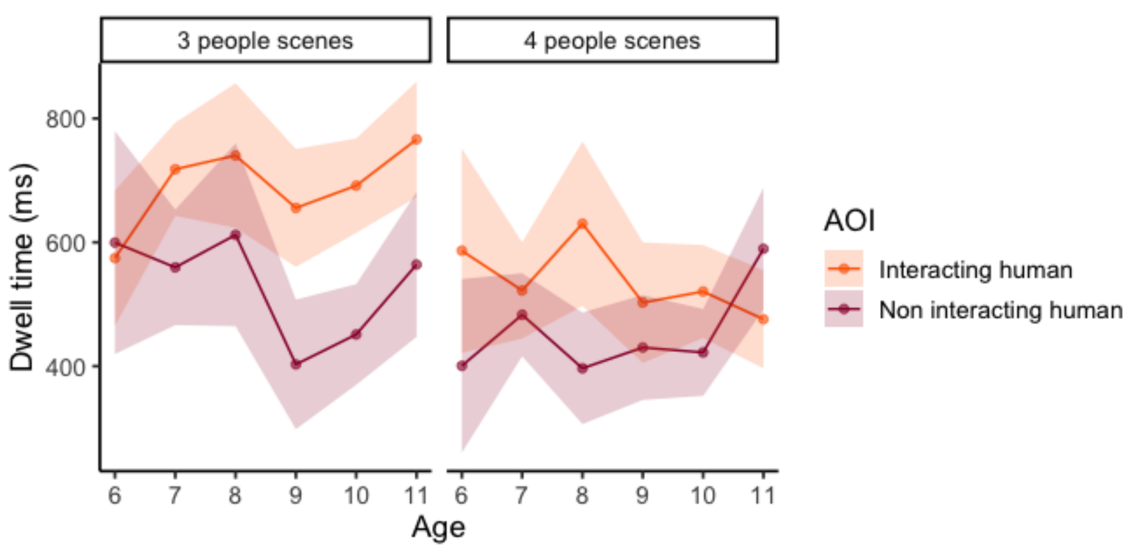

Figure 12. Mean dwell time (ms) for interacting and non-interacting humans in three and four people pictures across childhood years. Width of bands represent 95% confidence intervals.

Attentional capture

Time to first fixation was transformed (logarithm in base 10) to meet multilevel modelling assumptions, but as the model and the results did not change as a consequence of the transformation, we present the untransformed data (see Supplementary materials n. S5 – 5d for transformed data).

The measure of attentional capture showed variance at participant, scene, and AOI level, (*SD* = 177.93, $\chi^2(3) = 8.72$, *p* = .003). The model included age as a continuous predictor and number of people in the scene and type of human AOI – i.e., interacting or not-interacting – as categorical predictors.

Developmental increase in the speed of orienting to the scenes reached a trend level – age main effect ($F(1,52) = 3.67$, *p* = .06, $\eta^2{}_p$ = .07). Additionally age did not interact significantly with the type of scene (not significant interaction $F(1,52) = 0.20$, *p* = .66, $\eta^2{}_p$= .004) or with the type of AOI ($F(1,104) = 0.28$, *p* = .60, $\eta^2{}_p$= .003).

Participants were equally fast to orient to the pictures depicting 3 (*M* = 1034.77, *SD* = 1032.84) or 4 people (*M* = 1068, *SD* = 1020.43), ($F(1,52) = 0.66$, *p* = .42, $\eta^2{}_p$ = .01), but within the scenes, as expected, children were on average faster to look at the interacting humans (*M* = 882.45, *SD* = 911.41) than the non-interactors (*M* = 1251.97, *SD* = 1120.01), $F(1,104) = 73.57$, $p < .001$, $\eta^2{}_p$ = .41. Finally, this effect was different for the two types of pictures, $F(1, 104)$ = 23.12, $p < .001$, $\eta^2{}_p$ = .18. In particular, the difference in capture by interacting (*M* = 801.71, *SD* = 834.77) versus non-interacting humans (*M* = 1355.23, *SD* = 1183.20), was seen only in the three-people pictures, $t(104) = -9.59$, $p < .001$, *d* = -0.94, while there was no significant difference in the four-people pictures (interacting human: *M* = 1000.15, *SD* = 1002.07; non-interacting human: *M* = 1138.38, *SD* = 1035.49), $t(104) = -2.14$, *p* = .13, *d* = -0.21) (Figure 13).

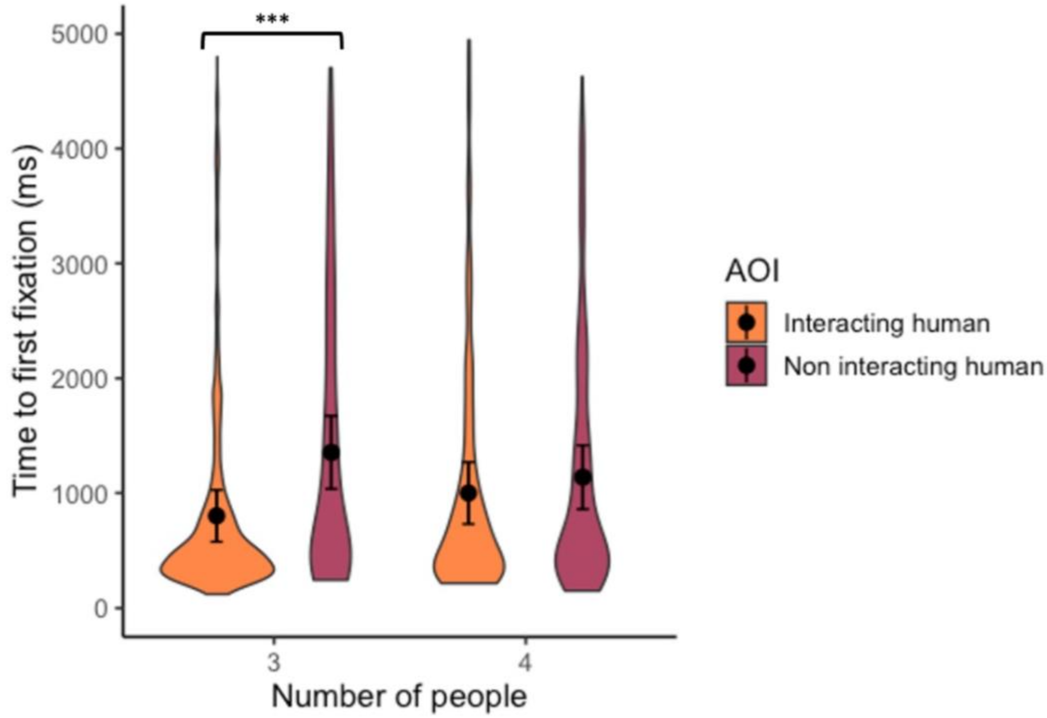

Figure 13. Violin plot for mean time to first fixation (ms) for interacting and non-interacting humans in three and four people pictures. Error bars represent 95% confidence intervals.

The three way interaction between age, number of people in the scene, and type of human did not improve the model fit and was not significant, $F(1,104) = 0.001$, $p = .98$, $\eta^2{}_p < .001$ (Figure 14), suggesting this advantage in capture by interacting humans was constant across childhood.

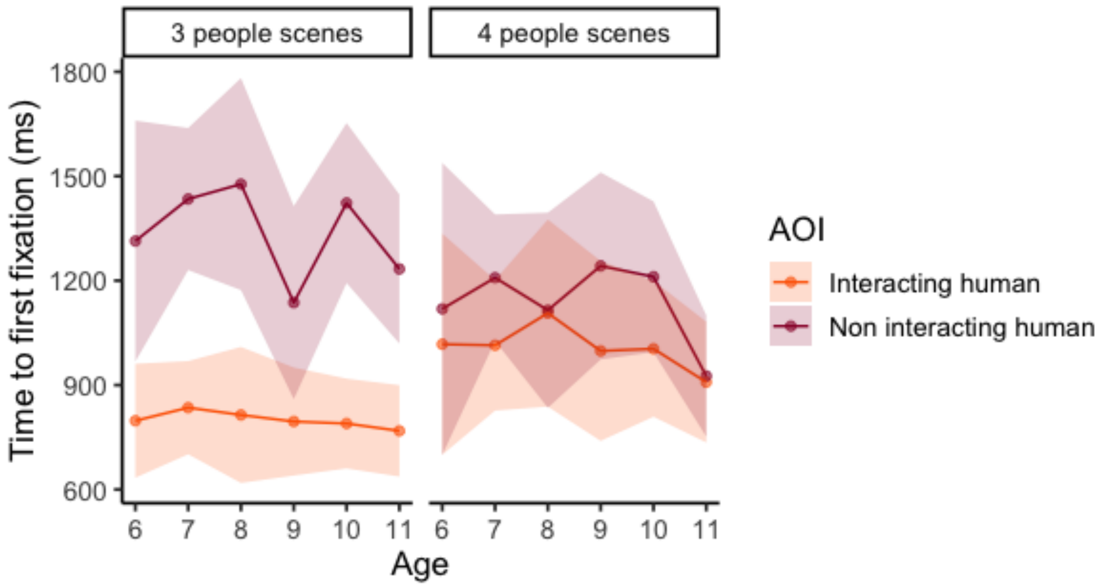

Figure 14. Mean time to first fixation (ms) for interacting and non-interacting humans in three and four people pictures across childhood years. Width of bands represent 95% confidence intervals.

c. **Developmental changes in the attentional competition of social interactions**

In this section, for both measures, age was modelled as a categorical predictor – children vs adults – while number of people in the scene and type of human were the between subjects predictors.

Results

Attentional engagement

Dwell time data was transformed (square root) to meet multilevel modelling assumptions (see Supplementary materials n. S5 – 5e for details). Dwelling time showed variance at participant, scene and AOI level, (SD = 1.65, $\chi^2(3) = 16.01$, $p < .001$).

General attentional engagement with the scenes was significantly shorter for children (M = 20.07, SD = 12.76) compared to the adults (M = 21.91, SD = 10.44), ($F(1,150) = 24.20$, $p < .001$, ${\eta^2}_p = .14$), and this was true for both types of scenes, $F(1,150) = 0.18$, $p = .67$, ${\eta^2}_p = .001$.

Additionally more attention was given to the pictures depicting 3 people (M = 22.04, SD = 12.03) compared to the four people pictures (M = 20.24, SD = 10.23), $F(1,150) = 50.33$, $p < .001$, $\eta^2{}_p = .25$, and this was similar for the two groups.

Unsurprisingly, in general more attention was given to the humans included in a social interaction (M = 22.73, SD = 10.09) compared to the independent humans (M = 19.82, SD = 12.29), $F(1,7492) = 135.79$, $p < .001$, $\eta^2{}_p = .02$. This was moderated by age, $F(1,7492) = 18.16$, $p < .001$, $\eta^2{}_p = .002$. Unpacking this interaction revealed that although both groups show a strong attentional priority to interacting humans (children: interacting human: M = 22.27, SD = 11.60; non-interacting human: M = 17.91, SD = 0.31; $t(7492) = 9.33$, $p < .001$, $d = 0.11$; adults - interacting human: M = 22.98, SD = 9.18; non-interacting human: M = 20.85, SD = 11.47; $t(7492) = 6.33$, $p < .001$, $d = 0.07$)), adults actually attended *more* to non-interacting humans than children did, $t(150) = 6.13$, $p < .001$, $d = 0.50$. There was, however, no significant difference between groups attention to interacting humans ($t(150) = 1.75$, $p = .29$, $d = 0.14$) (Figure 15).

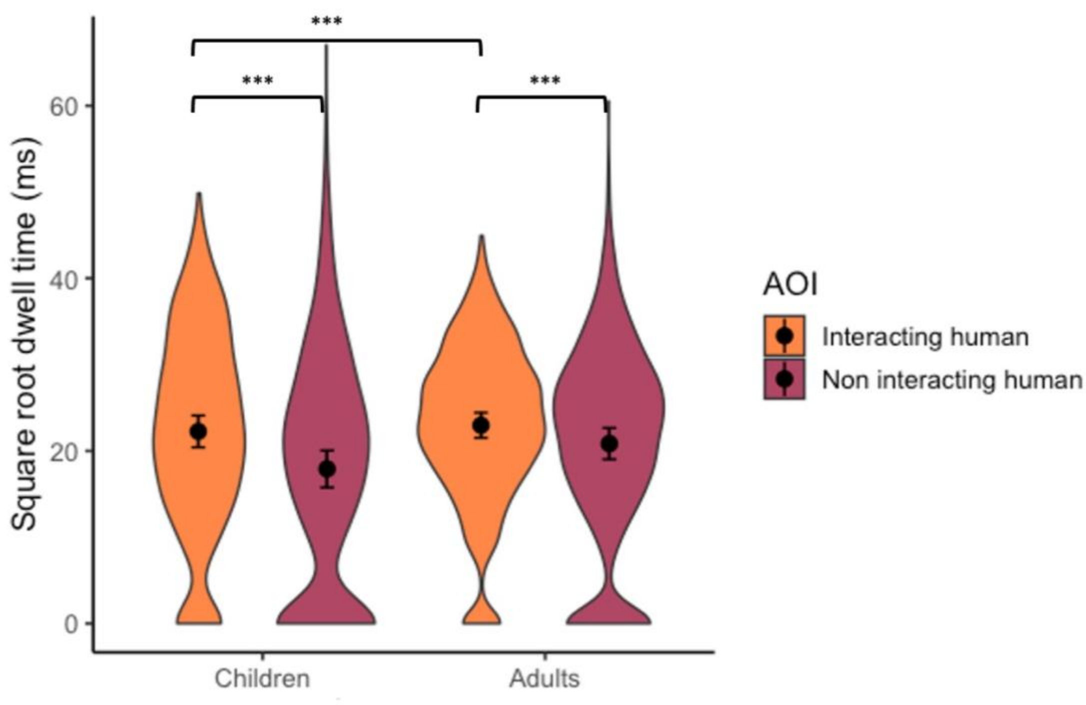

Figure 15. Violin plot for transformed mean dwell time (ms) in the two age-groups, for interacting and non-interacting humans. Error bars represent 95% confidence intervals.

This attentional advantage was also different depending on the number of people in the scene, $F(1,7492) = 36.26$, $p < .001$, $\eta^2_p = .01$, with less attention to the interacting humans in the four people pictures ($M = 20.82$, $SD = 10.20$) compared to the three-people pictures ($M = 24.14$, $SD = 9.78$), $t(150) = -9.38$, $p < .001$, $d = -0.77$) but no difference in the attention given to the non-interacting humans (3 people pictures: $M = 19.94$, $SD = 13.60$; 4 people pictures: $M = 19.66$, $SD = 10.23$), $t(150) = -0.004$, $p = 1.0$, $d < 0.001$). However, while the difference between attention to interactive and non-interactive humans is smaller in the 4-person scenes, the pattern is similar (i.e., more attention to interactors and non-interactors) and when collapsing across children and adults, there is a significant difference for both 3-people ($t(7492) = 13.73$, $p < .001$, $d = 0.04$), and 4-people scenes, $t(7492) = 3.05$, $p = .01$, $d = 0.04$.

However, this effect was different between age-groups, $F(1,7492) = 36.26$, $p < .001$, $\eta^2_p = .001$ (Figure 16; see Supplementary materials n. S5 – 5e for descriptive statistics of transformed and untransformed data). Indeed while there are no developmental differences in the attention given to the interactors in either three person ($t(150) = 0.24$, $p = 1.0$, $d = 0.02$) or 4-person scenes ($t(150) = 2.35$, $p = .16$, $d = 0.19$), children pay less attention than adults to non-interactors in the 3-person scenes ($t(150) = 6.67$, $p < .001$, $d = 0.54$), an effect that is smaller in the 4-person scenes ($t(150) = 3.05$, $p = .03$, $d = 0.25$). Additionally, when looking at the two age groups separately, neither children ($t(7492) = 2.24$, $p = .19$, $d = 0.03$) nor adults ($t(7492) = 2.11$, $p = .24$, $d = 0.02$) show significant attentional priority in the 4 people pictures, while both do in the three people pictures (children: $t(7492) = 11.69$, $p < .001$, $d = 0.14$; adults: $t(7492) = 7.26$, $p < .001$, $d = 0.08$).

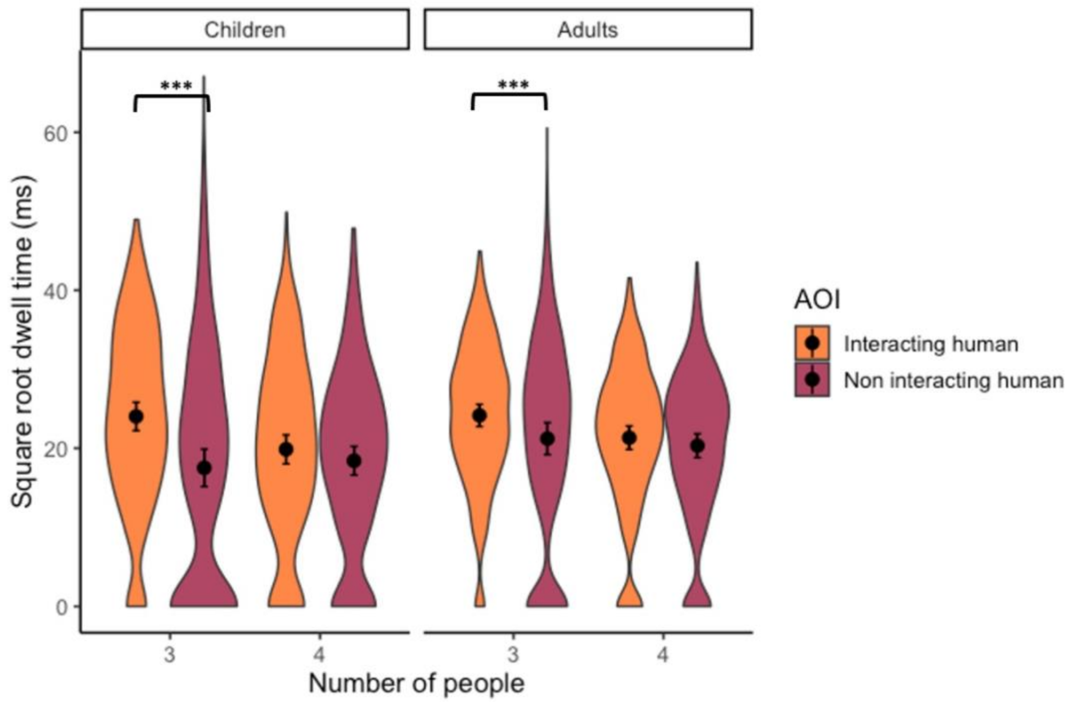

Figure 16. Violin plot for transformed mean dwell time (ms) in the two age-groups, for interacting and non-interacting humans in three and four people pictures. Error bars represent 95% confidence intervals.

Attentional capture

Time to first fixation was transformed (logarithm in base 10) to meet multilevel modelling assumptions, but as the model and the results did not change as a consequence of the transformation, we present the untransformed data (see Supplementary materials S5 – 5f for details on transformation and model details with transformed data). The measure of attentional capture showed variance at participant, scene and AOI level, (*SD* = 201.1, $\chi^2(3) = 54.72$, $p <$ .001).

After setting participant, condition and AOI as random effects, the fixed effects were the same as in the dwell-time analysis and included number of people in the scene and type of human AOI – i.e., interacting or not-interacting.

General speed of orienting to the scenes was statistically similar for children (*M* = 1049.33, *SD* = 1027.32) and adults (*M* = 1013.61, *SD* = 945.17), ($F(1,150) = 2.13$, $p = .15$, η^2_p = .01) and the type of scene had no effect on the time to first fixation – $F(1,150) = 0.27$, $p = .60$, $\eta^2_p = .002$ – with similar time taken to orient to the scenes depicting 3 (*M* = 1019.96, *SD*=

992.89) and 4 (M = 1032.42, SD = 946.15) people, which was similar for the two groups, $F(1,150) = 0.16$, $p = .69$, $\eta^2_p = .001$.

Both groups showed an advantage for interacting individuals in attentional capture (M = 840.88, SD = 839.30) compared to the non-interacting individuals (M = 1236.60, SD = 1068.08), $F(1, 300) = 295.37$, $p < .001$, $\eta^2_p = .50$, and this was not moderated by age, $F(1, 300) = 0.70$, $p = .40$, $\eta^2_p = .002$. This effect did, however, vary with the number of people in the scene, $F(1, 300) = 58.54$, $p < .001$, $\eta^2_p = .16$) where the difference in time-to-first-fixation between interacting humans and non-interactors was larger for the three people pictures (interacting humans: M = 775.05, SD = 794.06; non-interacting humans: M = 1325.64, SD = 1123.17; $t(300) = -16.78$, $p < .001$, d = - 0.97) than for the four people pictures (interacting humans: M = 936.25, SD = 892.51; non-interacting humans: M = 1130.84, SD = 988.74; $t(300) = -4.86$, $p < .001$, d = - 0.28) (Figure 17).

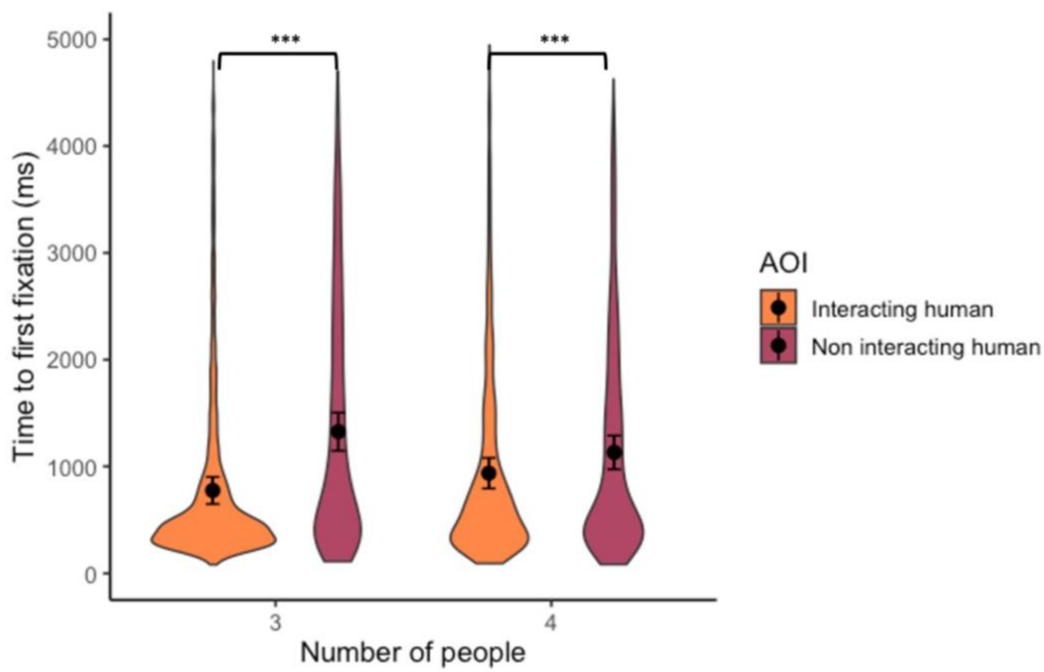

Figure 17. Violin plot for untransformed mean time to first fixation (ms) for interacting and non-interacting humans in three and four people pictures. Error bars represent 95% confidence intervals.

Finally, the three-way interaction between age, number of people in the scene and type of human did not improve the model fit and was not significant, $F(1,300) = 0.77$, $p = .38$, η^2_p =

.003 (Figure 18; see Supplementary materials n. S1 – 1c for descriptive statistics). Despite this, planned post-hoc comparisons show that while both groups show an attentional advantage for interacting individuals in the three people pictures (children: $t(300) = -10.2$, $p < .001$, $d = -0.59$; adults: $t(300) = -14.79$, $p < .001$, $d = -0.85$), only adults showed this effect in the four people pictures as well (children: $t(300) = -2.28$, $p = .18$, $d = -0.13$; adults: $t(300) = -5.2$, $p < .001$, $d = -0.30$).

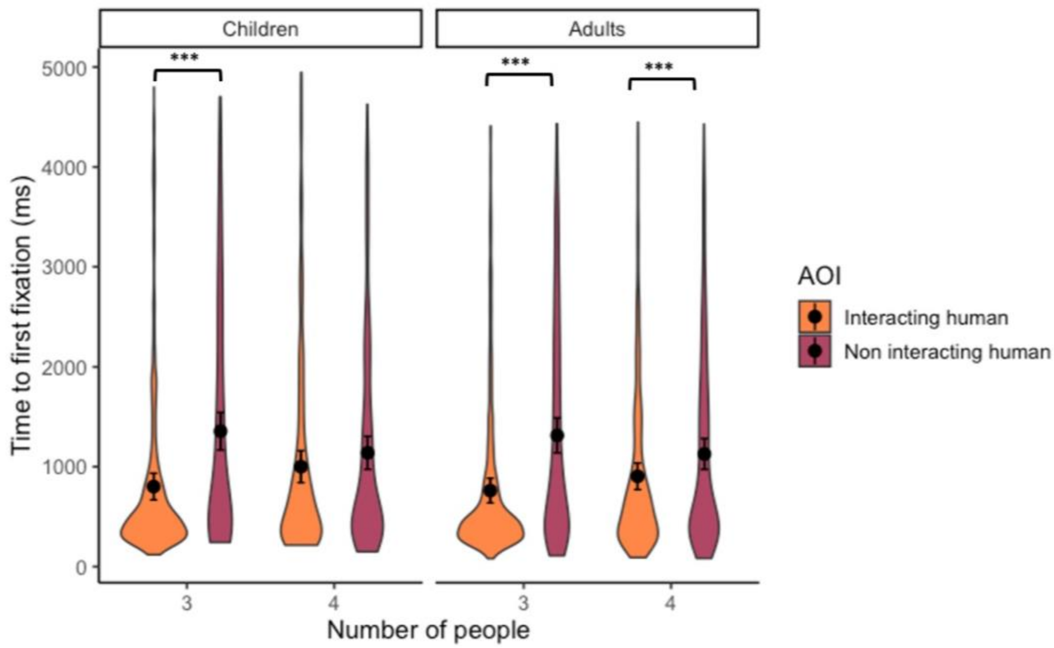

Figure 18. Violin plot for untransformed mean time to first fixation (ms) in the two age-groups, for interacting and non-interacting humans in three and four people pictures. Error bars represent 95% confidence intervals.

Exploratory analyses

However, for consistency, given that in part 1 we also considered the size of the AOIs across conditions to verify our effects, we decided to do the same here (see Supplementary materials n. S3 – 3d for details on the size of the human AOIs in the 3 and 4 people pictures).

Since in the three people pictures the interacting humans seem to show much variability and potentially drive the effects of attentional priority seen in part 2, we decided to re-run the developmental analyses for dwelling time (Supplementary materials S2 – 2d) and time to first fixation (Supplementary materials S2 – 2e) without 3 of the pictures depicting three people where the interacting humans were much greater in average size than the non-interacting human. These analyses show that nothing changes in the effects, lending strength to the idea that this attentional priority did not depend on the size of the interacting AOIs compared to the non-interacting.

General discussion and conclusions

In this work we reconfirm a strong attentional bias to human information in complex scenes in adults and children, but find little indication of developmental changes to social information, nor age-related differences in the influence of social interactions when comparing scenes containing social interactions with scene that contain only independent agents. Crucially, when social interactions have to compete for attentional resources with other social target in the same scene, they capture attention first and engage it for longer in both groups, although this effect is smaller in children and is reduced as the number of social targets increases. Exploratory analyses suggest some evidence that interactions might be perceived/processed as a single unit/gestalt, a mechanism that may not be fully developed in childhood.

This research first of all extends the literature on social orienting literature (Birmingham et al., 2008; Doherty et al., 2017, 2019; Sue Fletcher-Watson et al., 2008; Rösler et al., 2017; Stagg et al., 2014) by showing that pre-adolescent children and adults are extremely similar in the way they orient to and engage with social information, even in complex naturalistic scenes with multiple social targets. This work also largely supports prior work showing an attentional advantage for interacting dyads (Papeo et al., 2019; Skripkauskaite et al., n.d.; Stagg et al., 2014; Vestner et al., 2020), although this advantage is more clearly demonstrated in multi-person scenes when interactive information is in direct competition is other social information. Indeed, both groups manifest an apparent lack of sensitivity to the presence of a social interaction when scenes that contain a social interaction are compared to scenes depicting only independent agents, congruent with some prior findings (Birmingham et al., 2008). Follow-up exploratory analyses, however, suggest that this insensitivity to interactive content might be explained at least partially by holistic processing of interacting dyads, where interacting agents are processed and scanned together, while independent humans must be attended to, and processed, separately. Although these exploratory results must be interpreted cautiously and will need to be replicated, they are coherent with prior studies showing that interactions are processed holistically and as 'more than the sum of their parts' (Ding et al., 2017; Papeo, 2020; Walbrin & Koldewyn, 2019). Importantly, exploratory analyses suggest similar processes for both children and adults, although these results are underpowered.

As for attentional engagement, it would seem that social interactions' ability to hold attention weakens with the addition of other targets in the scene (i.e. pattern seen in the 4 people pictures). This attentional priority is in line with the stream of research showing attentional advantages of social interactions (Papeo et al., 2017, 2019; Vestner et al., 2019; Villani et al., 2015). Moreover, we add novel findings by showing a similar pattern in pre-adolescent childhood. Indeed, also 6-12 years old children give attentional priority to social interactions in a constant trend across childhood, by showing capture and engagement of attention when one other socially relevant target is in the scene. When the competitors are two, it would seem that children's attentional capture is not sensitive enough to show this interaction advantage as in adults. There has been very little research on the developmental changes of processing of social interactions (Augusti et al., 2010; Hamlin et al., 2007; Handl et al., 2013; Stagg et al., 2014; Walbrin et al., 2020) and our findings extend greatly these investigations. Additionally, Stagg et al. (2014) found a preference of children of 9 years of age in looking at two facing figures when presented together with two non-facing figures, while we don't find such effect in the four people pictures. This might be because our stimuli are much more complex and naturalistic, suggesting that clutter in the scene makes this interaction bias harder to break through attentionally in childhood. Future studies should investigate the role of social interactions for attention when visual perception resources are reduced. Interestingly, some research might suggest this effect could be driven by an attentional "hot-spot" facing dyads create (Vestner et al., 2020) but our naturalistic displays are too various and heterogeneous in the structure of the interactions – i.e. often the interacting agents don't even face each other or are also far away from each other – to drive this attentional priority effect.

It is surprising, and contrary to our original hypotheses, that we find so little difference between children and adults in how they allocate spontaneous attention to these complex multi-person scenes. Indeed, we know from prior work that social attention is not fully 'adult-like' in pre-adolescent childhood (Amso & Scerif, 2015) and that regions of the 'social brain' that support the perception of social interactions are not yet fully tuned (Mills et al., 2014; Sapey-

Triomphe et al., 2017; Walbrin et al., 2020). Social interactions seem to override these early developmental features, and are given priority even in childhood, manifested by capturing and holding attention even when the scene is complex and cluttered. This suggests a hierarchy of social information, where a single individual has attentional priority over objects and background elements but less priority than a social interaction between two individuals. That attention to social interactions is already 'adult-like' in middle childhood lends strength again to the idea that social interactions play a fundamental role early in life, even when attentional tools are not fully developed. Indeed social learning, predicting others' behaviours and decisions, impression formation, emotional recognition and social decisions all rely heavily on the ability to orient to and attentionally select social cues, and perhaps particularly *interactive* cues, from noisy environment (Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017; Quadflieg & Westmoreland, 2019; Skinner et al., 2017). It is important, however, to note that our task here captures spontaneous attention rather than goal or task-directed attention. Differences between children and adults in social attention to social interactions may emerge in the context of attentional tasks or when attentional resources are taxed.

We did see developmental differences between children and adults in two areas. First, children appeared to pay more attention to background information compared to the adults, potentially reflecting either the need for more contextual information to process complex scenes, or a less developed executive control system (Amso & Scerif, 2015; Federico et al., 2017). It could, however, also reflect children's natural viewing behaviour (Açık et al., 2010), where children are more likely to explore *all* of a scene than are adults. The overall slower orientation to scenes in children and the greater attention to background compared to adults may suggest weaker attentional control. Similarly, Stoesz et al. (2014) showed that when multiple elements were added to the scene or motion was included, children showed a pattern of disengaging attention from relevant but potentially too complex elements of the scene, by for example looking off screen more often for more complex stimuli. Additionally, later stages of social attention might be influenced by factors like 'top-down' social knowledge about the 'usual' configuration of social scenarios, which are likely not fully developed in childhood. This might indicate that more attentional resources need to develop to be able to cope with all the social information in the four people pictures. Future studies could investigate the role of social knowledge in parsing complex social scenes, as well as how the influence of such knowledge on social attentional orienting changes across development.

There are a few limits to the current research that need consideration. In the first place our developmental sample would benefit from a greater size – for example completing the sample originally planned for in our a priori power analyses. Therefore, the developmental findings should be interpreted with caution. Additionally, the analyses investigating attentional competition in the 4-person scenes is relies on eye-gaze data form only 11 scenes. Thus, the different findings from 3-person and 4-person scenes will require replication. Future investigations into mechanisms of social attention to social interactions in such competitive situations should consider increasing the number of stimuli, especially if using a diversity of contexts and environments, as we have here. Finally, future studies should consider more carefully the relative size of AOIs when designing experiments looking at attentional capture by social information, especially when using naturalistic complex scenarios.

Chapter 5. Development of attention to ambiguous social scenes and the role of social knowledge

Abstract

Human attention supports exploration of complex and noisy environments through facilitating the processing of important elements and filtering out distractors. Starting in the very first hours of life, humans display an attentional preference for social information. The attentional prioritisation of social information has been shown to be automatic in nature, but can be moderated by top-down knowledge that contributes to the interpretation and prioritization of what is being observed. The influence of top-down knowledge is especially apparent when scenes are noisy and ambiguous. Despite this, we don't yet know how social attention to compex scenes is influenced by ambiguous information during childhood. One source of top-down social knowledge for parsing complex scenes is people's understanding of the usual 'characteristics' that identify when individuals are engaged in a social interaction. Understanding the influence of such knowledge is important because not only are social interactions a ubiquitous feature of social scenes but prior research suggests an attentional preference for social interactions that is similar for children and adults. However, research has not yet looked into how the interplay between social knowledge and the perception of social interactions can influence attentional orienting. In this study we recorded the eye movements of 73 adults and 54 children in a free viewing experiment. Ambiguous naturalistic scenes depicted a dyad that could either be interpreted as interacting or not. After the eye-tracking session, participants indicated whether they had perceived each scene as interactive or not. Here, we aimed to investigate developmental changes in the way top-down information about social events (e.g., what a social interaction usually looks like) might influence social attention. Areas of interest (AOIs) were divided between "social" (entire human figures in the scene) and "non-social" (all other elements). We show that adults show a stronger bias to interpret ambiguous scenes as interactive compared to children, but that this categorisation had no influence on attention to social information in the scenes. We show a general attentional social bias in engagement and capture in both groups, across scenes but whether a scene was seen as interactive or not did not change how attention

was allocated for either children or adults. Implications for social attention and the development of social knowledge are discussed.

Introduction

As social animals, human beings are constantly immersed in complex social information, and must learn to make decisions and predictions based on social cues. To do so, we must be able to quickly and readily detect, extract, and understand the most relevant social cues, even when they are embedded in complex scenes. We have thus developed cognitive systems since early in life that are well tuned for such processes and their reflexive – i.e. automatic – nature is well established (e.g. Flechsenhar & Gamer, 2017; Rösler et al., 2017). Additionally, these automatic processes that select and process social information are in constant communication with ‘higher-order’ processes that inform our social beliefs, knowledge and behaviour. Our knowledge and beliefs about people and social contexts then feeds back to guide our perceptual and voluntary attentional systems (Collins & Olson, 2014; Dolan et al., 1997; Wiese et al., 2012). In the real world, the information surrounding us is most often noisy and ambiguous so that we often need to use our knowledge to disambiguate social situations. One excellent source of knowledge about complex social scenes that can guide our understanding are the social interactions that we have observed across our lives. Indeed, social interactions seem to carry unique value for social learning across development, informing action prediction, social decision making, and more general social scene understanding (Quadflieg & Westmoreland, 2019). At the same time, recent evidence suggests that our attention is captured and held by social interactions (Papeo et al., 2019; Skripkauskaite et al., n.d.; Stagg et al., 2014; Vestner et al., 2019). Additionally, as seen in chapter 3 of this work, this effect holds strongly also in complex naturalistic scenes. Yet, despite these attentional effects and the demonstrated value of social interactions in building social context knowledge (Quadflieg & Westmoreland, 2019), most of what we know about the interplay between reflexive attentional mechanisms and higher order knowledge in social attention comes from depictions of single agents and mostly through the study of adult populations.

Much prior research has shown that human attention is strongly tuned to social information both when individuals are presented individually and in naturalistic, cluttered scenes. Adult viewers show a strong social attentional bias as indicated by more time spent looking at faces, bodies, and whole human figures in a variety of attentional tasks (Bindemann et al., 2010; Birmingham et al., 2009a; Doherty et al., 2017; Sue Fletcher-Watson et al., 2008). Findings from developmental work are heterogeneous, but despite mostly focusing on infants (Bertenthal & Boyer, 2015), developmental work generally supports the idea of a similar social preference being present also in childhood (Soto-Icaza et al., 2015; Van Der Geest et al., 2002) and even from the very start of life (Bertenthal & Boyer, 2015; Reynolds & Roth, 2018; Taylor et al., 2004). Despite this, depending on the stimuli and the experimental paradigm, children are either more sensitive than adults to social information (Doherty et al., 2019), show a weaker attentional preference for social stimuli than do adults, or show a mild developmental increase in such preferences across childhood and adolescence (Amso et al., 2014). However, this body of research has mostly focused on attention to isolated humans, and has not yet looked at the influence of more complex social information, such as social interactions, which can provide a much richer source of social information.

Indeed, recent and ongoing research suggests the human visual system may be tuned to social interactions across a variety of experimental designs (Papeo, 2020; Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017). Regions in the human brain also appear to be strongly sensitive to interactive content, as opposed to simply human information, across several stimulus types (e.g. Walbrin & Koldewyn, 2019), and this is true for both adults and children, although regions sensitive to interactive information are not fully 'tuned' until adulthood (Sapey-Triomphe et al., 2017; Walbrin et al., 2020). Attention research has shown a search advantage for facing dyads compared to non-facing agents (Papeo et al., 2019; Vestner et al., 2020), and in naturalistic scenes, interacting humans to hold attention for longer compared to non-interacting agents (Skripkauskaite et al., n.d.; Stagg et al., 2014; Villani et al., 2015). How such 'tuning' takes place across development, however, is not yet clear as very little research has looked at the perception of social interactions in childhood or adolescence, although many studies have used interacting dyads to look at other research questions (e.g., Chevallier et al., 2015; Riby & Hancock, 2008; Stoesz & Jakobson, 2014). One relevant study does, however, suggest that the preference for attending to interacting dyads is also present in childhood, relative to non-facing agents (Stagg et al., 2014). Similarly, we also know that as early as 5 years of age, children are able to learn and draw social inferences from third party encounters (Brey & Shutts, 2015; Over & Carpenter, 2015; Skinner et al., 2017), indicating that even young children naturally attend to and are able to parse observed social interactions. That social interactions are preferentially attended to and readily understood by even young children lends strength to the proposal that social interactions have biological and evolutionary value, as containing a great deal of unique information to supports processes that are important in social species like social learning and social discrimination (i.e., choosing social partners) across development (Over & Carpenter, 2015; Quadflieg & Penton-Voak, 2017; Quadflieg & Westmoreland, 2019; Skinner et al., 2017).

Another crucial aspect of the human attentional bias is that research has shown that it might be purely reflexive and automatic in nature as well as being relatively independent of low-level features of the stimulus (Birmingham et al., 2009a; End & Gamer, 2017; Langton et al., 2000; Ristic & Kingstone, 2005; Rösler et al., 2017). Whether the 'social bias' is purely automatic is still, however, a matter of debate. Indeed, research suggests that while this bias does not appear to be sensitive to the kind of task that participants are performing (Doherty et al., 2017; Flechsenhar & Gamer, 2017) top-down processes such as beliefs and knowledge about what the viewer is seeing are able to modulate both perception and the details of automatic social orienting. Indeed, feature attribution to the same ambiguous stimulus – namely thinking the stimulus is a car with wheels vs. a face with eyes – can influence social orienting, by inducing a face-like cuing effect only when the stimulus is thought to be a face (Ristic & Kingstone, 2005). Similarly, this kind of attentional cuing effect was found to be strongest when the viewed stimuli – a face or a robot – were thought to have intentionality compared to just being inanimate (Wiese et al., 2012). Finally, the modulation of perception by top-down knowledge can go so far as enhancing the response of category-specific brain structures after participants learn new visual information that helps identify initially meaningless stimuli (Dolan et al., 1997). Together, this research suggests an interesting interplay between bottom-up automatic processes and top-down knowledge so that visual attention and visual perception can operate together to evaluate the social and biological relevance of the attended stimulus.

Crucially, although the social value and the attention-driving qualities of interactive information have both been demonstrated, the interplay between automatic social orienting and top-down 'knowledge' modulation has not been investigated in complex scenes that include possible social interactions. Additionally, it is unknown how, and how early these processes start operating, therefore investigating them across development appears crucial. Indeed, we know that social understanding develops since early in life (Soto-Icaza et al., 2015) and in Chapter 3 we have seen that children are similar to adults in the way their attention is preferentially captured and engaged by social interactions, but it is not known how and when social knowledge extends to more complex information like social interactions nor its relationship with visual attention.

In this study we ask whether pre-existing top-down knowledge and consequent interpretation of a human dyad can modulate orienting of attention in complex scenes. In other words, if participants see a scene as containing a social interaction, will that change how they attend to the social information in the scene compared to how they would have attended to the same scene if it did not depict a social interaction? Can knowledge and the way visual information is interpreted activate interaction specific social orienting by increasing attentional capture and engagement to humans that are perceived as interacting compared to when they are not (Skripkauskaite et al., n.d.)? The answer to these questions could have fundamental implications not only for the way our social-cognitive system works, but would also highlight the importance of understanding how social knowledge and social exposure refine and tune social *attention* across development.

Thus, the aim of this work was to 1. Investigate attention to socially ambiguous scenes, and assess whether pre-existing top-down knowledge and beliefs about a social scene can modulate social orienting~~.~~ 2. Explore whether children and adults interpret ambiguous scenes similarly as well as how this interpretation influences social attentional orienting in ambiguous scenes across development.

In Experiment 1, we investigated whether considering a picture as interactive or not influences social orienting in adults, then investigated the same phenomenon in children aged 6 to 12 years to explore developmental changes in Experiment 2, and finally we investigate possible differences between children and adults. For both groups, we expected that categorising pictures as interactive would moderate social orienting, resulting in more attention to social information in the scenes categorized as interactive compared to the ones considered to be non-interactive (Skripkauskaite et al., n.d.). Additionally, while we have shown in chapter 3 that children are very similar to adults in the way they engage and orient attention to social interactions, those scenes were chosen to be unambiguous and were based on adult judgments, therefore we don't know whether children deal with social ambiguity in the same way. We expected developmental differences in the way the scenes are interpreted and expected children would categorise fewer scenes as interactive compared to adults.

General methods

Both studies pre-registered on AsPredicted and power analyses were performed in G power (Erdfelder et al., 2009; Faul et al., 2007) meant to reach 80% of power ($\alpha \leq .05$) and to detect a large effect size (Cohen's f = .40) of social categorization and AOI on participants' attention. All participants had normal or corrected-to-normal vision.

Stimuli & apparatus

All stimuli were selected from an on-line open database (SUN database, Xiao et al., 2010). We chose the pictures to represent various ordinary life contexts (e.g. schools, shops, markets), to be as emotionally neutral as possible, and to depict two people either having a social interaction or not – i.e., involved in independent actions. The initial stimulus set included 127 pictures, and they were rated by 26 naïve participants for the level of "interactiveness" on a Likert scale from 1 ("non-interactive at all") to 7 ("very interactive"). The 30 pictures that were most 'in the middle' in the ratings were selected for this study (cfr. Skripkauskaite et al., n.d.) as raters were not able to categorise them as clearly interactive or non-interactive. Mean interactiveness scores were between 2.92 and 4.08 ($M = 3.59$, $SD = 0.45$) (see "Materials" section in General Methods chapter for details; Appendix B3 for full set of stimuli).

Each picture was pre-processed with Photoshop (version CC 2019) where it was neutralized to remove colour cast ("adjustments – match color – neutralize color"), automatically matched for colour with a picture sample and finally, sharpened ("filter – sharpen").
The scenes were presented on a grey background, using Psychopy 2 (Peirce et al., 2019), on a 380 x 215 mm (1920 x 1080 px) screen. Each stimulus had a size of 860 x 860 pixels (13.6° x 13.6° visual angle) and was presented with the margin closest to the centre either shifted to the left or right from the fixation cross by 60 pixels (0.85° visual angle) so that participants had to move their eyes from the central fixation in order to view the picture. All data were collected with an EyeLink Portable Duo Tracker (EyeLink x, SR Research, Ontario, Canada) with remote

binocular set up at a 1000hz sampling rate. Data were collected from both eyes, but only data from one eye was used for the analysis. We choose each participant's 'best' eye by looking at the calibration accuracy.

Procedure

Each participant sat comfortably on a stable, still chair at 80 cm from the screen and freely viewed 142 pictures: 30 pictures were part of this experiment, while the other 112 pictures belonged to two other experiments and will not be further discussed here. However, all pictures were emotionally neutral, naturalistic scenes that all contained between 2 and 4 human figures. All stimuli were fully randomized to ensure a different presentation order for each participant, and presentation side for individual pictures was counterbalanced across the sample.

Participants were asked to keep their attention on the screen and freely view each picture for 5 seconds with no other specific request, except to return their gaze to the central fixation cross between trials in order to move to the next picture. After participants had viewed all the stimuli and taken a short break, they were shown a physical copy of each of the 30 'ambiguous' pictures in this experiment, and were asked to categorize them into two separate piles by answering the question "Are the people in the scene having a social interaction or not?". How each scene was categorized by each participant was then recorded by the experimenter. Before the free viewing session started, a 13-point calibration procedure was carried out for each participant.

Data preparation

We created hand-drawn areas of interest (AOIs) for each picture using the "freehand" option in Eyelink Data Viewer (SR Research, 2013): one social AOI that included all visible parts of the two humans in the scene, and one non-social AOI that including everything else in the scene excluding the human AOI (i.e., all background elements and all objects).

Dwelling time, which for our purposes included both fixations and saccades, measured the time spent looking at each AOI in milliseconds (ms) and was extracted as a measure of attentional engagement with social and non-social information in the scene. Attentional capture was measured through the time to first fixation for each AOI, defined as the time in milliseconds that the eye took to look at a specific AOI for the first time in each scene (i.e., how long before a participant looks at the social vs. the non-social information).

For each participant and for each picture, we therefore had two measures of engagement and two of capture, divided by attention to social vs. non-social information as well as the post-hoc judgment of the stimulus by each participant as interactive or non-interactive.

Unless stated otherwise, time to first fixation data in all 3 parts of the analysis were transformed (logarithm in base 10) to meet multilevel modelling assumptions. In the manuscript we present the untransformed data for ease of understanding, and details about all transformations and the analyses using models to assess transformed data are included in the Supplementary materials n. S2 (Appendix E).

Part 1 – Attention to ambiguous social scenes

Aim

The main goal of the first part of this work was to investigate the social bias in socially ambiguous naturalistic scenes and the role of feature attribution – i.e. perceiving a scene as depicting a social interaction or not – in moderating attentional orienting and engagement. Besides re-confirming a social attentional bias – i.e. more attention to the social information compared to the background – across all scenes, we expected that scenes later categorized as interactive would show an increased attention to human information compared to when that same scene was categorized as non-interactive (cfr. Skripkauskaite et al., n.d.).

Participants

The a priori power analysis indicated a sample of 70 adults (pre-registered on AsPredicted: https://aspredicted.org/HB2_J6Y) (Appendix A5), calculated based on pilot dwell time data showing increased attention to human figures in photographs judged as interactive by independent judges (η^2_p = .68). We recruited 73 participants but data from two participants were eliminated because they were out of our desired age-range (18-35) and one because of poor tracking data quality due to sleepiness. The final sample thus consisted of 70 participants (*M* = 21.07, *SD* = 2.63, range = 18-35; 47 female and 1 other). All participants provided informed consent and received either monetary compensation or university credit for their participation.

All procedures were approved by the ethical committee at Bangor University (ethics protocol number: 2018-16360).

Procedure

Participants were instructed to freely explore the pictures, both through on-screen visually-presented instructions and orally. Participants did not know that they would be asked to categorise the pictures before viewing them and researchers did not mention interactive content, simply stating that the pictures would be shown one at a time and that they should keep their attention on the pictures. Before the presentation of each stimulus, the participant performed a drift correction procedure, in which they had to fixate a calibration point at the centre of the screen and then press the space bar to proceed to the next scene. These between-trial procedures served to draw participants' gaze back to the centre of the screen before the start of each trial. The entirety of the eye-tracking procedure lasted around 20 minutes and consisted of 4 blocks of 35 trials each. Participants could take breaks between blocks and rest their eyes if necessary, and another drift correction procedure was performed before each re-start.

Data analysis

Trials with less than 33% of total time of engagement with the stimulus were treated as missing (S. Fletcher-Watson et al., 2009) – which including both off-screen time and missing data because of poor signal or blinks. In the adult group, this produced the loss of 0.42 % trials, which represents between 0 -2 trials per participant.

We specified separate models for dwelling time and time to first fixation. After assessing the variance in the dataset, we analysed the data using multilevel modelling with a 2 x 2 structure of predictors (nlme package (Pinheiro et al., 2016) in a hierarchical model. The model structure was pre-registered as a four-level model, with participant information at the highest level, trial information with the participants' categorization (interacting or non-interacting) nested within participant information at the 3rd level, AOI type (human or background) at the second level, and the dependent variable (dwell-time or time-to-first-fixation) at the first level. Despite this, the model building procedure showed that participant's scene categorization and participant information were in fact not nested. In particular, categorization of the scene and participant information were revealed to be at the same level, which is not surprising considering that every participant had their own categorization of any given scene as interacting or not. Therefore with participant variance, the categorization scores also varied. Additionally, for each model we added the size of the AOI as a random factor to account for variance in the size of AOIs across pictures and the fact that the background AOI is always bigger in size than the human AOIs. We then compared the originally planned model – a three-level model with participant information at the highest level, AOI type at the second, and the dependent for each AOI at the first level, nested within trial and participant – with the model including size as a random factor, and in every case the two models were not statistically different (all *p*'s > .99). Post-hoc pairwise comparisons were performed using Tukey's HSD (emmeans package (Lenth et al., 2018)).

Results

Categorization results

On average adults categorized a higher proportion of scenes as interactive ($M = 60.67$, $SD = 12.43$) compared to non-interactive ($M = 38.90$, $SD = 12.21$; $t(69) = 7.4$, $p < .001$, $d = 0.89$) (Figure 1).

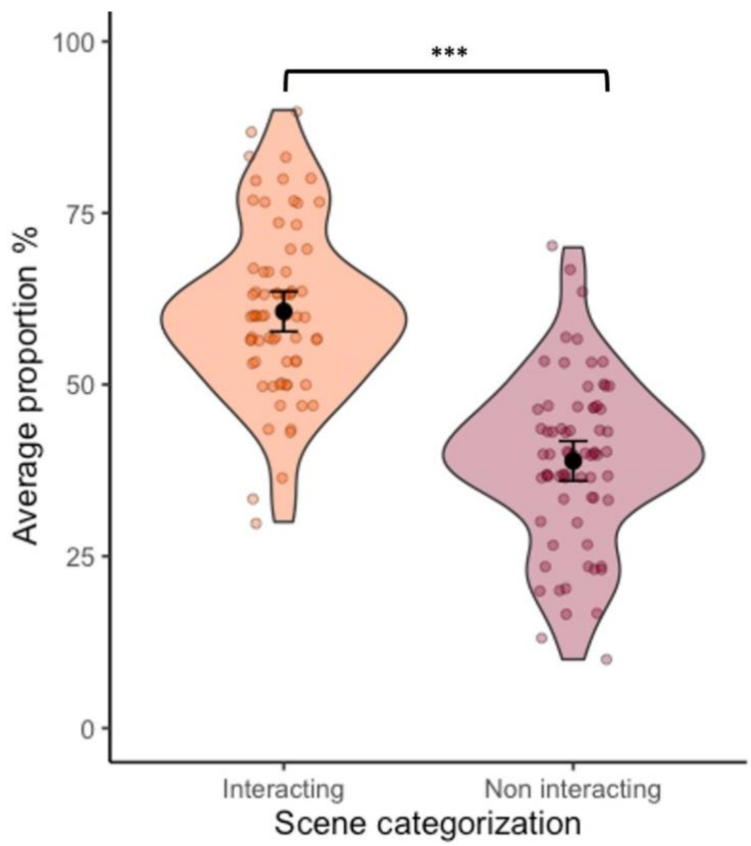

Figure 1. Average proportion of scenes categorized as interacting vs non-interacting. Error bars represent 95% confidence intervals.

Attentional engagement

The model building procedure indicated that there was significant variance of intercepts across participants and type of AOI – social or background – *SD* = 257.26, $\chi^2(2) = 185.21$, $p < .001$, and adding the area of the AOIs in pixels as a random effect did not alter the model significantly from the above mentioned – *SD* = 843.41, $\chi^2(3) = 0$, $p = .99$. Therefore, in the final model, participant, AOI and size of the AOI were set as random effects, and participant's scene categorization and the type of AOI were set as fixed effects.

Dwelling time was not significantly different for scenes categorized as interactive (*M* = 1802.10, *SD* = 887.53) compared to the ones categorized as non-interactive (*M* = 1790.33, *SD* = 876.14) –$F(1,4040) = 0.09$, $p = .76$, $\eta^2_p < .001$. As expected, more time was spent looking at the human AOIs (*M* = 1864.05, *SD* = 890.47) compared to the background (*M* = 1730.95, *SD* = 870.63). In other words, there was a main effect of AOI, $F(1,69) = 7.29$, $p = .01$, $\eta^2_p = .07$. Finally, contrary to our prediction, this social attentional bias was not moderated by the type of scene. In other words, scenes categorized as interactive (social: *M* = 1869.89, *SD* = 895.70; background: *M* = 1734.31, *SD* = 874.39) were not attended to differently to those that were categorised as non-interactive (social: *M* = 1854.93, *SD* = 882.72; background: *M* = 1725.73, *SD* = 865.23) – $F(1,4040) = 0.004$, $p = .95$, $\eta^2_p < .001$ (Figure 2). Although the interaction between categorization and type of AOI was not significant, we performed planned post-hoc comparisons which did not reveal any difference between scene-types but instead show that when the scenes are split by their categorization, the human bias weakens to trend levels in *both* types of scenes (interacting: $t(69) = 2.45$, $p = .06$, $d = 0.30$; non-interacting: $t(69) = 2.27$, $p = .08$, $d = 0.27$).

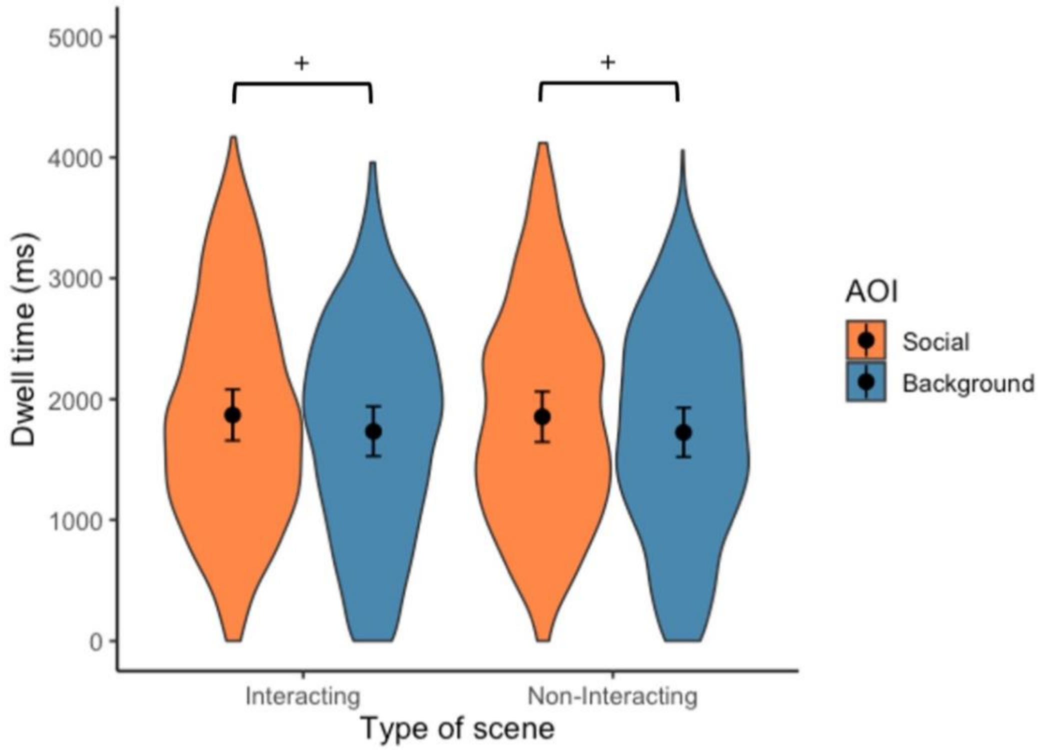

Figure 2. Violin plot for mean dwelling time (ms) to social and background AOIs in scenes categorized as interacting and non-interacting. Error bars represent 95% confidence intervals.

Attentional capture

In the model for the time to first fixation (see Supplementary materials n. S2- 2a for details on the log-transformation and the full analysis and results from transformed data), participant and AOI showed significant variance of intercepts – SD = 202.72, $\chi^2(2) = 179.77$, $p < .001$, and adding area of the AOIs in pixels as a random effect did not change the model significantly from the originally planned one – SD = 658.52, $\chi^2(3) = 0$, $p = .99$.

Participants were equally fast to orient to the scenes categorized as interactive (M = 655.74, SD = 708.57) as to the ones categorized as non-interactive (M = 653.77, SD = 660.38) – $F(1,3958) = 0.0001$, $p = .99$, $\eta^2_p < .001$. As predicted, attentional capture was faster for the human AOIs (M = 486.35, SD = 472.50) compared to the background (M = 827.24, SD = 822.26) characterised by a main effect of AOI, $F(1,69) = 152.54$, $p < .001$, $\eta^2_p = .69$. As in the dwell-time analysis, this social attentional bias was not moderated by the type of scene so that scenes categorized as interactive (social: M = 484.10, SD = 472.94; background: M = 831.12, *SD* = 852.09) showed a similar social bias to those categorised as non-interacting (social: *M* = 489.87, *SD* = 472.06; background: *M* = 821.18, *SD* = 773.92); *F*(1,3958) = 0.07, *p* = .80, η^2_p< .001 (Figure 3).

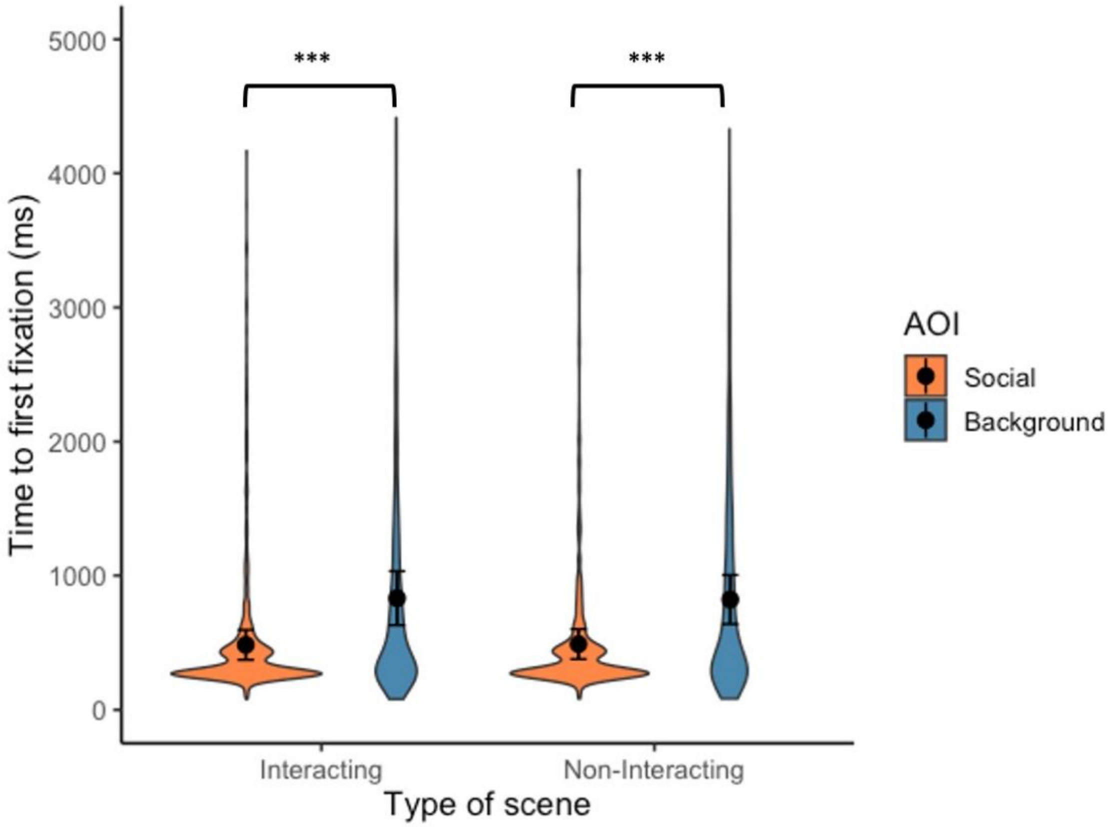

Figure 3. Violin plot for mean time to first fixation (ms) to social and background AOIs in scenes categorized as interacting and non-interacting. Error bars represent 95% confidence intervals.

Interim discussion

In this section we show that when asked to decide whether two people are having a social interaction in a socially ambiguous scene, adults seem to be more likely to consider the scene as interacting. Against our predictions, despite showing an attentional bias to social information in both engagement and capture, this bias did not change depending on whether the viewer considered the scene to be interactive or not. On one hand, this is not surprising because the scenes *are the same* regardless of whether they are categorised as interactive or not. On the other hand, this result was contrary to our expectations because non-ambiguous interactive cues clearly drive engagement and capture in prior work. One interpretation of this result is that top-down semantic judgement about the scene might not modulate visual attention and that the visual properties of the scene, including social objects and cues, primarily drive attention, particularly during this kind of 'free-exploration' task. On the other hand, this result may reflect our choice to measure "spontaneous" knowledge rather than openly manipulating interactive categorisation, for example through priming procedures (Ristic & Kingstone, 2005; Wiese et al., 2012).

Interestingly, although we cannot directly compare the two, the social bias in attentional engagement appears to be milder than that seen for less ambiguous scenes (cfr. Chapter 3) but just as robust for capture. This indicates that participants' attention was first strongly captured by the human component of the scene, in line with previous social scenes research (Birmingham et al., 2009a; Sue Fletcher-Watson et al., 2008) but perhaps, given the ambiguity of the scene, more attention was then given to the background, resulting in a milder engagement bias. This suggests that when the scene is harder to disambiguate or interactive information is less diagnostic, there is more need for contextual information to understand the scene.

Part 2 – Attention to ambiguous social scenes in pre-adolescent childhood

Participants

Our a priori power analysis (pre-registered on AsPredicted: https://aspredicted.org/L9K_KRS) (Appendix A5) for the developmental group indicated a sample of 90 children and teenagers between 6 and 18 years to have enough power to detect a three-way interaction, and pre-registered the intention to stop data collection at 90 participants or by the 31st December 2020. Unfortunately, we were forced to interrupt data collection due to the COVID-19 pandemic and were unable to resume data collection before the end of 2020. Thus, we were only able to collect data from 6-12 year-old children and were unable to collect data from our planned adolescent sample. Therefore, this sample included 54 participants across a narrower-than-planned age-range (*M* = 8.76, *SD* = 1.72; range = 6-12; 28 female). Each child gave informed assent to participate, and each child's guardian(s) signed a consent form. Children received a small toy of their choice as compensation as well as small prizes, such as stickers and pens. All procedures were approved by the School of Psychology's Ethics committee at Bangor University (ethics protocol number: 2019-16586).

Procedure

Our procedure here was nearly identical to that used with the adults. To minimize written content, children were verbally instructed regarding the task, and received additional instruction about staying (relatively) still and keeping their attention focused on the screen. One significant change in procedure is that children did not have a drift correction procedure before each trial, as we found that this caused confusion and frustration. Instead, we carried out drift correction procedures only between blocks. In order to draw participants' gaze back to the centre of the screen before each stimulus, children were shown an animated gif at the centre of the screen for approximately 2.5 seconds as a fixation point, without the need to press a key to proceed.
After the free-viewing session, children sorted pictures into interactive or non-interactive (two-alternative forced choice decision). They were instructed to look carefully at each picture and decide whether the two people in the scene were having a social interaction or not. With the youngest children, before the sorting task, we gave a brief explanation of what a social interaction is, where children were encouraged to think about whether people in the scene were doing something with each other or not. Finally, as a way to increase motivation and engagement with the task, children completed a "sticker chart" as they progressed through different steps in the procedure.

Data analysis

As in the adult data, trials with less than 33% of total engagement with the stimulus were treated as missing. This led to the loss of 2.35 % of the total trials with a range of 0 – 9 trials per participant. Similar modelling procedures were carried as in the adult analysis. The pre-registered analysis was a four-level hierarchical model with a 2 x 2 structure of predictors (nlme package (Pinheiro et al., 2016)) with participant information at the highest level, and nested within each participant the scene categorization – interacting or non-interacting – followed by the AOI information and finally at the first level our measure, nested within trial and participant. As for the adult sample analysis, the hierarchical model building procedure indicated that scene categorization and participant information were at the same level (thus not truly nested), therefore the final model included only three-levels: participant information at the third level, AOI type at the second level, and the dependent variable – dwelling time or time or time to first fixation – at the first level, nested within trial and participant. Post-hoc pairwise comparisons were performed using Tukey's HSD (emmeans package; Lenth et al., 2018). Additionally, to allow investigation of developmental change across the included age-range, participants' age was modelled as a continuous predictor.

Results

Categorization results

On average, children categorised an equal proportion of pictures as interactive (M = 52.16, SD = 16.10) as non-interactive (M = 46.67, SD = 16.58), $t(53) = 1.61$, $p = .11$, $d = 0.21$), but the proportion of scenes categorised as interactive increase with age (Figure 4). Indeed, age interacted significantly with scene categorization in the children's group, $t(104) = 2.10$, $p = .04$, $d = 0.21$, with an increasing proportion of scenes categorized as interactive from 9 years on (Supplementary materials n. S3 for descriptive statistics by age).

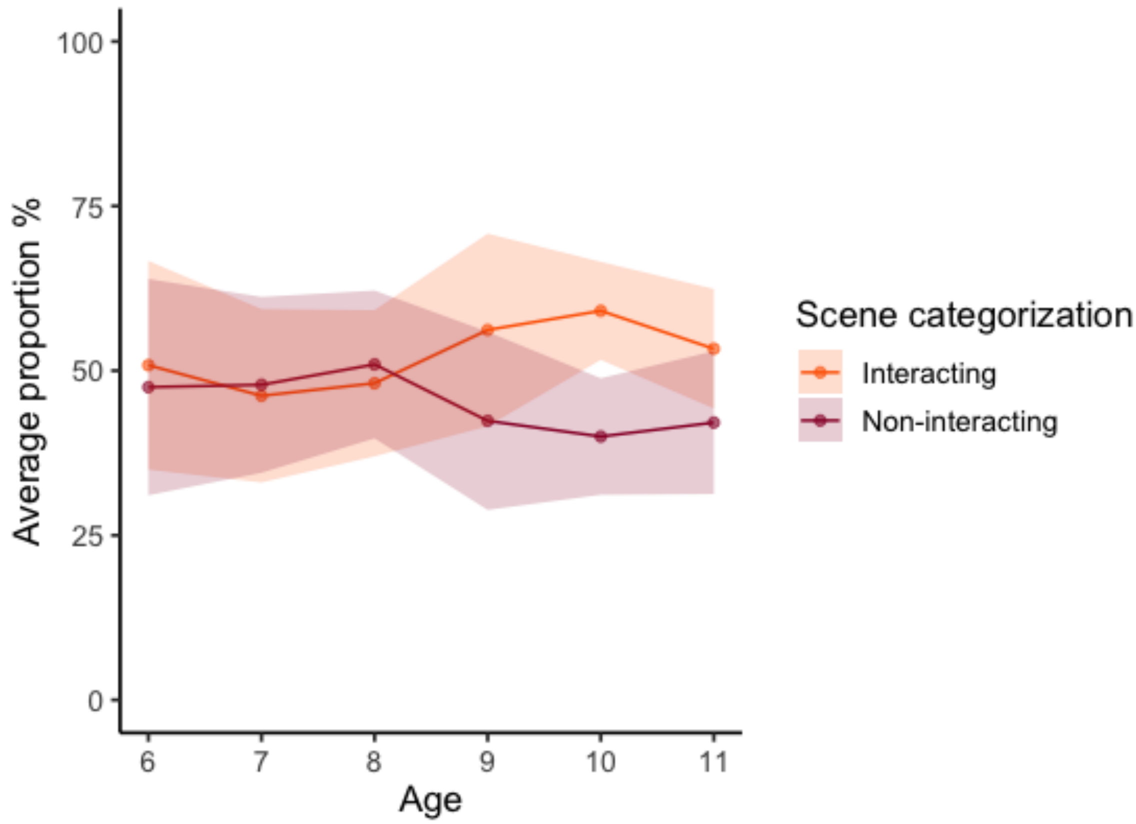

Figure 4. Average proportion of categorization of the scenes as interacting or non-interacting, in relation to age in the developmental group. Width of bands represents 95% confidence intervals.

Attentional engagement

In the developmental sample there was significant variance of the intercepts across participants and type of AOI – social or background - $SD = 256.84$, $\chi^2(2) = 68.12$, $p < .001$ and when adding the area of the AOIs in pixels as a random effect (see Supplementary materials n. S1 for details on AOI size), this model did not differ significantly from the originally planned one – $SD = 1072.12$, $\chi^2(3) = 0$, $p = .99$. Therefore, participant, AOI and size of the AOI were set as random effects, and participant's scene categorization and type of AOI were set as fixed effects while age was modelled as a continuous predictor. Additionally, age of participants was group mean centered.

The overall attention given to scenes, collapsed across AOI type, did not change with age, $F(1,52) = 2.64$, $p = .11$, $\eta^2_p = .05$. Additionally, overall dwelling time was not significantly different for scenes categorized as interactive ($M = 1852.42$, $SD = 1105.10$) compared to those categorized as non-interacting (M = 1839.62, SD = 1102.22) – $F(1,3028) = 0.01$, $p = .92$, $\eta^2_p < .001$, and this was not moderated by the age of the participants – $F(1,3028) = 0.003$, $p = .95$, $\eta^2_p < .001$.

Unsurprisingly, children spent more time looking at the human AOIs (M = 1927.62, SD = 1103.47) compared to the background (M = 1765.40, SD = 1098.13) with a main effect of AOI, $F(1,52) = 7.32$, $p = .01$, $\eta^2_p = .12$ that was similar across all ages, $F(1,52) = 1.49$, $p = .23$, $\eta^2_p = .03$. However, contrary to our predictions, but similar to the adult data, this social attentional bias was not moderated by the type of scene. Children did not attend to social information in interactive scenes (social: M = 1927.85, SD = 1107.07; background: M = 1777.00, SD = 1098.61) any differently than in non-interactive scenes (social: M = 1927.36, SD = 1100.03; background: M = 1751.89, SD = 1098.17); $F(1,3028) = 0.17$, $p = .68$, $\eta^2_p < .001$ (Figure 5). Although the interaction between categorization and type of AOI was not significant, exploratory post-hoc comparisons show that when the scenes are split by their categorization, the human bias is weaker in both scene types, much as in the adult data [interactive scenes: $t(52) = 2.20$, $p = .10$, $d = 0.30$; non-interactive scenes: $t(52) = 2.46$, $p = .06$, $d = 0.34$. Additionally, no differences were found in the attention to the two types of backgrounds (all p's > .99).

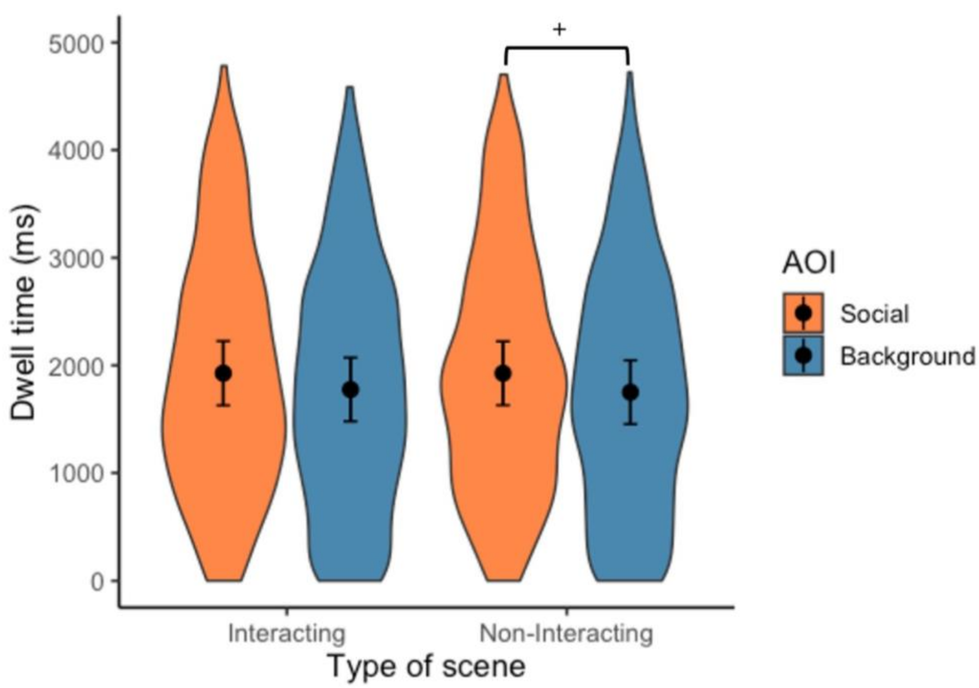

Figure 5. Violin plot for mean dwelling time (ms) to social and background AOIs in scenes categorized as interacting and non-interacting. Error bars represent 95% confidence intervals.

The three-way interaction between age, categorization of the scene, and AOI did not improve the model fit, only reached trend levels, and had a very small effect size – $F(1,3028) = 3.41$, $p = .07$, $\eta^2{}_p = .001$ (Figure 6). However, if this nearly significant interaction can be replicated in a study with sufficient power to properly assess a three-way interaction, this possible finding suggests that the social bias in engagement is slightly stronger in the youngest children compared to the older, echoing similar findings in Chapter 3.

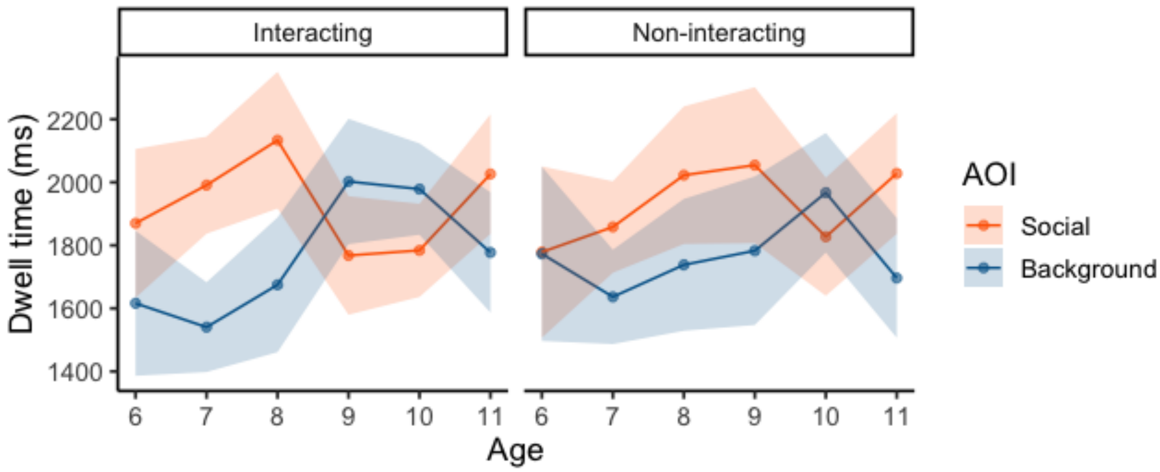

Figure 6. Average dwell time to social and background AOIs in scenes categorized as interacting and non-interacting in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

Attentional capture

In our time to first fixation model (see Supplementary materials n. S2 – 2b for details on transformation and analyses with log-transformed data), participant and AOI showed significant variance of intercepts – $SD = 171.74$, $\chi^2(5) = 61.94$, $p < .001$. As in the dwelling time model, we added area of the AOIs in pixels as a random effect. This model did not differ significantly from the originally planned one – $SD = 722.84$, $\chi^2(5) = 0$, $p = .99$.

The overall speed of orienting to the scenes did not change with age, $F(1,52) = 0.04$, $p = .84$,. $\eta^2{}_p < .001$. Additionally, participants were equally fast to orient to the scenes categorized as interacting ($M = 707.68$, $SD = 726.14$) compared to the ones categorized as non-interacting

(M = 708.98, SD = 763.87; $F(1,2896) = 0.0003$, $p = .99$, $\eta^2_p < .001$), and this was not moderated by the age of the participants – $F(1,2896) = 06$, $p = .80$, $\eta^2_p < .001$. Unsurprisingly, participants were faster to look to the human AOIs (M = 557.68, SD = 596.92) compared to the background (M = 863.76, SD = 841.98; $F(1,52) = 104.16$, $p < .001$, $\eta^2_p = .67$), and this 'social capture' effect was similar for children across our age-range, $F(1,52) = 0.88$, $p = .35$, $\eta^2_p = .02$.

Contrary to our expectations, this social attentional bias was not moderated by the type of scene – i.e. whether children categorized a particular scene as interacting (social: M = 558.27, SD = 601.44; background: M = 861.01, SD = 807.07) or non-interacting (social: M = 557.00, SD = 592.06; background: M = 866.98, SD = 881.77); $F(1,2896) = 0.01$, $p = .92$, $\eta^2_p < .001$ (Figure 7).

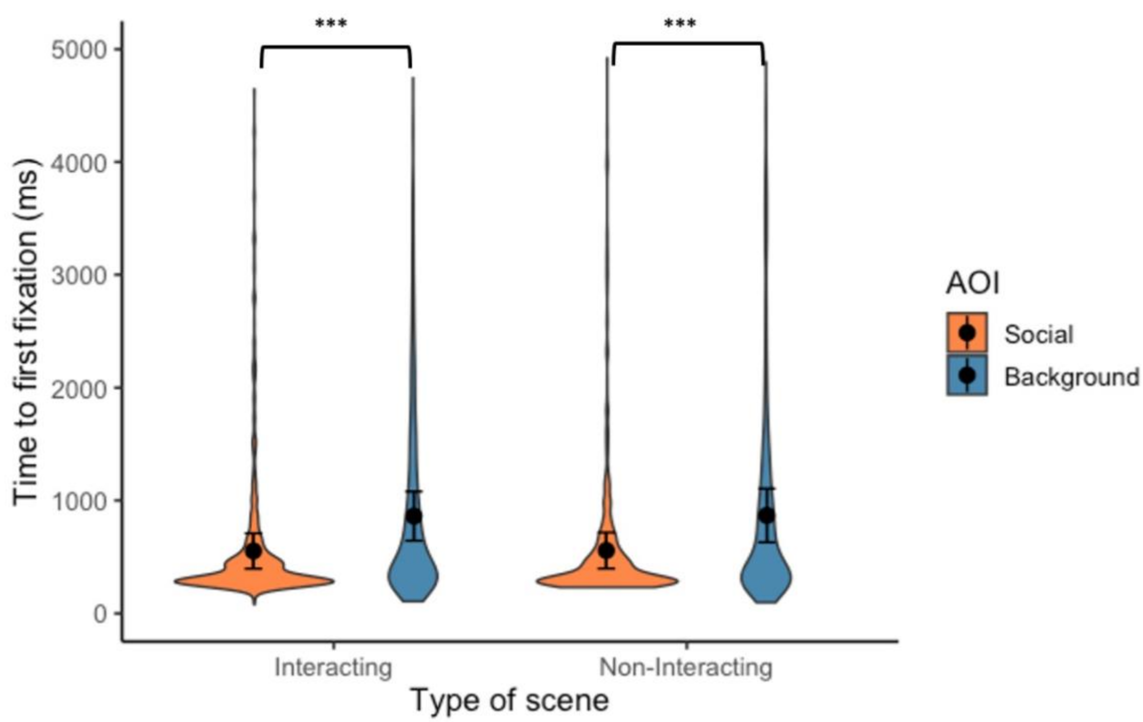

Figure 7. Violin plot for mean time to first fixation (ms) to social and background AOIs in scenes categorized as interacting and non-interacting. Error bars represent 95% confidence intervals.

Finally, the three-way interaction between age, categorization of the scene and the AOI did not improve the model fit, and when analysed, was not significant – $F(1,2896) = 1.60$, p =

.21, $\eta^2{}_p$ = .001 (Figure 8), meaning that the social bias in attentional capture did not change depending on the way the scene was categorised nor as a function of age.

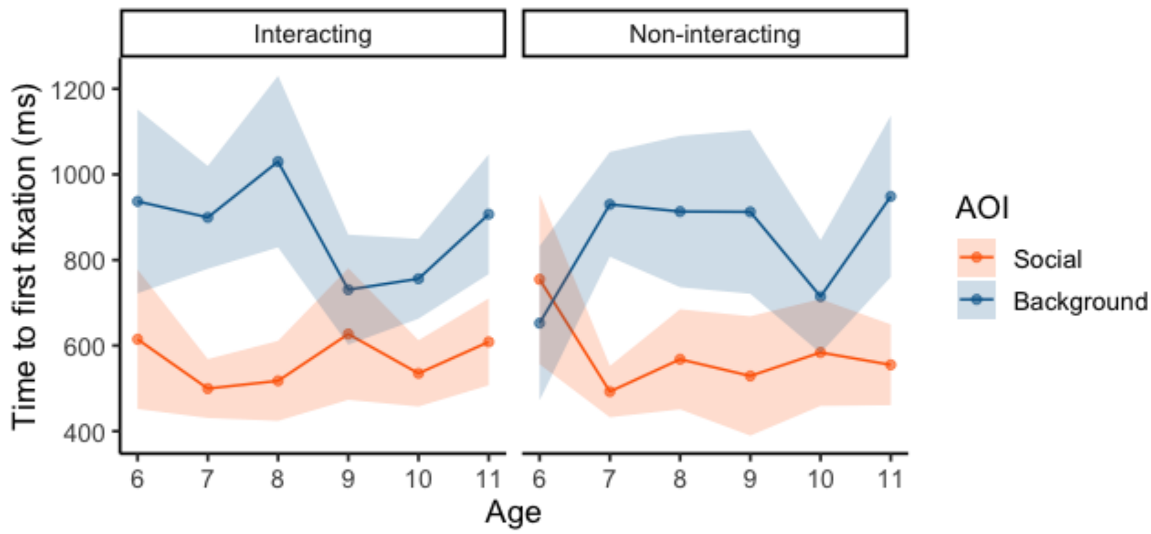

Figure 8. Average time to first fixation (ms) to social and background AOIs in scenes categorized as interacting and non-interacting in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

Part 3 – Development of attention to ambiguous social scenes

Data analysis

In this section of the work the aim is to assess developmental changes between children and adults. Given the gap in age between our children and adults, age in this case was modelled as a categorical predictor (children vs. adults) rather than a continuous predictor as in the child-only analyses. The models used here have a similar structure to the ones used in the previous sections with the final model including age-group, type of scene as categorized by each participant, and AOI as fixed effects, allowing intercept to vary at participant, AOI, and size of the AOIs level.

Results

Categorization results

As might be expected from prior analyses, on average, there were more scenes categorized as interactive (*M* = 56.96, *SD* = 14.71) compared to non-interactive (*M* = 41.45, *SD* = 14.51), $F(1,122) = 73.91$, $p < .001$, $\eta^2_p = .38$ and this was moderated by age (Figure 9). Indeed, age-group interacted significantly with scene categorization $F(1,122) = 15.56$, $p < .001$, $\eta^2_p =$ .11, with adults categorizing a much higher proportion of scenes as interactive (*M* = 60.67, *SD* = 12.43) compared to children (*M* = 52.16, *SD* = 16.10), $t(244) = 3.31$, $p = .004$, $d = 0.21$, but there was also a weaker developmental difference for scenes categorized as non-interacting (children: *M* = 44.75, *SD* = 16.58; adults: *M* = 38.90, *SD* = 12.21); $t(244) = -2.27$, $p = .08$, $d = -0.15$.

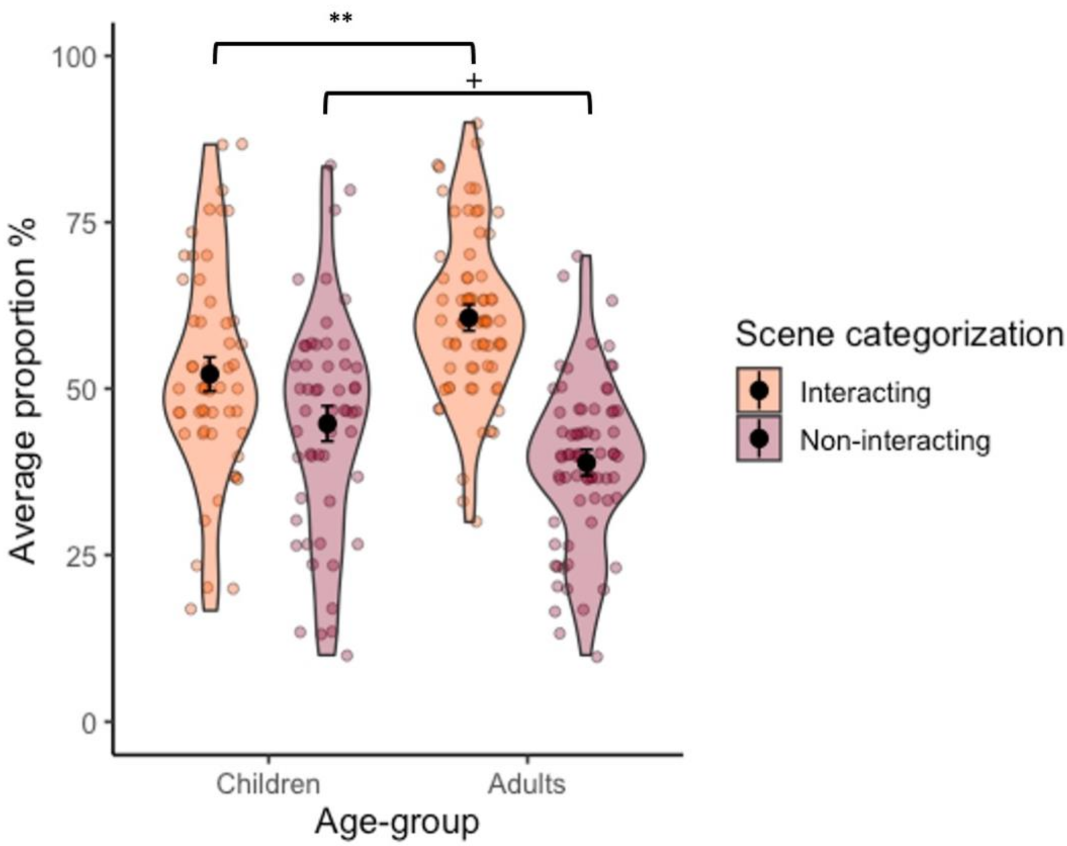

Figure 8. Average proportion of categorization of the scenes as interacting or non-interacting, children and adults. Error bars represents 95% confidence intervals.

Attentional engagement

When assessing dwelling time, there was significant variance of intercepts across participants and type of AOI, *SD* = 258.50, $\chi^2(2) = 234.83$, $p < .001$, and the model including the area of the AOIs in pixels as a random effect did not differ significantly from the originally planned one – *SD* = 947.03, $\chi^2(3) = 0$, $p = .99$.

The overall amount of attention given to the scenes did not change with age, *F*(1,122) = 1.34, $p = .25$, $\eta^2{}_p = .01$, with children (*M* = 1846.51, *SD* = 1103.61) and adults (*M* = 1797.50, *SD* = 883.01) spending a similar amount of time attending to the scenes. As might be expected from prior analyses but contrary to our original hypothesis, dwelling time was not significantly different for scenes categorized as interacting (*M* = 1822.17, *SD* = 980.28) compared to the ones categorized as non-interacting (*M 1813.50*, *SD* = 989.04) – *F*(1,7070) = 0.11, $p = .74$, $\eta^2{}_p < .001$, and this was true for both children and adults – *F*(1,7070) = 0.002, $p = .96$, $\eta^2{}_p < .001$.
All participants spent more time looking at the human AOIs (*M* = 1891.31, *SD* = 987.82) compared to the background (*M* = 1745.73, *SD* = 974.72) – main effect of AOI, *F*(1,122) = 14.32, $p = .0002$, $\eta^2{}_p = .11$, and this was similar for children (social: *M* = 1927.62, *SD* = 1103.47; background: *M* = 1765.40, *SD* = 1098.13) and adults (social: *M* = 1864.05, *SD* = 890.47; background: *M* = 1730.95, *SD* = 870.63), *F*(1,122) = 0.11, $p = .74$, $\eta^2{}_p < .001$. This social attentional bias was not moderated by whether the participants saw the scene as interacting (social: *M* = 1893.01, *SD* = 985.60; background: *M* = 1751.33, *SD* = 970.01) or non-interacting (social: *M* = 1888.98, *SD* = 991.18; background: *M* = 1738.03, *SD* = 981.42) – *F*(1,7070) = 0.13, $p = .72$, $\eta^2{}_p < .001$. Finally, this was similar for children and adults, *F*(1,7070) = 0.11, $p = .74$, $\eta^2{}_p < .001$ (Figure 9; see Supplementary materials n. S3 for descriptive statistics), but when we analyse the three-way interaction, the social bias weakens, as every social > background contrast loses power, underlining the similarity between children and adults.

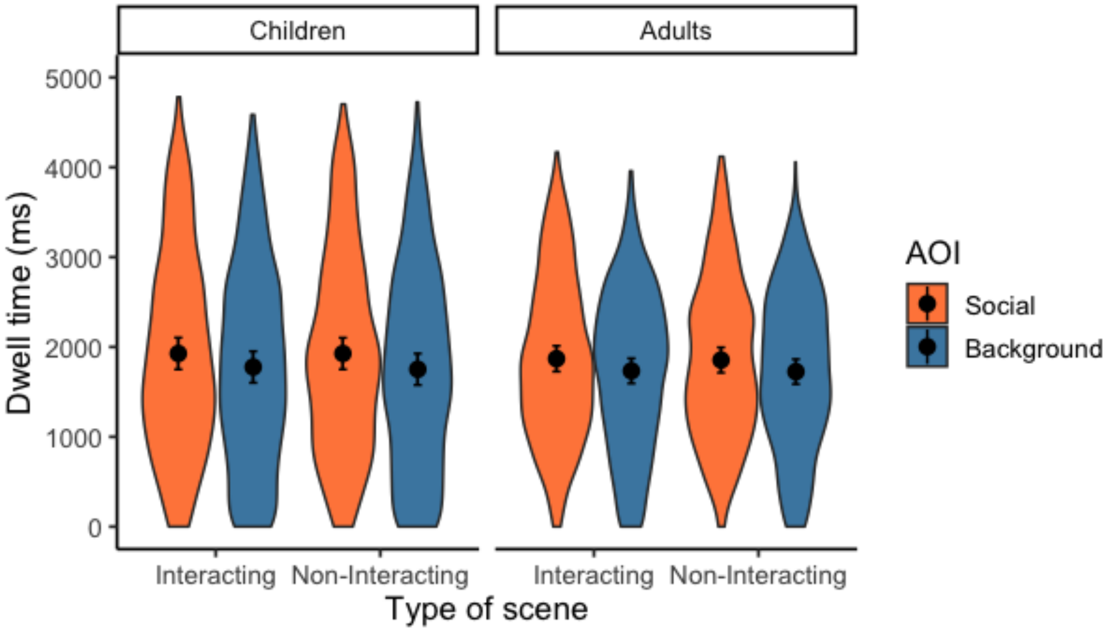

Figure 9. Violin plot for mean dwelling time (ms) to social and background AOIs in scenes categorized as interacting and non-interacting in the two groups. Error bars represent 95% confidence intervals.

Attentional capture

In our time to first fixation model (see Supplementary materials S2 – 2c for details on log-transformation and details on analyses with transformed data), participant and AOI showed significant variance of intercepts – *SD* = 192.22, $\chi^2(2) = 236.91$, $p < .001$ and when size of the AOI was added as a random effect, the model did not change significantly (*SD* = 685.94, $\chi^2(3) = 0$, $p = .99$).

Although the overall speed of orienting to the scenes was significantly slower for children (*M* = 708.28, *SD* = 743.68) compared to adults (*M* = 654.97, *SD* = 690.07), $F(1,122) = 7.03$, $p = .01$, $\eta^2{}_p = .05$, participants were equally fast to orient to the scenes categorized as interacting (*M* = 676.16, *SD* = 715.89) as to the ones categorized as non-interacting (*M* = 679.42, *SD* = 710.75) – $F(1,6856) = 0.001$, $p = .98$, $\eta^2{}_p < .001$, and this was similar for children (interacting: *M* = 707.68, *SD* = 726.14; non-interacting: *M* = 708.98, *SD* = 763.87) and adults (interacting: *M* = 655.74, *SD* = 708.57; non-interacting: *M* = 653.77, *SD* = 660.38); $F(1,6856) = 0.00$, $p = .99$, $\eta^2{}_p = .00$.

Unsurprisingly, more capture was driven by the human AOIs (*M* = 516.63, *SD* = 529.99) compared to the background (*M* = 842.65, *SD* = 830.72) indexed by a main effect of AOI, ($F(1,122) = 255.37$, $p < .001$, $\eta^2{}_p = .68$), but this effect was not strongly moderated by age, $F(1,122) = 0.70$, $p = .41$, $\eta^2{}_p = .01$, but exploratory post-hoc comparisons show that children were slower to orient to social information compared to adults (children: *M* = 557.68, *SD* = 596.92; adults: *M* = 486.35, *SD* = 472.50; $t(122) = 2.46$, $p = .05$, $d = 0.22$) but no developmental difference was found in the capture by background information (children: *M* = 863.76, *SD* = 841.98; adults: *M* = 827.24, *SD* = 822.26; $t(122) = 1.30$, $p = .46$, $d = 0.12$). However, this likely simply reflects the generally slower orienting for children compared to adults, combined with the fact that the first fixation for both groups is almost always to the human information in the scene.

As might be expected from prior analyses, the social attentional bias was not moderated by the type of scene [interactive (social: *M* = 513.30, *SD* = 528.38; background: *M* = 842.86, *SD* = 834.63) or non-interactive (social: *M* = 521.20, *SD* = 532.33; background: *M* = 842.36, *SD* = 825.56); $F(1,6856) = 0.04$, $p = .85$, $\eta^2{}_p < .001$] and it was not different between age-groups as there was no evidence of a three-way interaction between scene-type, AOI-type, and age-group, $F(1,6856) = 0.06$, $p = .80$, $\eta^2{}_p < .001$ (Figure 10; see Supplementary materials n. S3 for descriptive statistics).

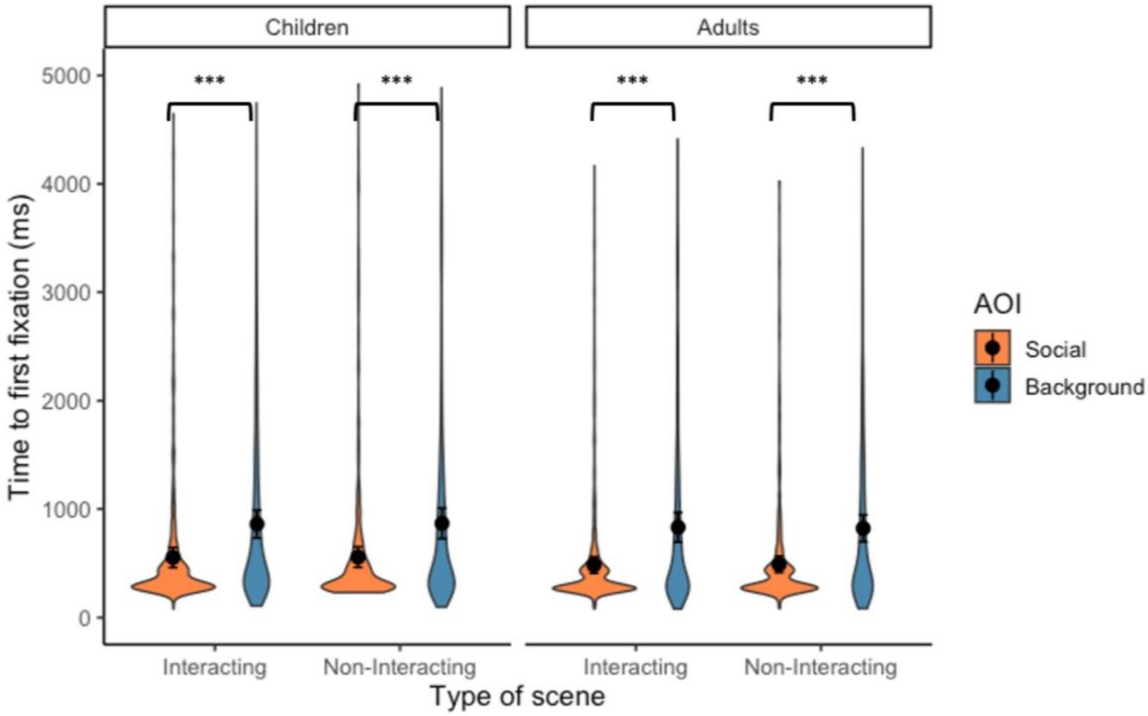

Figure 10. Violin plot for mean time to first fixation (ms) to social and background AOIs in scenes categorized as interacting and non-interacting in the two groups. Error bars represent 95% confidence intervals.

General discussion

In this work, we aimed to investigate the role of spontaneous social knowledge in modulating the social attentional bias in adults and pre-adolescent children, specifically investigating if seeing an ambiguous scene as depicting a social interaction (vs. being non-interactive), would influence how participants attended to social information in the scene. Against our predictions, we found no clear influence of seeing the same scene as interactive vs. non-interactive on social attention in either children or adults. There is, however, a developmental difference in how scenes were categorised, where adults were much more likely to 'see' scenes as interactive compared to the children as a group, with this difference likely being driven primarily by the youngest (< 9 years old) children. Additionally, while the human attentional bias was large and reliable in our attentional capture measure, this bias was weaker and less reliable in our measure of attentional engagement than has been reported in previous work looking at unambiguously interactive vs. non-interactive scenes. Although we could not make that comparison directly in the current data set, this suggests that ambiguous social scenes are explored differently than scenes that are less socially ambiguous, with participants attending more to background information in ambiguous scenes.

That adults categorized more scenes as interactive compared to children might be indicative of adults having either more complex social knowledge, more experience with and exposure to social interactions, or a larger bias to interpret social content as interactive even in ambiguous situations. On the other hand, we cannot rule out that younger children struggled more with the concept of "interaction" and categorised the pictures more randomly than did either older children or adults. However, that children look more 'adult-like' in how they categorise scenes from 9 years of age is consistent with research showing that 9-12-year-old children were more similar to adults in the way category specific brain structures responded to social interactions, while the same regions in younger children were less "tuned" to interactive information (Walbrin et al., 2020). Additionally, structural development of temporal social brain regions appears to peak at around 9 years of age (Mills et al., 2014) suggesting that it is at this age that a wider network is available for processing social information in more complex scenarios such as ambiguous scenes that might contain social interactions.

The lack of modulation by spontaneous social knowledge about social interactions on orienting to social information contrasts somewhat with past research on the topic (Ristic & Kingstone, 2005; Wiese et al., 2012). In these studies feature attribution or interpretation of visual information was deliberately manipulated, resulting in moderated attentional patterns while in our data scene categorisation is spontaneous and had no effect on the way social attention was deployed. This is suggestive of two possibilities: either top-down information cannot modulate attentional mechanisms towards social interactions, or that attentional modulation requires a clear manipulation of knowledge or top-down information about the stimuli. It may also matter that we ask for categorisation *after* participants' first viewing of the scenes, when eye-tracking data was collected. While we deemed this necessary, so as not to bias participants to *look for interactions*, we cannot be certain that they 'saw' the stimuli the same way when viewed for a 2nd time. How likely is it that participants might see the scenes differently the 2nd time than they did the first? How might our results have been different if they were tasked with categorising the scenes in the first place? Looking at attentional patterns *while* participants are assessing ambiguous dyads for 'interactiveness' could, in future work, help to clarify these questions.

Finally, the finding of a social bias in capture and engagement also in complex and socially ambiguous scenes is consistent with past research investigating attention to social information in naturalistic scenes (Bindemann et al., 2010; Birmingham et al., 2008; Doherty et al., 2017; Flechsenhar & Gamer, 2017; Sue Fletcher-Watson et al., 2008). Interestingly, our findings extend this research by showing that attentional engagement to human information is weakened statistically compared to what we saw in the Chapter 3, suggesting perhaps a need to use more contextual information to interpret and disambiguate the scene. This finding makes intuitive sense, as it supports the idea that social information always captures attention but then is engaged differently depending on the type of social information: if there is a clear social interaction, social objects hold attention for longer compared to when interactive content is absent (Skripkauskaite et al., n.d.) but if the scene is interactively ambiguous, this bias to social information is milder because more contextual (background) exploration is necessary to extract meaning.

The current research has several limitations. First, similarly to the other chapters, analyses evaluating developmental changes between children and adults could benefit from a greater sample size. Similarly, the addition of the originally planned adolescent sample is needed to clarify how and when changes in the interpretation of ambiguous dyads becomes 'adult-like'. Additionally, future research should consider directly manipulating the knowledge around the ambiguous stimuli to better disentangle the influence of top-down processes in attention to social interactions, for example by using priming methods. Indeed, inducing feature attribution participants by explicitly telling them that what they are viewing is a social interaction (or not), may be a better measure of how knowledge and interpretation of visual information influences social attention.

Despite the limitations, this study extends research on the interplay between social knowledge and attentional orienting by demonstrating that while there are indeed developmental changes in the way social information is interpreted, those interpretation differences do not greatly modify social attention across age group, informing future directions in developmental research.

Chapter 6. General discussion

Much research has established the exquisitely social nature of humans and their brains, by showing how tuned we are towards social information, especially faces, since very early in life. However, the vast majority of available research looking at social perception across development has focused on face and gaze perception in infancy, therefore it is unknown how attention to and perception of more complex social information develops in childhood. What's more, humans act socially in noisy and cluttered environments. Thus, we do not only have to be able to process isolated, individual faces, but are bombarded by a variety of social information and cues embedded in complex scenes. Recent research investigating 'people watching' has suggested a special place for observed social interactions in human vision. Yet, research into how our attention orients us to these complex sources of information is limited. How do humans become so skilled at detecting and understanding interactive cues in adulthood? How do we deal with the flood of visual information we're constantly bombarded with and how do we extract the social cues/information we need to be successful throughout life? Interestingly, although the social brain in middle childhood is not yet as specialized for social processing as it is in adulthood, there is evidence that children can use the social interactions they observe to make informed social decisions. How does attention support these processes in childhood? How does attention to social interactions change across development?

The general purpose of this work was to explore the development of attentional orienting to social information in complex naturalistic scenarios and investigate the role of social interactions in those processes. In particular, the research questions addressed across the three empirical chapters were as it follows:

1. How does attention to complex naturalistic social scenes change across middle childhood, and does it differ from how social attention operates in young adulthood?
2. Does the social attentional bias change across pre-adolescent childhood, and does it differ between children and adults?
3. What is the role of social interactions in the way naturalistic scenes are attended to?
4. Does attention to social interactions change across development and is it any different in adulthood?

5. How do social interactions compete for attention with other social targets across development?
6. Does an individual's social knowledge influence their attention to social information, and does the extent of this influence change across pre-adolescent childhood, and between childhood and adulthood?

In this chapter I will summarise the findings from each empirical chapter and then provide a synthesis of the findings across the three chapters. This will be followed by a discussion of the main findings around the social attentional bias across both age groups, and the conditions in which an attentional bias to social interactions can occur, as well as the implications of these findings for human vision and development. The potential implications of this research for scene exploration research across development and social development will then be discussed. Finally, I will discuss some of the strengths and limitations of the research presented here, and suggest potential future lines of research and open questions, to then conclude with a brief summary.

1. Summary of findings

In study I (chapter 3) the aim was to investigate developmental changes in attention to naturalistic social scenes as well as the influence of social interactions on attentional orienting to social information. We found a consistent social bias in attentional engagement and capture in complex scenes, and that this bias towards social information increases in the presence of a social interaction in a similar way in both children and adults. Additionally, we found that the youngest children are in general slower to orient to all scenes as well as being more engaged by social information. In general, however, children and adults were similar in the extent to which they prioritised social information generally, and interactive information specifically.

In study II (chapter 4) we looked at social attention in scenes with multiple human agents (3 or 4), and at the attentional competition between social interactions and other social targets in the same scene. We find a strong social attentional bias in both children and adults, even in these more complex social scenes. Indeed, contrary to our expectations, we found no developmental changes in social attention across childhood or between children and adults, although children spent more time looking at the background compared to the adults. As in study II, we also find that children are slower than adults in orienting to the scenes. Exploratory analyses suggest that social interactions might be processed in a holistic way, and that – if true – this holistic processing occurs similarly in children and adults. As this finding was post-hoc and exploratory, however, it needs to be replicated in a follow-up study designed to test that hypothesis more directly. Importantly, when social interactions have to compete for attentional resources with other social targets in the same scene, they capture attention first and engage it for longer in both groups, although this effect is (somewhat) smaller in children and is reduced in effect size as the number of social targets increases.

In study III (chapter 5) we investigated developmental changes in attention to ambiguous social scenes, the role of pre-existing top-down social knowledge in orienting attention in ambiguous social scenes, and how the effect of this knowledge changed across childhood, and as compared to adulthood. Children, especially the youngest children, categorise fewer scenes as interactive than do adults. Additionally, as in both other studies, we found children to be slower to orient to the scenes compared to the adults. However, contrary to our expectations, interpreting a scene as interactive or non-interactive had no effect on how social attention was deployed for either adults or children. In other words, unlike our findings in scenes of unambiguous social interactions, how a scene was categorised did not influence either how quickly or for how long social information in the scene captured and engaged attention. Although our data demonstrate strong attentional capture by social information across age-groups and scenes, the social bias in attentional engagement was milder than was seen in non-ambiguous interactive scenes, with participants spending more time spent looking at the background.

2. Synthesis of findings across the three studies

Here I will synthesize the main findings as grouped by each research question, before proceeding to interpretation of the findings.

1. *How does attention to complex naturalistic social scenes change across middle childhood, and does it differ from how social attention operates in young adulthood?*

Consistently throughout the three studies, we find that children are in general slower to orient to the scenes compared to the adults, and this effect seems to be driven mainly by the youngest children. With age, children orient more quickly to both social and non-social information, potentially reflecting general improvement in processing skills. When the scene is especially complex – namely when it contains multiple people – children seem to look more at the background compared to the adults.

2. *Does the social attentional bias change across pre-adolescent childhood, and does it differ between children and adults?*

When a scene is socially unambiguous (Study I), the youngest children look for longer at social information compared to adults, a difference that slowly decreases with age. However, in more complex scenes, and in ambiguous scene, there were no developmental differences between groups in the attentional bias towards social information.

3. What is the role of social interactions in the way naturalistic scenes are attended to?

When only two people are in a scene and they are clearly interacting, looking times and attentional capture by human information (the social bias) are stronger, compared to when the two people are not interacting. Additionally, when an interacting dyad competes for attention with other social targets, the interacting dyad captures attention more quickly and holds it for longer. If the scene is socially ambiguous – i.e. if it is not clear whether the two individuals are having a social interaction or not – and interactive cues are missing or unclear, this modulation of attention is not seen, even if a participant categorises the scene as interactive.

4. Does attention to social interactions change across development and is it any different from adulthood?

Both children and adults are engaged in a similar way by social interactions, but only adults show this bias in attentional capture as well. This difference is likely the consequence of children being generally slower to orient their attention. Both children and adults may process interacting dyads holistically, and they both prioritize social interactions when they are competing for attention with other social targets, although this effect is weaker for children when more than one additional person is present in a scene.

5. How do social interactions compete for attention with other social targets in children and adults?

Both children and adults prioritize social interactions when they compete for attention, although this effect is weaker for 4-people pictures than for three, and slightly more variable in children than adults. Interactive information is prioritised by both groups, for both attentional capture and engagement, for 3-people pictures. This effect does not reach significance for children in 4-people pictures, and is not significant for attentional engagement for adults. When the two groups are merged, however, the prioritisation of social interaction reaches significance for both measures and both 3 and 4 person pictures, suggesting these developmental differences might be due to lack of statistical power to detect this effect rather than a true between-groups difference.

6. *Does an individual's social knowledge influence their attention to social information, and does the extent of this influence change across pre-adolescent childhood, and between childhood and adulthood?*

Adults are more likely to see ambiguous human dyads as interacting compared to children, and this difference is mainly driven by the youngest children (<9 years of age). Despite this difference in categorisation, there was both no influence of social knowledge/categorisation on the way social attention was deployed, and no developmental differences.

3. The current research within the Posner attentional framework

In the current work attention was conceptualized as a system subtending three main different functions as indicated by the tripartite model of attention (M. I. Posner & Petersen, 1990; Michael I Posner & Boies, 1971): alerting, orienting and executive control. In particular, the three experiments discussed here explored the development of the orienting of attention to social interactions in complex, naturalistic scenarios.

In this light, attention plays a role in filtering out information that is not needed, preparing individuals for action, action control and conflict monitoring (Rueda et al., 2015). In particular, the orienting system flexibly engages and shifts attention between different regions of visual space, selecting relevant information from the environment, based on both stimulus-driven factors (exogenous) and internal factors like goals and priorities (endogenous) (Colombo, 2001; M. I. Posner & Petersen, 1990).

Previous research on developmental changes in the three attentional systems has shown that while there are changes across childhood within the alerting and executive control systems, the orienting functions are qualitatively pretty stable between 6 and 10 years of age (Mullane et al., 2016; Rueda et al., 2004). Our data is consistent with this, as indicated by there being little-to-no evidence of developmental differences between middle childhood and adulthood in the way attention was engaged and captured by information across the three experiments. The main developmental change presented in this work, namely the generally slower capture by scenes in the children's group compared to the adults, is also consistent with developmental research in attentional orienting (Pozuelos et al., 2014; Rueda et al., 2004) and research findings that show slower saccade initiation in this age-group across stimuli and paradigms (Fukushima et al., 2000; Luna et al., 2008; Plude et al., 1994).

Another important aspect in the work presented here is the consistent finding of a social attentional bias across all three studies and that this bias was similar across the two age-groups, including a strong social interaction bias in study I. The ability to orient to relevant information in the environment has important consequences for development, from both a social and cognitive perspective (Amso & Scerif, 2015; Frischen et al., 2007; Soto-Icaza et al., 2015). Therefore, the early orienting to and capture of attention by social information and especially social interactions in a way similar to adults, provides strong evidence of the importance of observed social information for cognitive and social development as well as for social learning (see below for a more detailed interpretation of these results in terms of social development). Within this framework, the current results provide at least a partial answer to the question of how humans become so skilled at detecting and understanding interactive cues by adulthood. Indeed, the evidence this work provides of an orienting system already seemingly highly tuned to complex social situations strongly suggests that such tuning might have high developmental and biological value (Papeo, 2020; Quadflieg & Westmoreland, 2019).

Additionally, attention in the tripartite model of attention is not only conceptualized as a tool to regulate the selection of information and prioritize information based on its relevance at specific moments (M. I. Posner & Petersen, 1990), but also as a tool of action selection and preparation (M. I. Posner & Petersen, 1990; Michael I Posner, 2012; Rueda et al., 2015). Attention is especially important in development, for thoughts and actions regulation (Michael I Posner & Fan, 2008). Why is social information - especially pertaining to social interactions - so relevant for the exchange between individuals and their environment?

In the adult attentional system, the Posner model posits a balanced interplay between automatic attentional orienting, the selection of information through stimulus-based activation of the visual system (exogenous), and direction of attention to specific information because of its importance for future actions and goals (endogenous) (Henderson, 2003; Knudsen, 2007; Petersen & Posner, 2012; Rueda et al., 2015). Interestingly, in study III we don't have a clear indication of whether top-down factors influence the orienting of attention to social interactions, therefore it may be that the attentional bias to social interactions found in the other two studies might be more driven by stimulus-based cues to interaction rather than semantic knowledge or cognitive concepts around social interactions. However, the fact that the attentional bias becomes much weaker when a scene is socially ambiguous suggests that actually, when stimulus-based cues are not enough to interpret the scene, more contextual information is needed. This goal-oriented drive to further explore the scene suggests an involvement of top-down mechanisms that increase exploration rather than being strongly different between scenes that participants eventually decide are interactive vs. not. Interesting questions remain open around the interplay between exogenous cues and top-down factors that would deserve more future investigations.

Finally, this work investigated basic orienting abilities to complex social scenarios, however, as mentioned below in the "Future research and remaining questions", the next natural step would be to investigate whether, when, and how information about social interactions overrides the control systems of attention, but also, explore the strength of the orienting network, by probing its shifting and disengagement functions with social information. Indeed, orienting systems are involved in tasks such as information selection, shifting of attention, engagement and disengagement, and all these functions might be influenced by social information to different extents, which also provides material for future investigations.

4. Why is human information so important?

The most consistent result is the presence of a strong social attentional bias in all three studies and in both age-groups. This is unsurprising, considering that a great deal of past research has demonstrated that people show a preference to attend social information across a variety of stimuli and tasks (Bindemann et al., 2010; Doherty et al., 2017; Sue Fletcher-Watson et al., 2008; Rösler et al., 2017; Sasson & Touchstone, 2014; Van Der Geest et al., 2002). We do extend this research, however, by examining this social bias in childhood and demonstrating that mostly there are no substantial developmental changes in this preference, with the exception of when children view unambiguous two-person scenes. Indeed, our findings in this case are in contrast with prior research: we find that children spend more time looking at social information in unambiguous dyadic social scenes compared to the adults (independently of whether dyads are interacting, or not), while Amso (2014) shows an increase in engagement across childhood and into young adulthood. This contrast might be due to methodological differences as we consider the whole human AOI in the scene instead of just faces. Additionally, while Doherty et al. (2019) show greater capture by social information in children, we don't find this developmental difference. Again, however, this might reflect important methodological differences between paradigms: we measure spontaneous orienting in social attention through a free-viewing paradigm, while more executive functions are necessary in the task used by Doherty et al. (2019). In fact, the differences they report in speed of processing and executive functioning are mirrored in our finding of children being slower in general to orient to the scene, in line with research showing that processing speed and saccade initiation are still developing during childhood (Luna et al., 2008).

That social information is prioritised so strongly across a wide variety of scenes, in different social content and across levels of complexity and that this prioritisation is so strongly similar in children and adults makes a great deal of sense considering the evolutionary value of social information for human beings (Dunbar & Shultz, 2007). Indeed, starting from very early in life our attentional systems are tuned to social information in order to orient us to learn from other people (Gliga & Csibra, 2007; Papeo, 2020; Quadflieg & Westmoreland, 2019; Soto-Icaza et al., 2015). What is striking here is that this interest holds steady across childhood and doesn't differ from adults despite the fact that social brain structures (Mills et al., 2014; Walbrin et al., 2020) and attentional abilities (Pozuelos et al., 2014) are still developing, supporting the importance of attentional orientation to social information for observational learning in early childhood (Carpendale & Lewis, 2004; Eccles, 1999; Lee & Rutherford, 2018).

5. Why are social interactions so important?

We also find that two humans having a social interaction attract more attention than two people not interacting, both in scenes where they are the only social targets and when interactions are competing for attention with other social information. This is consistent with research showing that social interactions might be processed differently than two people acting independently (Papeo & Abassi, 2019; Vestner et al., 2019; Walbrin & Koldewyn, 2019) in several ways. First, we show that social interactions increase engagement of attention, and more importantly they do this across development with no developmental differences between children and adults. While the developing social brain is not yet fully tuned to social interactions (Walbrin et al., 2020), children are able to orient attention promptly to social interactions. Again, the lack of developmental change between children and adults suggests that the information collected during the observation of social interactions might be crucial in development of a child's "higher order" social skills, especially through facilitating learning from observed interactions (Brey & Shutts, 2015; Over & Carpenter, 2015; Skinner et al., 2017). Indeed, interacting dyads are a rich source of information about social norms, social cues, relationships, and personalities that can inform future social decisions and social learning (Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017; Quadflieg & Westmoreland, 2019; Skinner et al., 2017). The fact that the developing brain might be tuned to attend to this information preferentially supports the idea that social interactions have biological and evolutionary value.

One strong contribution of this work is that it investigated the conditions under which social interactions might attract attention or be processed differently. When social interactions are directly competing with other social targets (Study II), we find that they capture and hold attention more than competing social targets. However, while in the Study I we find a clear moderation of social attention by social interactions, in the most complex scenes in Study II this doesn't seem to be true, contrasting with prior literature suggesting the important role that social interactions play in visual attention in adults (Papeo, 2020; Papeo et al., 2017, 2019; Vestner et al., 2019) and in children (Stagg et al., 2014). One reason for these differences might be the complexity of the scenes and the diversity of the interactive cues that were present in our stimulus set. Additionally, exploratory post-hoc analyses suggested that this pattern of results may be the result of holistic processing of interacting agents where more attention is dedicated to the interpersonal space between interactive, but not non-interactive, individuals. These results are in line with previous research showing a holistic processing of interacting dyads compared to two non-interactive individuals (Papeo et al., 2019; Papeo & Abassi, 2019; Walbrin & Koldewyn, 2019). However, these results must be treated with caution as they were post-hoc and exploratory and, especially when investigated developmentally, lack sufficient power to be reliably interpreted.

A social hierarchy

Importantly, the investigation of social interactions in these scenes can also inform the processing of social information more broadly. Indeed, that the social interaction bias and social bias differ in strength across different scenes, suggests a potential hierarchy of social information in scenes. Indeed, we found attentional capture by human information across all conditions and participants, but the measure of attentional engagement was moderated more strongly by different conditions. This suggests an automatic mechanism of early attentional orienting to social information which can be only mildly modified by social interactive content. At later attentional stages, social information will hold attention for the longest when people in the scene are interacting, at last if the scene is not ambiguous and not too socially cluttered. If the scene becomes crowded with social targets, social interactions have to “fight” for attention with other social targets, While interactions *are* prioritised, this effect becomes smaller as more targets are added. Future studies should investigate to what extent and under what circumstances social interactions can win the fight for attentional resources. The work here presented suggests that when competing with one other human, interactions will capture and hold attention for longer, suggesting an attentional hierarchy of social information, where a single individual has attentional priority over objects and background elements but less priority than a social interaction between two individuals. When competing with two other humans, however, interactions will capture attention more quickly, but later engagement will be shared to a greater extent with non-interacting people in the same scene. Future studies could investigate this potential hierarchy of social information more carefully in order to disentangle the role of the dyads vs individuals, as well as the number of the people in the scene and the type of social content. Finally, when the scene is socially ambiguous, attention is shifted more to the background (as indicated by a milder social bias seen in engagement) after initially being captured by the social info, potentially suggesting the importance of context in interpreting a socially ambiguous situation. The fact that some of these effects in the attentional hierarchy are milder in childhood (namely the ability to hold attention weakens with the addition of two other targets in the scene), suggests that at least a part of these processes depend on the social experience of the child. Future studies should both investigate this hierarchy with more power in childhood and investigate how this hierarchy may change during adolescence.

6. Insights on scene exploration across development

The findings in this work consistently inform also on some aspects of scene processing, and especially on the developmental changes in these processes.

The fact that children were often slower to orient to scenes compared to adults is in line with research showing that saccade initiation is slower in childhood and continues developing until adulthood (Fukushima et al., 2000; Luna et al., 2008), as well as replicating prior studies that showed children to have slower performance in tasks investigating the orienting network (e..g. Pozuelos et al., 2014)

Our results from study III (Chapter 5) indicates that when a scene is hard to interpret or ambiguous socially, contextual information is crucial. This is in line with research on scene perception (Henderson & Hayes, 2018; Henderson & Hollingworth, 1999; Oliva et al., 2003; Torralba et al., 2006) suggesting that context is fundamental to understand relations between objects in the scene. In the case of our study, these 'objects' are likely the humans depicted in the scene, and participants may have looked more at background elements in order to disambiguate whether people were interacting, or not. This finding also again highlight the importance of using naturalistic stimuli to investigate these processes, especially in childhood.

Indeed, in study II (Chapter 4) involving the exploration of multiple people scenes, the fact that children look for longer at background information compared to the adults might either suggest that children need more contextual information when a scene is too complex, or simply they have not yet fully developed the ability to filter out irrelevant info, which is in line with prior research on attention in development (Amso & Scerif, 2015; Federico et al., 2017; Pozuelos et al., 2014). However, it could also reflect children's natural viewing behaviour (Açık et al., 2010), where they are more likely to explore *all* of the scene compared to adults.

7. Insights on social understanding across development

Study III greatly informs on the developmental differences in what is considered to be a social interaction or at least ambiguous social scenes interpretation.

The fact that in Study III we find that adults are more biased to interpret an ambiguous dyad as interacting compared to children, and especially children that are younger than 9 years, suggests either that a bias to see dyads as interactive is gained through social experience or that there are developmental differences in the concept itself (i.e., the internal representation of social interactions). Research suggests that by age 6 children are already equipped with a fair amount of social knowledge that gives them good insight into social situations and the ability to learn from such scenarios (Brey & Shutts, 2015; Carpendale & Lewis, 2004; Over & Carpenter, 2015; Soto-Icaza et al., 2015). However, the results in Study III suggest that higher order social information processing is still developing and being tuning in middle childhood.

8. Strengths and limitations

The strengths of the research presented here are mainly in the novelty of the aims, and in the use of naturalistic stimuli to investigate attention to social interactions in complex, close-to-real-life scenarios. Additionally, we investigated these processes across development in an area that has not yet received much scientific interest. Thus, this works makes a strong contribution to both social interaction research and, more generally, to the developmental social attention literature.

The use of naturalistic stimuli contributes to the ongoing debate about which factors could drive attentional orienting towards social interactions, namely the disconnect between research suggesting that the attentional bias to interacting dyads is due to the added biological value of social interactions (e.g. Abassi & Papeo, 2020; Papeo, 2020; Papeo et al., 2019) and other research suggesting an attentional cueing account, where social interactions drive attentional capture and engagement by creating an attentional “hot-spot” because of the converging attentional cues generated by the facing direction of the agents (e. g. Vestner et al., 2020; Vestner, Gray, et al., 2021; Vestner, Over, et al., 2021). Interaction is not cued by facing direction alone in our stimulus set. Indeed, our social scenarios are too varied and heterogeneous and facing direction is not always the strongest cue to whether two people are interacting or not. Indeed, often the interacting agents don’t directly face or look at each other. This suggests that the attentional cueing account is not the *only* factor to influence attentional priority given to interactions, though attentional cues from facing and eye-gaze direction are likely influential in at least some scenarios.

Although the use of heterogeneous, naturalistic scenes was a strength of this set of studies, the choice of stimuli also had some negative consequences. Variability in the size of AOIs, the distance between agents, and the complexity of backgrounds all likely contributed to variability in how attention was allocated across scenes. Although interactive and non-interactive scenes usually did not differ substantially on important measures like AOI size, they were not always matched in terms of the variance in such measures. Even small systematic differences in some visual aspects of the scenes could change attention. Indeed, we have seen that the size of the AOIs sometimes can drive early processes of attention (Chapter 4).

The developmental results might be more strongly supported by a higher sample size, as originally planned for in pre-registered power analyses. This was not possible in the current set of studies due to the interruption of data collection by the ongoing COVID-19 pandemic. Thus, where we found no developmental difference between groups, we cannot be certain of a true null effect.

9. Future research and remaining questions

While the research presented here has added to our understanding of attentional orienting to social interactions across development, there are open questions left to address to fully understand these mechanisms.

First, our stimuli were purposefully chosen to be emotionally neutral. What role might emotional content play in moderating attention to social interactions?

Secondly, we found developmental differences in the way ambiguous social scenes were categorized. How might the categorisation of ambiguous scenes (as interactive or not) change during adolescence? What are the factors that might influence this development – e.g. amount of social exposure, size of social network, development of social semantic cognition? Are there individual differences in social knowledge that might increase (or decrease) the bias to see dyads as interactive?

Perceptually, it is different to explore a scene depicting two, three, or four people and the amount of 'social content' is unlikely to scale linearly with each additional agent. What is the role of the social hierarchy in overriding attentional control limits, and especially, under different cognitive load manipulations (e.g Lavie, 1995, 2010; Lavie et al., 2014). How good is a single face, body, or human figure compared to an interacting dyad – or a non-interacting dyad – in breaking through attentional control under different cognitive load conditions?

Lastly, we don't find an effect of pre-existing social interaction knowledge on how attention is allocated in ambiguous social scenes. Would this be different if such scenes were investigated through priming methods (e.g. Ristic & Kingstone, 2005), specifically through priming participants to interpret the same scene as interactive or not?

10. Conclusion

Results across three free exploration experiments show that children and adults are extremely similar in the way they exhibit an attentional bias to social information across a variety of naturalistic social scenes. What's more, this research demonstrates the importance of social interactions across childhood, by showing that children have increased visual sensitivity to interacting individuals in a similar way to adults. Both children and adults show increased attention to humans when they are interacting, and prioritise them in a scene with multiple people. These processes don't seem to depend, at least strongly, on the amount of social information, although children differ from adults in this respect. Altogether, these results suggest a social hierarchy of attention, where interactions might be sitting at the top, as that this hierarchy is already in place in childhood. Future investigations should extend these findings to development during adolescence, and to investigation of this social hierarchy and social interactions when other components of social attention are taxed, like attentional control.

References

Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience*, *40*(4), 852–863. https://doi.org/10.1523/JNEUROSCI.1378-19.2019

Abdel Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin & Review*, *15*(6), 1055–1063. https://doi.org/10.3758/PBR.15.6.1055

Açık, 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 . In *Frontiers in Psychology* (Vol. 1, p. 207). https://www.frontiersin.org/article/10.3389/fpsyg.2010.00207

Adibpour, P., Hochmann, J.-R., & Papeo, L. (2021). Spatial Relations Trigger Visual Binding of People. *Journal of Cognitive Neuroscience*, *33*(7), 1343–1353. https://doi.org/10.1162/jocn_a_01724

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*(7), 267–278. https://doi.org/10.1016/S1364-6613(00)01501-1

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PLoS ONE*, *9*(1), 1–7. https://doi.org/10.1371/journal.pone.0085701

Amso, D., & Scerif, G. (2015). The attentive brain: Insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience*, *16*(10), 606–619. https://doi.org/10.1038/nrn4025

Astle, D. E., & Scerif, G. (2009). Using developmental cognitive neuroscience to study behavioral and attentional control. *Developmental Psychobiology*, *51*(2), 107–118. https://doi.org/10.1002/dev.20350

ATKINSON, J., & BRADDICK, O. (2012). Visual attention in the first years: typical development and developmental disorders. *Developmental Medicine & Child Neurology*, *54*(7), 589–595. https://doi.org/https://doi.org/10.1111/j.1469-8749.2012.04294.x

Augusti, E. M., Melinder, A., & Gredebäck, G. (2010). Look who's talking: Pre-verbal infants' perception of face-to-face and back-to-back social interactions. *Frontiers in Psychology*, *1*(OCT), 1–7. https://doi.org/10.3389/fpsyg.2010.00161

Bente, G., Krämer, N. C., Petersen, A., & de Ruiter, J. P. (2001). Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *Journal of Nonverbal Behavior*, *25*(3), 151–166. https://doi.org/10.1023/A:1010690525717

Bertenthal, B. I., & Boyer, T. W. (2015). *The Development of Social Attention in Human Infants BT - The Many Faces of Social Attention: Behavioral and Neural Measures* (A. Puce & B. I. Bertenthal (eds.); pp. 21–65). Springer International Publishing. https://doi.org/10.1007/978-3-319-21368-2_2

Bindemann, M., Scheepers, C., Ferguson, H. J., & Burton, A. M. (2010). Face, Body, and Center of Gravity Mediate Person Detection in Natural Scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1477–1485. https://doi.org/10.1037/a0019057

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *Quarterly Journal of Experimental Psychology*, *61*(7), 986–998. https://doi.org/10.1080/17470210701410375

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009a). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, *49*(24), 2992–3000. https://doi.org/10.1016/j.visres.2009.09.014

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009b). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, *17*(6–7), 904–924. https://doi.org/10.1080/13506280902758044

Birmingham, E., & Kingstone, A. (2009). Human social attention. In *Progress in Brain Research* (Vol. 176, Issue 626). Elsevier. https://doi.org/10.1016/S0079-6123(09)17618-5

Blakemore, S.-J. (2017). Social-Cognitive Development During Adolescence. In *Child Psychology and Psychiatry* (pp. 91–95). https://doi.org/https://doi.org/10.1002/9781119170235.ch11

Brey, E., & Shutts, K. (2015). Children Use Nonverbal Cues to Make Inferences About Social Power. *Child Development*, *86*(1), 276–286. https://doi.org/https://doi.org/10.1111/cdev.12334

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

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(1), 79–96. https://doi.org/DOI: 10.1017/S0140525X04000032

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/https://doi.org/10.1016/j.visres.2011.04.012

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 34, Issue 3, pp. 660–675). American Psychological Association. https://doi.org/10.1037/0096-1523.34.3.660

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage*, *12*(3), 314–325. https://doi.org/https://doi.org/10.1006/nimg.2000.0612

Centelles, L., Assaiante, C., Etchegoyhen, K., Bouvard, M., & Schmitz, C. (2013). From action to interaction: Exploring the contribution of body motion cues to social understanding in typical development and in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(5), 1140–1150. https://doi.org/10.1007/s10803-012-1655-0

Centelles, L., Assaiante, C., Nazarian, B., Anton, J. L., & Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0015749

Chevallier, C., Parish-Morris, J., McVey, A., Rump, K. M., Sasson, N. J., Herrington, J. D., & Schultz, R. T. (2015). Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Research*, *8*(5), 620–628. https://doi.org/https://doi.org/10.1002/aur.1479

Chevallier, C., Parish-morris, J., Mcvey, A., Rump, K., Noah, J., Herrington, J., & Schultz, R. T. (2016). Disorder using eye-tracking : Stimulus type matters. *Autism Research*, *8*(5), 620–628. https://doi.org/10.1002/aur.1479.Measuring

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, *48*, 79–93.

https://doi.org/10.1016/j.ridd.2015.10.011

Choi, W., Chao, Y.-W., Pantofaru, C., & Savarese, S. (2014). *Discovering Groups of People in Images BT - Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (eds.); pp. 417–433). Springer International Publishing.

Choudhury, S., Blakemore, S.-J., & Charman, T. (2006). Social cognitive development during adolescence. *Social Cognitive and Affective Neuroscience*, *1*(3), 165–174. https://doi.org/10.1093/scan/nsl024

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2010). A Taxonomy of External and Internal Attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Cloutier, J., & Gyurovski, I. (2014). Ventral medial prefrontal cortex and person evaluation: Forming impressions of others varying in financial and moral status. *NeuroImage*, *100*, 535–543. https://doi.org/https://doi.org/10.1016/j.neuroimage.2014.06.024

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, *21*(4), 843–860. https://doi.org/10.3758/s13423-013-0564-3

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Van Essen, D. C., & Shulman, G. L. (1998). A Common Network of Functional Areas for Attention and Eye Movements. *Neuron*, *21*(4), 761–773. https://doi.org/https://doi.org/10.1016/S0896-6273(00)80593-0

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science*, *13*(1), 120–133. https://doi.org/https://doi.org/10.1111/j.1467-7687.2009.00864.x

Crone, E. A. (2009). Executive functions in adolescence: inferences from brain and behavior. *Developmental Science*, *12*(6), 825–830.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*(6–7), 790–811. https://doi.org/10.1080/13506280902793843

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cerebral Cortex*, *25*(11), 4596–4609. https://doi.org/10.1093/cercor/bhv111

Deubel, H. (2008). The time course of presaccadic attention shifts. *Psychological Research*, *72*(6), 630. https://doi.org/10.1007/s00426-008-0165-3

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. https://doi.org/https://doi.org/10.1016/0042-6989(95)00294-4

Ding, X., Gao, Z., & Shen, M. (2017). Two Equals One: Two Human Actions During Social Interaction Are Grouped as One Unit in Working Memory. *Psychological Science*, *28*(9), 1311–1320. https://doi.org/10.1177/0956797617707318

Doherty, B. R., Fraser, A., Nobre, A. C., & Scerif, G. (2019). The functional consequences of social attention on memory precision and on memory-guided orienting in development. *Developmental Cognitive Neuroscience*, *36*(February). https://doi.org/10.1016/j.dcn.2019.100625

Doherty, B. R., Patai, E. Z., Duta, M., Nobre, A. C., & Scerif, G. (2017). The functional consequences of social distraction: Attention and memory for complex scenes. *Cognition*, *158*, 215–223. https://doi.org/https://doi.org/10.1016/j.cognition.2016.10.015

Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S. J., & Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, *389*(6651), 596–599. https://doi.org/10.1038/39309

Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93*(1), B27–B38. https://doi.org/https://doi.org/10.1016/j.cognition.2003.10.010

Downing, P. E., & Peelen, M. V. (2016). Body selectivity in occipitotemporal cortex: Causal evidence. *Neuropsychologia*, *83*, 138–148. https://doi.org/10.1016/j.neuropsychologia.2015.05.033

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag.

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*(5843), 1344–

1347. https://doi.org/10.1126/science.1145463

Eccles, J. S. (1999). The Development of Children Ages 6 to 14. *The Future of Children*, *9*(2), 30–44. https://doi.org/10.2307/1602703

Elison, J. T., Sasson, N. J., Turner-Brown, L. M., Dichter, G. S., & Bodfish, J. W. (2012). Age trends in visual exploration of social and nonsocial information in children with autism. *Research in Autism Spectrum Disorders*, *6*(2), 842–851. https://doi.org/10.1016/j.rasd.2011.11.005

End, A., & Gamer, M. (2017). Preferential Processing of Social Features and Their Interplay with Physical Saliency in Complex Naturalistic Scenes . In *Frontiers in Psychology* (Vol. 8, p. 418). https://www.frontiersin.org/article/10.3389/fpsyg.2017.00418

Erdfelder, E., FAul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Federico, F., Marotta, A., Martella, D., & Casagrande, M. (2017). Development in attention functions and social processing: Evidence from the Attention Network Test. *British Journal of Developmental Psychology*, *35*(2), 169–185. https://doi.org/https://doi.org/10.1111/bjdp.12154

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications.

Field, A. P., & Wright, D. B. (2011). A Primer on Using Multilevel Models in Clinical and Experimental Psychopathology Research. *Journal of Experimental Psychopathology*, *2*(2), 271–293. https://doi.org/10.5127/jep.013711

Flechsenhar, A. F., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PLoS ONE*, *12*(8), 1–20. https://doi.org/10.1371/journal.pone.0183799

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder.

Neuropsychologia, *47*(1), 248–257. https://doi.org/10.1016/j.neuropsychologia.2008.07.016

Fletcher-Watson, Sue, Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*(4), 571–583. https://doi.org/10.1068/p5705

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170. https://doi.org/10.1016/j.cognition.2008.11.010

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

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A Meta-Analysis of Gaze Differences to Social and Nonsocial Information Between Individuals With and Without Autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(7), 546–555. https://doi.org/10.1016/j.jaac.2017.05.005

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694.

Frith, C. D. (2007). The social brain? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 671–678. https://doi.org/10.1098/rstb.2006.2003

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements: I. Age-related changes in normal children. *Brain and Development*, *22*(3), 173–180. https://doi.org/https://doi.org/10.1016/S0387-7604(00)00101-7

Georgescu, A. L., Kuzmanovic, B., Santos, N. S., Tepest, R., Bente, G., Tittgemeyer, M., & Vogeley, K. (2014). Perceiving nonverbal behavior: Neural correlates of processing movement fluency and contingency in dyadic interactions. *Human Brain Mapping*, *35*(4), 1362–1378. https://doi.org/10.1002/hbm.22259

Gliga, T., & Csibra, G. (2007). Seeing the face through the eyes: a developmental perspective on face expertise. *Progress in Brain Research*, *164*(October), 323–339. https://doi.org/10.1016/S0079-6123(07)64018-7

Gobbini, M. I., Gors, J. D., Halchenko, Y. O., Hughes, H. C., & Cipolli, C. (2013). Processing of invisible social cues. *Consciousness and Cognition*, *22*(3), 765–770. https://doi.org/10.1016/j.concog.2013.05.002

Gray, K. L. H., Barber, L., Murphy, J., & Cook, R. (2017). Social interaction contexts bias the perceived expressions of interactants. *Emotion*, *17*(4), 567–571. https://doi.org/10.1037/emo0000257

Gray, K. L. H., Haffey, A., Mihaylova, H. L., & Chakrabarti, B. (2018). Lack of Privileged Access to Awareness for Rewarding Social Scenes in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(10), 3311–3318. https://doi.org/10.1007/s10803-018-3595-9

Greene, M. R., & Oliva, A. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science*, *20*(4), 464–472. https://doi.org/10.1111/j.1467-9280.2009.02316.x

Hafri, A., & Firestone, C. (2021). The Perception of Relations. *Trends in Cognitive Sciences*, *25*(6), 475–492. https://doi.org/https://doi.org/10.1016/j.tics.2021.01.006

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880–905. https://doi.org/10.1037/a0030045

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. https://doi.org/10.1038/nature06288

Handl, A., Mahlberg, T., Norling, S., & Gredebäck, G. (2013). Facing still faces: What visual cues affect infants' observations of others? *Infant Behavior and Development*, *36*(4), 583–586. https://doi.org/10.1016/j.infbeh.2013.06.001

Hanley, M., McPhillips, M., Mulhern, G., & Riby, D. M. (2013). Spontaneous attention to faces in Asperger syndrome using ecologically valid static stimuli. *Autism*, *17*(6), 754–761. https://doi.org/10.1177/1362361312456746

Hein, G., & Knight, R. T. (2008). Superior Temporal Sulcus—It's My Area: Or Is It? *Journal of Cognitive Neuroscience*, *20*(12), 2125–2136. https://doi.org/10.1162/jocn.2008.20148

Helo, A., van Ommen, S., Pannasch, S., Danteny-Dordoigne, L., & Rämä, P. (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behavior and Development*, *49*, 248–266. https://doi.org/https://doi.org/10.1016/j.infbeh.2017.09.008

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/https://doi.org/10.1016/j.tics.2003.09.006

Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, *18*(6), 10. https://doi.org/10.1167/18.6.10

Henderson, J. M., & Hollingworth, A. (1999). HIGH-LEVEL SCENE PERCEPTION. *Annual Review of Psychology*, *50*(1), 243–271. https://doi.org/10.1146/annurev.psych.50.1.243

Hessels, R. S., & Hooge, I. T. C. (2019). Eye tracking in developmental cognitive neuroscience – The good, the bad and the ugly. *Developmental Cognitive Neuroscience*, *40*, 100710. https://doi.org/https://doi.org/10.1016/j.dcn.2019.100710

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84. https://doi.org/10.1038/71152

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. https://doi.org/10.3758/BF03192848

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye Tracking : A Comprehensive Guide to Methods and Measures.* Oxford University Press. https://doi.org/http://ukcatalogue.oup.com/product/9780199697083.do

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*, *21*(3), 1167–1173. https://doi.org/10.1016/j.neuroimage.2003.11.013

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145 LP-E9152. https://doi.org/10.1073/pnas.1714471114

Itti, L. (2005). *CHAPTER 94 - Models of Bottom-up Attention and Saliency* (L. Itti, G. Rees, & J. K. B. T.-N. of A. Tsotsos (eds.); pp. 576–582). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012375731-9/50098-7

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506. https://doi.org/https://doi.org/10.1016/S0042-6989(99)00163-7

Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, *68*(November 2016), 269–297. https://doi.org/10.1146/annurev-psych-010416-044242

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*(7), 475–483. https://doi.org/10.1038/35081509

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. https://doi.org/https://doi.org/10.1016/j.visres.2007.09.013

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. In *Psychological Review* (Vol. 87, Issue 4, pp. 329–354). American Psychological Association. https://doi.org/10.1037/0033-295X.87.4.329

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. In *Psychological Bulletin* (Vol. 109, Issue 3, pp. 490–501). American Psychological Association. https://doi.org/10.1037/0033-2909.109.3.490

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience*, *1*(2), 91–100. https://doi.org/10.1038/35039043

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review*, *27*(3), 283–348. https://doi.org/https://doi.org/10.1016/j.dr.2007.06.006

Kaspar, K. (2013). What Guides Visual Overt Attention under Natural Conditions? Past and Future Research. *ISRN Neuroscience*, *2013*, 1–8. https://doi.org/10.1155/2013/868491

Kirby, L. A., Moraczewski, D., Warnell, K., Velnoskey, K., & Redcay, E. (2018). Social network size relates to developmental neural sensitivity to biological motion. *Developmental Cognitive Neuroscience*, *30*(March), 169–177. https://doi.org/10.1016/j.dcn.2018.02.012

Kirkorian, H. L., Anderson, D. R., & Keen, R. (2012). Age Differences in Online Processing of

Video: An Eye Movement Study. *Child Development*, *83*(2), 497–507. https://doi.org/https://doi.org/10.1111/j.1467-8624.2011.01719.x

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809

Knudsen, E. I. (2007). Fundamental Components of Attention. *Annual Review of Neuroscience*, *30*(1), 57–78. https://doi.org/10.1146/annurev.neuro.30.051606.094256

Konrad, K., Neufang, S., Thiel, C. M., Specht, K., Hanisch, C., Fan, J., Herpertz-Dahlmann, B., & Fink, G. R. (2005). Development of attentional networks: An fMRI study with children and adults. *NeuroImage*, *28*(2), 429–439. https://doi.org/https://doi.org/10.1016/j.neuroimage.2005.06.065

Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, A. (2017). Emotion Recognition in Context. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1667–1675.

Kujala, M. V, Carlson, S., & Hari, R. (2012). Engagement of amygdala in third-person view of face-to-face interaction. *Human Brain Mapping*, *33*(8), 1753–1762. https://doi.org/https://doi.org/10.1002/hbm.21317

Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*(1), 330–342. https://doi.org/https://doi.org/10.1016/j.cognition.2007.07.012

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59. https://doi.org/https://doi.org/10.1016/S1364-6613(99)01436-9

Lavie, N. (1995). Perceptual Load as a Necessary Condition for Selective Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*(3), 143–148. https://doi.org/10.1177/0963721410370295

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B:*

Biological Sciences, *369*(1641). https://doi.org/10.1098/rstb.2013.0205

Lee, V., & Rutherford, M. D. (2018). Sixteen-month-old infants are sensitive to competence in third-party observational learning. *Infant Behavior and Development*, *52*, 114–120. https://doi.org/https://doi.org/10.1016/j.infbeh.2018.07.001

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14. https://doi.org/https://doi.org/10.1016/S1364-6613(99)01418-7

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

Ma, Y., Paterson, H. M., & Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods*, *38*(1), 134–141. https://doi.org/10.3758/BF03192758

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0014594

Mayer, K. M., Vuong, Q. C., & Thornton, I. M. (2015). Do people “pop out”? *PLoS ONE*, *10*(10), 1–15. https://doi.org/10.1371/journal.pone.0139618

Mihai, I., Skripkauskaite, S., & Koldewyn, K. (2020). Development of the attentional priority of social interactions in scenes. *On-Line Experimental Psychology Society Meeting*, *poster*. https://osf.io/qs3kx/

Miller, L. T., & Vernon, P. A. (1997). Developmental changes in speed of information processing in young children. In *Developmental Psychology* (Vol. 33, Issue 3, pp. 549–554). American Psychological Association. https://doi.org/10.1037/0012-1649.33.3.549

Mills, K. L., Lalonde, F., Clasen, L. S., Giedd, J. N., & Blakemore, S.-J. (2014). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience*, *9*(1), 123–131. https://doi.org/10.1093/scan/nss113

Mondloch, C. J., Geldart, S., Maurer, D., & Grand, R. Le. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology*, *86*(1), 67–84. https://doi.org/https://doi.org/10.1016/S0022-0965(03)00102-4

Mullane, J. C., Lawrence, M. A., Corkum, P. V, Klein, R. M., & McLaughlin, E. N. (2016). The development of and interaction among alerting, orienting, and executive attention in children. *Child Neuropsychology*, *22*(2), 155–176. https://doi.org/10.1080/09297049.2014.981252

Nakayama, K. (2011). Introduction: Vision Going Social. *The Science of Social Vision*, 3–17.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*(9), 1186–1192. https://doi.org/10.1038/nn1759

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, *395*(6705), 894–896. https://doi.org/10.1038/27661

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, *13*(3), 135–143. https://doi.org/https://doi.org/10.1016/j.tics.2008.12.006

Nurmsoo, E., Einav, S., & Hood, B. M. (2012). Best friends: children use mutual gaze to identify friendships in others. *Developmental Science*, *15*(3), 417–425. https://doi.org/https://doi.org/10.1111/j.1467-7687.2012.01143.x

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, *1*, I–253. https://doi.org/10.1109/ICIP.2003.1246946

Over, H., & Carpenter, M. (2015). Children infer affiliative and status relations from watching others imitate. *Developmental Science*, *18*(6), 917–925. https://doi.org/https://doi.org/10.1111/desc.12275

Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B., & Lagopoulos, J. (2014). A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience*, *9*(6), 610–632. https://doi.org/10.1080/17470919.2014.934966

Papeo, L. (2020). Twos in human visual perception. *Cortex*, *132*(xxxx), 473–478. https://doi.org/10.1016/j.cortex.2020.06.005

Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human-human interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(7), 877–888. https://doi.org/10.1037/xhp0000646

Papeo, L., Goupil, N., & Hochmann, J.-R. (2020). Visual perception grounding of social cognition in preverbal infants. *OSF Preprints*. https://doi.org/https://doi.org/10.31219/osf.io/v7y9x

Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual Search for People Among People. *Psychological Science*, *30*(10), 1483–1496. https://doi.org/10.1177/0956797619867295

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychological Science*, *28*(3), 369–379. https://doi.org/10.1177/0956797616685769

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception and Psychophysics*, *62*(5), 889–899. https://doi.org/10.3758/BF03212075

Peelen, M. V, Glaser, B., Vuilleumier, P., & Eliez, S. (2009). Differential development of selectivity for faces and bodies in the fusiform gyrus. *Developmental Science*, *12*(6), F16–F25. https://doi.org/https://doi.org/10.1111/j.1467-7687.2009.00916.x

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*(1), 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, *68*(1), 101–109. https://doi.org/10.1001/archgenpsychiatry.2010.113

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2016). R Core Team (2015) nlme: linear and nonlinear mixed effects models. R package version 3.1–122. *URL: Http://CRANR-Projectorg/Package= Nlme*.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*(2), 227–272. https://doi.org/https://doi.org/10.1016/0001-6918(94)90004-3

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Posner, Michael I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, *78*(5), 391–408. https://doi.org/10.1037/h0031333

Posner, Michael I. (2012). Attention in a social world. In *Attention in a social world.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199791217.001.0001

Posner, Michael I, & Boies, S. J. (1971). Components of attention. *Psychological Review*, *78*(5), 391–408. https://doi.org/10.1037/h0031333

Posner, Michael I, & Fan, J. (2008). Attention as an organ system. *Topics in Integrative Neuroscience*, *31*.

Posner, Michael I, Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Pozuelos, J. P., Paz-Alonso, P. M., Castillo, A., Fuentes, L. J., & Rueda, M. R. (2014). Development of attention networks and their interactions in childhood. *Developmental Psychology*, *50*(10), 2405–2415. https://doi.org/10.1037/a0037469

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, *18*(6), 2188–2199. https://doi.org/10.1523/jneurosci.18-06-02188.1998

Quadflieg, S., Gentile, F., & Rossion, B. (2015). The neural basis of perceiving person interactions. *Cortex*, *70*, 5–20. https://doi.org/10.1016/j.cortex.2014.12.020

Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Annals of the New York Academy of Sciences*, *1396*, 166–182. https://doi.org/10.1111/nyas.13331

Quadflieg, S., & Penton-Voak, I. S. (2017). The Emerging Science of People-Watching: Forming Impressions From Third-Party Encounters. *Current Directions in Psychological Science*, *26*(4), 383–389. https://doi.org/10.1177/0963721417694353

Quadflieg, S., & Westmoreland, K. (2019). Making Sense of Other People's Encounters: Towards an Integrative Model of Relational Impression Formation. *Journal of Nonverbal Behavior*, *43*(2), 233–256. https://doi.org/10.1007/s10919-019-00295-1

Quattoni, A., & Torralba, A. (2009). Recognizing indoor scenes. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 413–420. https://doi.org/10.1109/CVPR.2009.5206537

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. In *Psychological Bulletin* (Vol. 124, Issue 3, pp. 372–422). American Psychological

Association. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., & Castelhano, M. S. (2008). Eye movements during reading, scene perception, visual search, and while looking at print advertisements. In *Visual marketing: From attention to action.* (pp. 9–42). Taylor & Francis Group/Lawrence Erlbaum Associates.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *32*(1), 123–142.

Reynolds, G. D., & Romano, A. C. (2016). The Development of Attention Systems and Working Memory in Infancy . In *Frontiers in Systems Neuroscience* (Vol. 10, p. 15). https://www.frontiersin.org/article/10.3389/fnsys.2016.00015

Reynolds, G. D., & Roth, K. C. (2018). The Development of Attentional Biases for Faces in Infancy: A Developmental Systems Perspective . In *Frontiers in Psychology* (Vol. 9, p. 222). https://www.frontiersin.org/article/10.3389/fpsyg.2018.00222

Riby, D. M., & Hancock, P. J. B. (2008). Viewing it differently: Social scene perception in Williams syndrome and Autism. *Neuropsychologia*, *46*(11), 2855–2860. https://doi.org/10.1016/j.neuropsychologia.2008.05.003

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*(1), 1–12. https://doi.org/10.1038/s41467-018-03399-2

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science: Real-World Social Attention and the Dual Function of Gaze. *Current Directions in Psychological Science*, *25*(1), 70–74. https://doi.org/10.1177/0963721415617806

Ristic, J., & Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition*, *94*(3), B55–B65. https://doi.org/https://doi.org/10.1016/j.cognition.2004.04.005

Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition*, *15*(3), 322–348. https://doi.org/10.1080/13506280600590434

Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLoS ONE*, *12*(7), 1–14. https://doi.org/10.1371/journal.pone.0182037

Ross, P. D., de Gelder, B., Crabbe, F., & Grosbras, M.-H. (2014). Body-selective areas in the visual cortex are less active in children than in adults. In *Frontiers in Human Neuroscience* (Vol. 8, p. 941). https://www.frontiersin.org/article/10.3389/fnhum.2014.00941

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, *12*(6), 852–877. https://doi.org/10.1080/13506280444000553

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2003.12.012

Rueda, M. R., Pozuelos, J. P., & Cómbita, L. M. (2015). Cognitive neuroscience of attention. *AIMS Neuroscience*, *2*(4), 183–202. https://doi.org/10.3934/Neuroscience.2015.4.183

Santos, N. S., Kuzmanovic, B., David, N., Rotarska-Jagiela, A., Eickhoff, S. B., Shah, J. N., Fink, G. R., Bente, G., & Vogeley, K. (2010). Animated brain: A functional neuroimaging study on animacy experience. *NeuroImage*, *53*(1), 291–302. https://doi.org/10.1016/j.neuroimage.2010.05.080

Sapey-Triomphe, L. A., Centelles, L., Roth, M., Fonlupt, P., Hénaff, M. A., Schmitz, C., & Assaiante, C. (2017). Deciphering human motion to discriminate social interactions: A developmental neuroimaging study. *Social Cognitive and Affective Neuroscience*, *12*(2), 340–351. https://doi.org/10.1093/scan/nsw117

Sasson, N. J., & Touchstone, E. W. (2014). Visual attention to competing social and object images by preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(3), 584–592. https://doi.org/10.1007/s10803-013-1910-z

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience*, *1*(3), 229–234. https://doi.org/10.1093/scan/nsl034

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain Regions for Perceiving and Reasoning About Other People in School-Aged Children. *Child Development*, *80*(4), 1197–1209. https://doi.org/https://doi.org/10.1111/j.1467-8624.2009.01325.x

Scerif, G. (2010). Attention trajectories, mechanisms and outcomes: at the interface between developing cognition and environment. *Developmental Science*, *13*(6), 805–812. https://doi.org/https://doi.org/10.1111/j.1467-7687.2010.01013.x

Scerif, G. (2020). Attention in Infancy and Childhood: A Focus on Developmental Dynamics. In *The Encyclopedia of Child and Adolescent Development* (pp. 1–12). https://doi.org/https://doi.org/10.1002/9781119171492.wecad114

Schultz, J., Friston, K. J., O'Doherty, J., Wolpert, D. M., & Frith, C. D. (2005). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron*, *45*(4), 625–635. https://doi.org/10.1016/j.neuron.2004.12.052

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van Der Gaag, C., Marois, R., & Skudlarski, P. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 415–427. https://doi.org/10.1098/rstb.2002.1208

Shaffer, R. C., Pedapati, E. V., Shic, F., Gaietto, K., Bowers, K., Wink, L. K., & Erickson, C. A. (2017). Brief Report: Diminished Gaze Preference for Dynamic Social Interaction Scenes in Youth with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *47*(2), 506–513. https://doi.org/10.1007/s10803-016-2975-2

Skinner, A. L., Meltzoff, A. N., & Olson, K. R. (2017). "Catching" Social Bias: Exposure to Biased Nonverbal Signals Creates Social Biases in Preschool Children. *Psychological Science*, *28*(2), 216–224. https://doi.org/10.1177/0956797616678930

Skripkauskaite, S., Mihai, I., & Koldewyn, K. (n.d.). Brief report: Attentional Bias towards Social Interaction during Viewing of Naturalistic Scenes. *Submitted*. https://doi.org/10.1101/2021.02.26.433078

Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive Ethology and exploring attention in real-world scenes. *Brain Research*, *1080*(1), 101–119. https://doi.org/https://doi.org/10.1016/j.brainres.2005.12.090

Soto-Icaza, P., Aboitiz, F., & Billeke, P. (2015). Development of social skills in children: Neural and behavioral evidence for the elaboration of cognitive models. *Frontiers in Neuroscience*, *9*(SEP), 1–16. https://doi.org/10.3389/fnins.2015.00333

Spunt, R. P., & Adolphs, R. (2015). Folk Explanations of Behavior: A Specialized Use of a Domain-General Mechanism. *Psychological Science*, *26*(6), 724–736. https://doi.org/10.1177/0956797615569002

SR Research, L. (2013). *Eyelink Data Viewer User's Manual* (p. No. 1.11.900).

Stagg, S. D., Linnell, K. J., & Heaton, P. (2014). Investigating eye movement patterns, language, and social ability in children with autism spectrum disorder. *Development and Psychopathology*, *26*(2), 529–537. https://doi.org/10.1017/S0954579414000108

Stoesz, B. M., & Jakobson, L. S. (2014). Developmental changes in attention to faces and bodies in static and dynamic scenes. *Frontiers in Psychology*, *5*(MAR), 1–9. https://doi.org/10.3389/fpsyg.2014.00193

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2016). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Su, J., Van Boxtel, J. J. A., & Lu, H. (2016). Social interactions receive priority to conscious perception. *PLoS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0160468

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics*. Pearson.

Taylor, M. J., Batty, M., & Itier, R. J. (2004). The Faces of Development: A Review of Early Face Processing over Childhood. *Journal of Cognitive Neuroscience*, *16*(8), 1426–1442. https://doi.org/10.1162/0898929042304732

Thompson, J., & Parasuraman, R. (2012). Attention, biological motion, and action recognition. *NeuroImage*, *59*(1), 4–13. https://doi.org/https://doi.org/10.1016/j.neuroimage.2011.05.044

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. In *Psychological Review* (Vol. 113, Issue 4, pp. 766–786). American Psychological Association. https://doi.org/10.1037/0033-295X.113.4.766

Twisk, J. W. R. (2006). Applied Multilevel Analysis: A Practical Guide for Medical Researchers. In *Practical Guides to Biostatistics and Epidemiology*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511610806

Van Der Geest, J. N., Kemner, C., Camfferman, G., Verbaten, M. N., & Van Engeland, H. (2002). Looking at Images with Human Figures: Comparison between Autistic and Normal Children. *Journal of Autism and Developmental Disorders*, *32*(2), 69–75.

https://doi.org/10.1023/A:1014832420206

Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, *200*, 104270. https://doi.org/10.1016/j.cognition.2020.104270

Vestner, T., Gray, K. L. H., & Cook, R. (2021). Visual search for facing and non-facing people: The effect of actor inversion. *Cognition*, *208*(December 2020). https://doi.org/10.1016/j.cognition.2020.104550

Vestner, T., Over, H., Gray, K., & Cook, R. (2021). Objects that direct visuospatial attention produce the search advantage for facing dyads. *Journal of Experimental Psychology: General*.

Vestner, T., Over, H., Gray, K. L. H., Tipper, S. P., & Cook, R. (2021). Searching for people: Non-facing distractor pairs hinder the visual search of social scenes more than facing distractor pairs. *Cognition*, *214*, 104737. https://doi.org/10.1016/j.cognition.2021.104737

Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S. A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*, *148*(7), 1251–1268. https://doi.org/10.1037/xge0000545

Villani, D., Morganti, F., Cipresso, P., Ruggi, S., Riva, G., & Gilli, G. (2015). Visual exploration patterns of human figures in action: An eye tracker study with art paintings. *Frontiers in Psychology*, *6*(OCT), 1–10. https://doi.org/10.3389/fpsyg.2015.01636

Wagner, D. D., Kelley, W. M., Haxby, J. V., & Heatherton, T. F. (2016). The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. *Journal of Neuroscience*, *36*(26), 6917–6925. https://doi.org/10.1523/JNEUROSCI.4220-15.2016

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*(February), 31–39. https://doi.org/10.1016/j.neuropsychologia.2018.02.023

Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, *198*(May), 296–302. https://doi.org/10.1016/j.neuroimage.2019.05.027

Walbrin, J., Mihai, I., Landsiedel, J., & Koldewyn, K. (2020). Developmental changes in visual responses to social interactions. *Developmental Cognitive Neuroscience*, *42*, 100774.

https://doi.org/https://doi.org/10.1016/j.dcn.2020.100774

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I See What You Mean: How Attentional Selection Is Shaped by Ascribing Intentions to Others. *PLoS ONE*, *7*(9), 1–7. https://doi.org/10.1371/journal.pone.0045391

Williams, T. A., Porter, M. A., & Langdon, R. (2013). Viewing social scenes: A visual scan-path study comparing fragile X syndrome and williams syndrome. *Journal of Autism and Developmental Disorders*, *43*(8), 1880–1894. https://doi.org/10.1007/s10803-012-1737-z

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970

Yang, D. Y.-J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews*, *51*, 263–275. https://doi.org/https://doi.org/10.1016/j.neubiorev.2015.01.020

Yarbus, A. (1967). *Eye movements and vision*. Springer, Boston, MA. https://doi.org/https://doi.org/10.1007/978-1-4899-5379-7

Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., & Boddaert, N. (2006). Autism, the superior temporal sulcus and social perception. *Trends in Neurosciences*, *29*(7), 359–366. https://doi.org/https://doi.org/10.1016/j.tins.2006.06.004

Appendix A – Pre-registrations

1. Pre-registration of "Attention to social interactions in naturalistic scenes" in Chapter 3 (#32797)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Does attentional social bias increase if the observed humans are part of a social interaction?

3) Describe the key dependent variable(s) specifying how they will be measured.
Eye-tracking data (time to first fixation and dwell time towards areas of interest in the scene (e.g., humans, background) will be collected during free-viewing of naturalistic scenes depicting social interactions or not.

4) How many and which conditions will participants be assigned to?
This is a within participant design with four repeated measures. Based on the ratings of 26 independent judges, 60 photographs depicting two agents have been selected and assigned to either the interactive (n = 30) or non-interactive (n = 30) condition. Two areas of interest (human and background AOIs) were further defined for each photograph.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
The time to first fixation data and dwell time data will be analysed separately using linear mixed-effect (multilevel) modelling in R (R Development Core Team, 2015), if applicable. To be more specific, a basic comparison of a model where the intercept is held constant and one where it is allowed to vary between participants will be conducted first to determine whether hierarchical structure should be taken into account.

If so, the eye-tracking data on interactive and non-interactive photographs will be analysed using a four-level hierarchical model with a 2x2 design. The participant information will be modelled at the highest (fourth) level. Nested within each participant, trial information with the social content of the scene (interacting or non-interacting) will be modelled as a third level predictor, whilst AOI type (human or background) as predictor will modelled at the second level. Finally, time to first fixation or dwell time per each AOI will be modelled at the first level, nested within each trial and participant.

If the initial comparison suggests that hierarchical modelling of random intercepts would not result in an increased model fit, two 2 (scene: interacting or non-interacting) x2 (AOI: human or background) repeated measures ANOVA would be conducted instead.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Participants who have failed the calibration procedure or otherwise did not produce accurate data (e.g. due to technical issues or sleepiness) will be fully excluded. Specific trials where participants have engaged with the target for less than 33% of the time will also be removed. If that results in the participant having less than 33% of usable trials, they will also be removed from any further analysis. Finally, if assumptions regarding a normal distribution of residuals are violated, the steps outlined by Tabachnick and Fidell (2007) to ensure the most efficient correction of distribution will be utilised.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about <u>exactly</u> how the number will be determined.

The recruitment will carry on till data is collected from 70 participants (not including sessions with technical issues). As there are no reliable power calculations for multilevel modelling, sample size was determined with an a priori power calculation based on ANOVA of the pilot dwell time data ($\eta2 = 0.68$). The power calculation was designed to reach 80% power ($\alpha \leq .05$) to detect a large effect size (Cohen's $f = .40$) of AOI and scene type on participants' time to first fixation or dwell time.

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Only pilot data has so far been collected, analysed, and presented in a poster form. Pilot data will not be included in the main analysis. The main data collection has been started, but no data has yet been extracted.

2. Pre-registration of "Development of attention to social interactions in naturalistic scenes" in Chapter 3 (#3821)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Does attentional bias to social interactions exist in childhood and adolescence? Does it change throughout development?

3) Describe the key dependent variable(s) specifying how they will be measured.
Eye-tracking data (time to first fixation and dwell time towards human and background) will be collected during free-viewing of naturalistic scenes depicting two agents.

4) How many and which conditions will participants be assigned to?
This is a within participant design with four repeated measures. Based on the ratings of 26 independent judges, 60 photographs depicting two agents have been selected and assigned to either the interactive (n = 30) or non-interactive (n = 30) condition. Two areas of interest (human and background AOIs) were defined for each photograph.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
The time to first fixation data and dwell time data will be analysed separately using multilevel modelling in R (R Development Core Team, 2015). The eye-tracking data on interactive and non-interactive photographs will be analysed using a four-level hierarchical model. The participant information will be modelled at the highest (fourth) level, where participants' age will be included as a predictor. Nested within each participant, trial information with the social content of the scene (interacting or non-interacting) will be modelled as a third level predictor, whilst AOI type (human or background) as predictor will modelled at the second level. Finally, time to first fixation or dwell time per each AOI will be modelled at the first level, nested within each trial and participant.
The participants' age at the time of data collection will be modelled either as a continuous, or a categorical variable. The decision will be made before the models are fitted based on the graphical visualisation of the data. If a potential linear relationship is observed, age will be used as a continuous variable. If a quadratic relationship between age and social interaction attention is suggested, instead, age will be used as a categorical variable.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Only participants who have failed the calibration procedure or otherwise did not produce accurate data (e.g. due technical issues or sleepiness) will be fully excluded. Otherwise, only specific trials where participants have engaged with the target for less than 33% of the time will be removed. Finally, if normal distribution of residuals were violated, the steps outlined by Tabachnick and Fidell (2007) to ensure the most efficient correction of distribution will be

utilised.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
The aim is to collect the data for at least 90 youths (< 18 years) in order to detect large effects (Cohen's $f = .40$, $\alpha \leq .05$, $1-\beta = 0.80$), but no more than the 222 children and adolescents necessary to detect a medium sized effect (Cohen's $f = .25$, $\alpha \leq .05$, $1-\beta = 0.80$). Data collection will cease at the end of the 2019/2020 academic year (July 21st), thus an exact number will depend on participants available by then.

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Questionnaire data from a measure of social communication difficulties will be collected for the adolescents in the sample. Therefore, sample size permitting, additional exploratory analysis predicting attentional engagement with and orienting to social interactions based on their questionnaire self-reports will be conducted.

3. Pre-registration of "Attentional competition of social interactions in naturalistic scenes" in Chapter 4 (#33540)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Is there an attentional bias towards social interactions in naturalistic scenes containing multiple people (i.e., more than two people)? If so, how is this attentional bias altered by increasing the attentional competition with other social stimuli by increasing the number of people in the scene?

3) Describe the key dependent variable(s) specifying how they will be measured.
Eye-tracking data (time to first fixation and dwell time in pre-defined human and background areas of interest) will be collected during a 5 second free-viewing of naturalistic scenes that either depict social interactions, or not. Scenes will contain 3 or 4 people.

4) How many and which conditions will participants be assigned to?
This is a whithin-subjects design where participants view all scenes. Based on the ratings of 26 independent judges, 30 photographs depicting three agents and 22 photographs depicting four agents have been selected and assigned in equal number to either the interactive or non-interactive condition. Thus, there are two relevant factors - number of people in the scene (3 or 4) and whether a two-person interaction is taking place in the scene (interactive or not).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Areas of interest (AOIs) have been defined for each picture. These include each human as separate AOIs and all background elements as a single AOI. For some analyses, we will combine all human AOIs into a single AOI (human). The time to first fixation data and dwell time data to these AOIs will be analysed separately using linear mixed-effect (multilevel) modelling in R (R Development Core Team, 2015), if applicable. To be more specific, a basic comparison of a model where the intercept is held constant and one where it is allowed to vary between participants will be conducted first to determine whether hierarchical structure should be taken into account.
If so, separate multilevel analyses will be conducted for each research question. For the first research question the model will have a 2 (scene: interacting or non interacting) x 2 (number of people in the picture: 3 or 4) x 2 (AOI: human or background) predictor structure. For the second research question only the interactive pictures will be analysed, and the model will have a 2 (number of people in the picture: 3 or 4) x 3 (AOI: interacting humans, non interacting humans and background) predictor structure.
If there's no variance at the participant level, ANOVAs with the same structure or predictors will be carried out.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Only participants who have failed the calibration procedure or otherwise did not produce accurate data (e.g. due to technical issues or sleepiness) will be fully excluded. Otherwise, only specific trials where participants have engaged with the target for less than 33% of the time will be removed. If that results in the participant having less than 33% of usable trials, they will also be removed from any further analysis. Finally, if assumptions regarding a normal distribution of residuals are violated, the steps outlined by Tabachnick and Fidell (2007) to ensure the most efficient correction of distribution will be utilised.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
Data collection will continue until 231 participants are reached or until 31st of July 2020, whichever comes first. Sample size was determined with a priori power calculation designed to reach 80% power ($\alpha \leq .05$) to detect a medium effect size (Cohen's $f = .25$) of AOI, scene type and people in the scene on participants' time to first fixation or dwell time.

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Only pilot data has so far been collected and analysed, and will not be included in the main analysis.

4. Pre-registration of " Development of the attentional priority of social interactions in naturalistic scenes" in Chapter 4 (#38336)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Does the attentional bias towards human information persist in naturalistic scenes containing more than two people, and is it moderated by the presence of interactive content in the scene? How is the attentional bias to social interactions, if any, altered by increasing the attentional competition with other social stimuli through the number of people in the scene? Furthermore, how do these attentional mechanisms change across development?

3) Describe the key dependent variable(s) specifying how they will be measured.
Eye-tracking data (time to first fixation and proportional dwell time in pre-defined human and background areas of interest - AOIs) are collected during a 5 second free-viewing of naturalistic scenes that either depict social interactions or not. Scenes will contain 3 or 4 people.

4) How many and which conditions will participants be assigned to?
This is a mixed design. All participants view all scenes. Based on the ratings of 26 independent judges, 30 photographs depicting three agents and 22 photographs depicting four agents have been selected and assigned in equal number to either the interactive or non-interactive condition. Thus, there are three relevant within-subject factors - number of people in the scene (3 or 4), whether a two-person interaction is taking place in the scene (interactive or not) and the AOI type (humans or background). The between-subject factor is age [adults (18-35) and developmental group (6-17 - see below)] (adults' data is discussed in AsPredicted#33540).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Separate multilevel modelling analyses will be conducted for each research question, if possible. The participant information will be modelled at the highest level where participants' age will be included as a predictor for both research questions. For both research questions the model will have other 2 predictors. For the first one these will be scene (interacting or non-interacting) and AOI (human or background), while for the second these will be AOI (interacting humans, non-interacting humans, within the interacting pictures) and number of people in the picture (3 or 4). If there is no variance at the participant level within each group, mixed ANOVAs with the same predictors will be carried out.
The participants' age at the time of data collection will be modelled either as a continuous or a categorical variable. The decision will be made based on the graphical visualisation of the data. If a linear relationship is observed, age will be used as a continuous variable. If a quadratic relationship between age and social interaction attention is suggested, age will be used as a categorical variable.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Only participants who have failed the calibration procedure or otherwise did not produce accurate data (e.g. due to technical issues or sleepiness) will be fully excluded. Otherwise, only specific trials where participants have engaged with the target for less than 33% of the time will be removed. If trial exclusion results in the participant having less than 33% usable trials, they will also be removed from any further analysis. Finally, if assumptions regarding a normal distribution of residuals are violated, the steps outlined by Tabachnick and Fidell (2007) to ensure the most efficient correction of distribution will be utilised.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
Data collection will continue until we have 90 participants in the developmental sample (for the adult sample power analysis, see AsPredicted#33540) in order to detect a large effect size (Cohen's $f = .40$) and reach 80% power ($\alpha \leq .05$), but we will not collect more than the 222 youths necessary to detect a medium effect size (Cohen's $f=.25$), or will cease at the end of the academic year, on 21st of July 2020 , whichever comes first.

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Data collection has been already started, but no data has yet been extracted and/or analysed.

5. Pre-registration of "Attention to ambiguous social scenes" in Chapter 5 (#32800)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Does gaze processing of ambiguous social scenes depend on whether they are perceived as depicting an interaction between two people or not?

3) Describe the key dependent variable(s) specifying how they will be measured.
Eye-tracking data (time to first fixation and dwell time) will be collected during free-viewing of naturalistic scenes depicting two human agents.

4) How many and which conditions will participants be assigned to?
This is a within participant design with 4 repeated measures. Based on the ratings of 26 independent judges, 30 photographs depicting two agents have been selected as ambiguous pictures. These pictures were chosen as they could not be categorised as clearly interactive or non-interactive. Each picture instead will be categorised by each participant as either interactive or not-interactive, in their own view. Two areas of interest (human and background AOIs) were further defined for each photograph.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
The time to first fixation data and dwell time data will be analysed separately using linear mixed-effect (multilevel) modelling in R, if applicable. A basic comparison of a model where the intercept is held constant and one where it is allowed to vary between participants will be conducted first to determine whether hierarchical structure should be taken into account.
If so, the eye-tracking data on ambiguous photographs will be analysed using a four-level hierarchical model with a 2x2 design. The participant information will be modelled at the highest (fourth) level. Nested within each participant, trial information with the participants' categorisation (interacting or non-interacting) will be modelled as a third level predictor, whilst AOI type (human or background) will modelled at the second level. Finally, raw data per each AOI will modelled at the first level, nested within each trial and participant.
If the initial comparison suggests that hierarchical modelling of random intercepts would not result in an increased model fit, two 2 (scene: interacting or non-interacting) x2 (AOI: human or background) repeated measures ANOVA would be conducted instead. Averages for interacting/non-interacting conditions would need to be calculated on person by person basis depending on their categorisation of the pictures.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Participants who did not produce accurate data (e.g. due to calibration, technical issues or sleepiness) will be fully excluded. Specific trials where participants have engaged with the target for less than 33% of the time will also be removed. If that results in the participant having less

than 33% of usable trials, they will also be removed from further analysis. Finally, if assumptions of normal distribution are violated, the steps outlined by Tabachnick and Fidell (2007) to ensure the most efficient correction of distribution will be utilised.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
Data will be collected from 70 participants. Sample size was determined with an a priori power calculation based on the pilot dwell time data showing increased attention to human figures in photographs judged as interactive by independent judges ($\eta2 = 0.68$). The power calculation was designed to reach 80% power ($\alpha \leq .05$) to detect a large effect size (Cohen's $f = .40$) of AOI and scene categorisation on participants' gaze behaviour.

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
The data collection has now started, but no data has yet been extracted or analysed.

6. Pre-registration of "Development of attention to ambiguous social scenes" in Chapter 5 (#43713)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Does the way participants allocate attention within ambiguous social scenes depend on whether the scene is perceived as containing a social interaction or not? Does the extent to which visual attention within a scene is influenced by the perceived social content change across development?

3) Describe the key dependent variable(s) specifying how they will be measured.
Eye-tracking measures (time to first fixation and dwell time in defined areas of interest – AOIs) collected during 5 seconds of free exploration of naturalistic scenes depicting two human agents.

4) How many and which conditions will participants be assigned to?
This is a mixed design with four repeated measures, and age as a between subjects factor. Based on the ratings of 26 independent judges, 30 photographs depicting two agents have been selected as ambiguous pictures. These pictures were chosen as they could not be categorised as clearly interactive or non-interactive. Each picture instead is categorised as interactive or non-interactive by each participant. Two areas of interest (human and background AOIs) were defined for each photograph. The between-subject factor is age [adults (18-35) and developmental group (6-17 - see below)] (adults' data is discussed in AsPredicted#32800).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
The time to first fixation data and dwell time data will be analysed separately using linear mixed-effect (multilevel) modelling in R, if applicable. The participant information will be modelled at the highest level where participants' age will be included as a predictor. Nested within each participant, trial information with the participants' categorisation (interacting or non-interacting) will be modelled as a third level predictor, whilst AOI type (human or background) will modelled at the second level. Finally, raw data per each AOI will modelled at the first level, nested within each trial and participant.
If there is no variance at the participant level within each group, mixed ANOVAs with the same predictors will be carried out.
The participants' age at the time of data collection will be modelled either as a continuous, or a categorical variable. The decision will be made before the models are fitted based on the graphical visualisation of the data. If a potential linear relationship is observed, age will be used as a continuous variable. If a quadratic relationship between age and social interaction attention is suggested, instead, age will be used as a categorical variable.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for

excluding observations.
Only participants who have failed the calibration procedure or otherwise did not produce accurate data (e.g. due to technical issues or sleepiness) will be fully excluded. Otherwise, only specific trials where participants have engaged with the target for less than 33% of the time will be removed. If trial exclusion results in the participant having less than 33% usable trials, they will also be removed from any further analysis. Finally, if assumptions regarding a normal distribution of residuals are violated, the steps outlined by Tabachnick and Fidell (2007) to ensure the most efficient correction of distribution will be utilised.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
Data collection qill continue until we have 90 participants in the developmental sample (for the adult power analysis, see AsPredicted#32800) in order to detect a large effect size (Cohen's f=.40) and reach 80% power ($\alpha \leq .05$).). Because of the Covid-19 pandemic, however, we are uncertain of our ability to recruit this sample in a reasonable time frame, and so may be forced to halt data collection before reaching 90 participants if we have been unable to collect a full sample before 31st December, 2020.

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Data collection has been already started, but no data has yet been extracted and/or analysed.

Appendix B – Stimuli

1. Stimuli used in Chapter 3

Figure S1. Stimuli set used in the interactive condition.

Figure S2. Stimuli set used in the non-interactive condition.

2. Stimuli used in Chapter 4

Figure S3. Stimuli set used in the interactive condition.

Figure S4. Stimuli set used in the non-interactive condition.

3. Stimuli used in Chapter 5

Figure S4. Stimuli set used in the ambiguous condition.

Appendix C – Supplementary materials for chapter 3

S1: AOI size information in pixels

Table S1a. Main effects and interactions for a 2 (type of scene) x 2 (type of AOI) mixed ANOVA on the AOI size in pixels.

Predictor	Df	*F* - value	*p*- value	$\eta^2{}_p$
Type of scene	1, 58	0.00	.99	< .001
AOI	1, 58	207.16	< .001	.78
scene * AOI	1, 116	7.99	.01	.06

Table S1b. Pairwise comparisons for the scene*AOI interaction, corrected with HSD Tukey for multiple comparisons.

Group	Contrast	Df	*t* - value	*p*- value	*d*
Interactive scenes	Social - Background	116	- 8.18	< .001	- 0.76
Non-Interactive scenes	Social - Background	116	- 12.18	< .001	- 1.13
Social AOI	Interactive – Non-interactive	116	1.99	.15	0.19
Background	Interactive – Non-interactive	116	- 2	.14	- 0.19

Table S1c. Descriptive statistics for AOI sizes across conditions.

Type of scene	AOI	Mean area (px)	SD
Interactive	Social	259379.99	104248.82
	Background	505413.42	104861.75
Non-interactive	Social	199358.06	127111.07

	Background	565631.96	127569.29

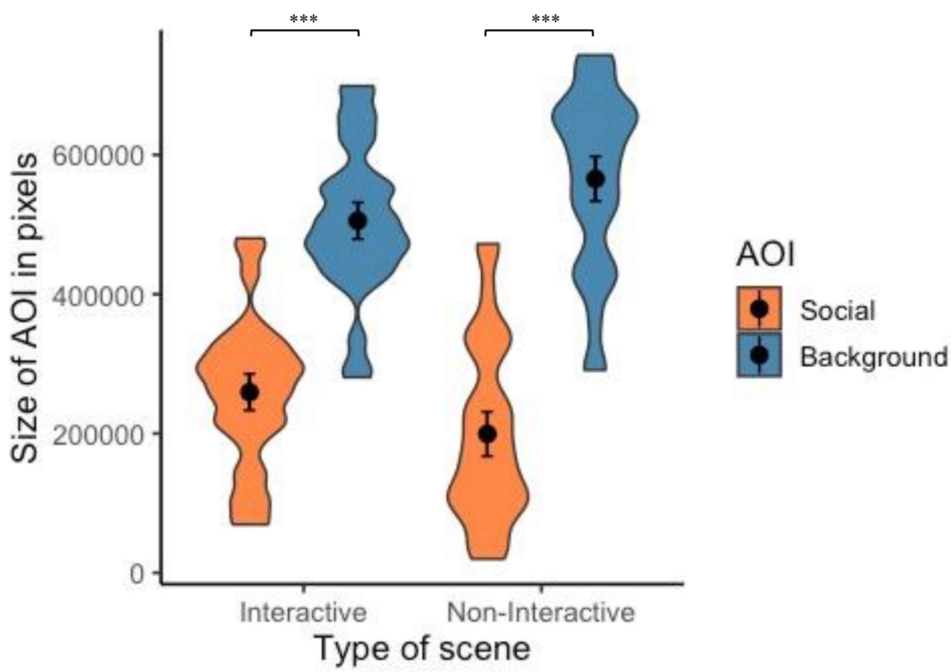

Figure S1a. Violin plots for mean size of AOIs (px) across interactive and non-interactive scenes. Error bars represent 95% confidence intervals. Only significant – at $p < .05$ level – comparisons shown.

S2: Data transformation information across the chapter

2a. Log10 transformation of the time to first fixation data in Experiment 1.
The Anderson-Darling test for normality ($A = 643.25$, $p < .001$) and visual inspection of residuals showed that this data was positively skewed, therefore a logarithm in base 10 transformation was applied to the data to reduce skewness. After transformation, however, the Anderson test for normality was still significant ($A = 141.24$, $p < .001$) although the transformation improved the graphical shape of the distribution.
In the model with the transformed data, results were similar to those found using the untransformed data, resulting in a main effect of AOI type but no main effect of type of scene, and an interaction between the two.

Participant, type of scene and AOI were set as random effects, $SD = 0.11$, $\chi^2(3) = 555.29$, $p < .001$, and when we added the size of AOIs as a random effect, the model did not change significantly, $SD = 0.31$, $\chi^2(3) = 0.00$, $p = .99$.

See Table S2a for main effects, Table S2b for descriptive statistics and Figure S2a for mean transformed time to first fixation across conditions.

Table S2a. Main effects and interactions in the model with a 2 (type of scene) * 2 (AOI) structure, for time to first fixation in experiment 1.

Predictor	numDF	denDF	F - value	p- value	η^2_p
Type of scene	1	69	0.0	.91	.00
AOI	1	138	559.8	< .001	.80
Type of scene*AOI	1	138	37.6	< .001	.21

Table S2b. Descriptive statistics for transformed time to first fixation to each AOI, in each condition in experiment 1.

Type of scene	AOI	Mean	SD
Interactive	Social	2.55	0.23
	Background	2.80	0.36
Non-interactive	Social	2.60	0.27
	Background	2.75	0.37

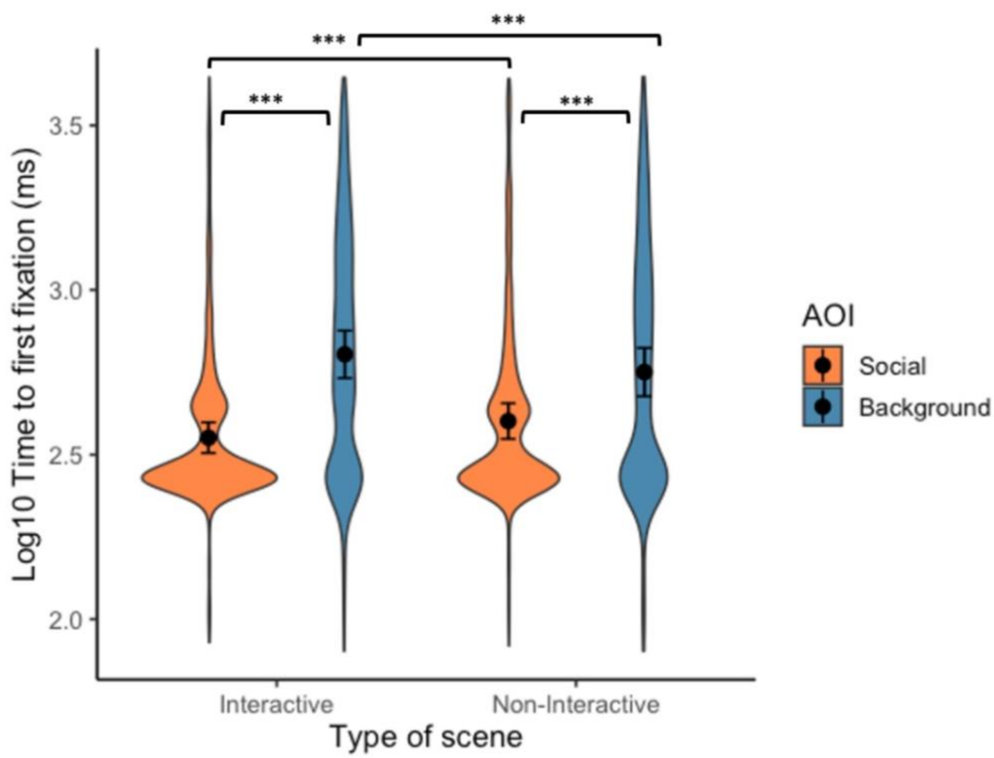

Figure S2a. Violin plots for transformed mean time to first fixation for human and background AOIs in interactive and non-interactive scenes. Error bars represent 95% confidence intervals.

2b. Log10 transformation of the time to first fixation data in Experiment 2.

The Anderson-Darling test for normality (A = 505.78, p < .001) and visual inspection of residuals showed that this data was positively skewed, therefore a logarithm in base 10 transformation was applied to the data to reduce skewness. After transformation, however, the Anderson test for normality was still significant (A = 129.78, p < .001) although the transformation improved the graphical shape of the distribution.

In the model with the transformed data, results were similar to those found using the untransformed data, resulting in a main effect of age and a main effect of AOI type but no main effect of type of scene, and an interaction between the age and AOI as well as an interaction between scene-type and AOI.

Participant, type of scene and AOI were set as random effects, SD = 0.11, $\chi^2(3)$ = 411.98, p < .001, and when we added the size of AOIs as a random effect, the model did not change significantly, SD = 0.30, $\chi^2(3)$ = 0.00, p = .99.

See Table S2c for main effects, Table S2d for descriptive statistics and Figures S2b, S2c and S2d for mean transformed time to first fixation across conditions and ages.

Table S2c. Main effects and interactions in the model with an age * 2 (type of scene) * 2 (AOI) structure, for time to first fixation in experiment 2.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
Age	1	52	5.35	.03	.09
Type of scene	1	52	0.05	.83	<.001
AOI	1	104	419.15	< .001	.80
Age*type of scene	1	52	0.16	.70	<.001
Age*AOI	1	104	7.96	< .001	.07
Type of scene*AOI	1	104	12.76	< .001	.11
Age*scene*AOI	1	104	0.25	.62	<.001

Table S2d. Descriptive statistics for transformed time to first fixation to each AOI, in each condition in experiment 2.

Type of scene	AOI	Mean	SD
Interactive	Social	2.60	0.24
	Background	2.82	0.35
Non-interactive	Social	2.63	0.26
	Background	2.79	0.35

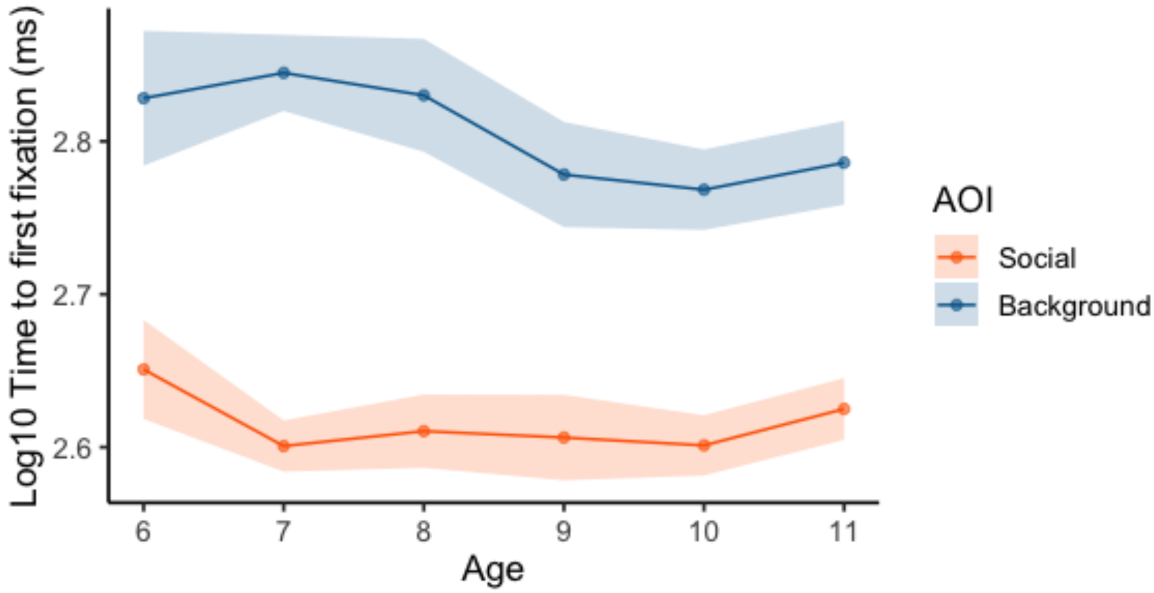

Figure S2b. Average transformed time to first fixation to social and background AOIs across scenes in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

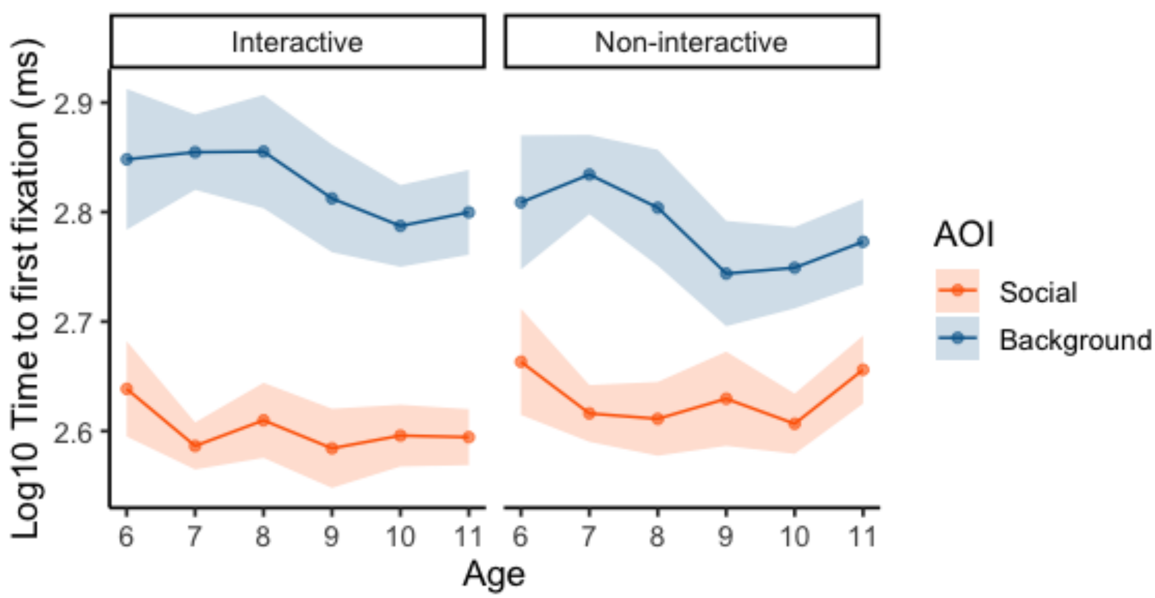

Figure S2c. Average transformed time to first fixation to social and background AOIs in interactive and non-interactive scenes in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

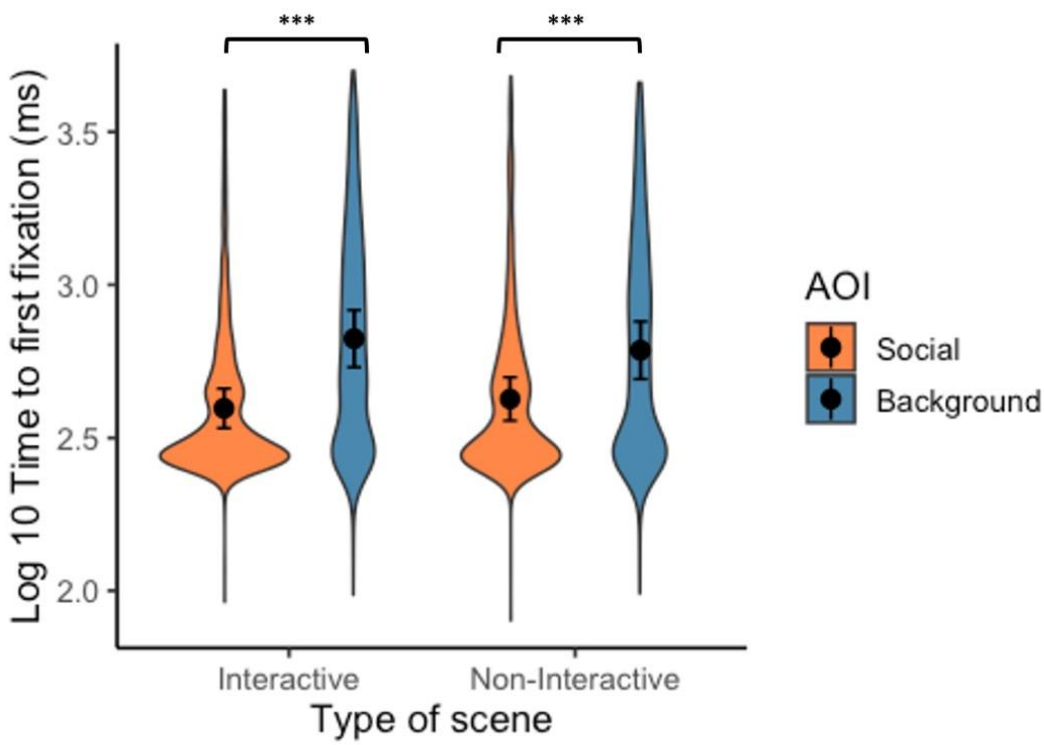

Figure S2d. Violin plots for transformed mean time to first fixation for human and background AOIs in interactive and non-interactive scenes in experiment 2. Error bars represent 95% confidence intervals.

2c. Log10 transformation of the time to first fixation data in analysis comparing children and adult.

The Anderson-Darling test for normality (A = 1144.7, $p < .001$) and visual inspection of residuals showed that this data was positively skewed, therefore a logarithm in base 10 transformation was applied to the data to reduce skewness. After transformation, however, the Anderson test for normality was still significant (A = 267.12, $p < .001$) although the transformation improved the graphical shape of the distribution.

In the model with the transformed data, results were similar to those found using the untransformed data, resulting in a main effect of age and a main effect of AOI type but no main effect of type of scene, an interaction between scene-type and AOI, and no moderation of AOI or scene by age.

See Table S2e for main effects, Table S2f for descriptive statistics and Figures S2e for mean transformed time to first fixation across conditions, AOIs and ages.

Table S2e. Main effects and interactions in the model with a 2 (age-group) * 2 (type of scene) * 2 (AOI) structure, for transformed time to first fixation in the developmental change analysis.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
Age	1	122	19.9	< .001	.14
Type of scene	1	122	0.1	.82	< .001
AOI	1	244	953.9	< .001	.80
Age*type of scene	1	122	0.0	.90	.00
Age*AOI	1	244	0.4	.55	< .001
Type of scene*AOI	1	244	47.9	< .001	.16
Age*scene*AOI	1	244	2	.16	.01

Table S2f. Descriptive statistics for transformed time to first fixation to each AOI, in each condition and age-group in the developmental change analysis.

Age-group	Type of scene	AOI	Mean	SD
Children	Interactive	Social	2.60	0.24
		Background	2.82	0.35
	Non-interactive	Social	2.63	0.26
		Background	2.79	0.35
Adults	Interactive	Social	2.55	0.23
		Background	2.80	0.36
	Non-interactive	Social	2.60	0.27
		Background	2.75	0.37

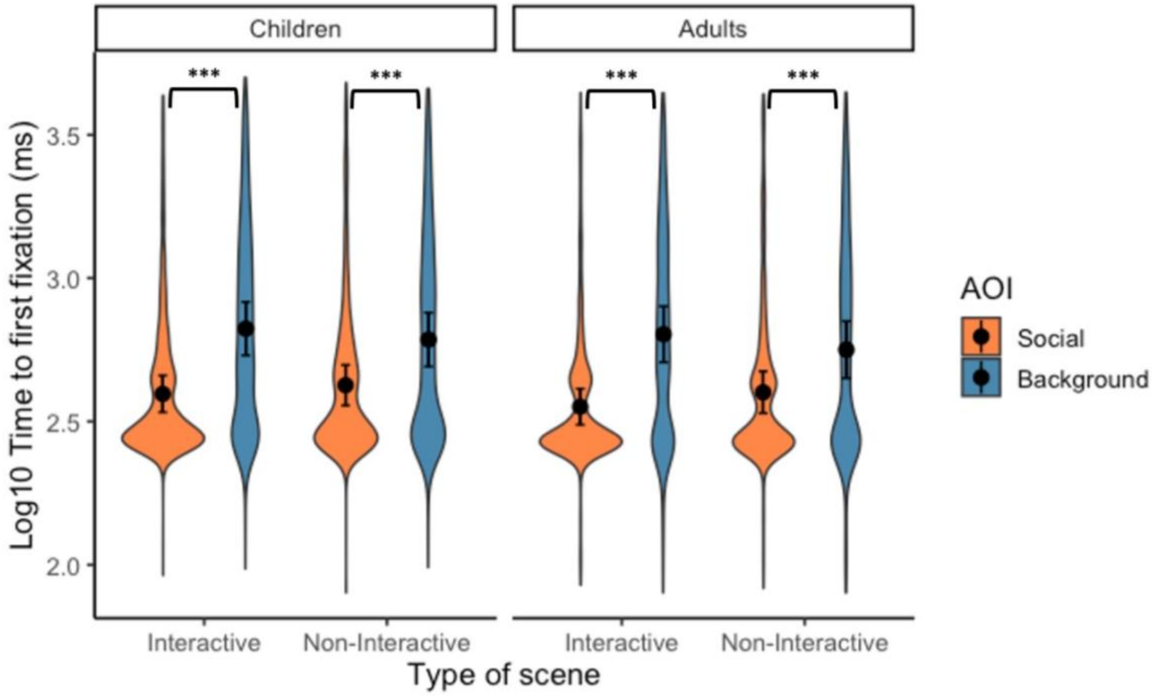

Figure S2e. Violin plots for transformed mean time to first fixation for human and background AOIs in interactive and non-interactive scenes, in the two age-groups. Error bars represent 95% confidence intervals.

S3: Missing trials modelling

3a. Models of the dwelling time and time to first fixation including the number of missing trials as a fixed effect in experiment 2.

For each of the two measures, we performed a model including age, type of scene, AOI, and missing trials as fixed effects, and participant, type of scene, and type of AOI as random effects. See Tables S3a and S3b for the main effects of the two models.

Table S3a. Main effects and interactions in the model with an age * 2 (type of scene) * 2 (AOI) * missing trials structure, for dwelling time in experiment 2.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
Age	1	51	1.98	.17	.04
Type of scene	1	52	0.04	.84	<.001
AOI	1	104	73.33	< .001	.41
Missing trials	1	51	17.87	< .001	.26
Age*type of scene	1	52	0.03	.86	<.001
Age*AOI	1	104	13.74	< .001	.12
Type of scene*AOI	1	104	33.17	< .001	.24
Age*scene*AOI	1	104	0.16	.69	<.001

Table S3b. Main effects and interactions in the model with an age * 2 (type of scene) * 2 (AOI) * missing trials structure, for time to first fixation in experiment 2.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
Age	1	51	3.99	.05	.07
Type of scene	1	52	0.001	.97	<.001
AOI	1	104	310.84	< .001	.75
Missing trials	1	51	0.33	.57	.01
Age*type of scene	1	52	0.04	.84	<.001
Age*AOI	1	104	7.31	.01	.07
Type of scene*AOI	1	104	7.64	.01	.07
Age*scene*AOI	1	104	0.02	.90	<.001

3b. Models of the dwelling time and time to first fixation including missing trials as a fixed effect in the developmental changes analysis.

For each of the two measures, we performed a model including age, type of scene, AOI and missing trials as fixed effects, and participant, type of scene, type of AOI as random effects. See Tables S3c and S3d for the main effects of the two models.

Table S3c. Main effects and interactions in the model with a 2 (age-group) * 2 (type of scene) * 2 (AOI) * missing trials structure, for dwelling time in the developmental changes analysis.

Predictor	numDF	denDF	F - value	p- value	η^2_p
Age	1	121	1.42	.24	.01
Type of scene	1	122	0.39	.54	<.001
AOI	1	244	77.25	< .001	.24
Missing trials	1	121	19.49	< .001	.16
Age*type of scene	1	122	0.53	.82	<.001
Age*AOI	1	244	12.46	< .001	.05
Type of scene*AOI	1	244	74.91	< .001	.23
Age*scene*AOI	1	244	0.00	.99	<.001

Table S3d. Main effects and interactions in a 2 (age-group) * 2 (type of scene) * 2 (AOI) * missing trials model, for time to first fixation in the developmental changes analysis.

Predictor	numDF	denDF	F - value	p- value	η^2_p
Age	1	121	5.73	.02	.05
Type of scene	1	122	0.01	.94	<.001
AOI	1	244	717.30	< .001	.75
Missing trials	1	121	1.77	.19	.01
Age*type of scene	1	122	0.02	.88	<.001
Age*AOI	1	244	0.00	.99	.00
Type of scene*AOI	1	244	27.73	< .001	.10
Age*scene*AOI	1	244	0.80	.37	<.001

S4: Developmental model descriptive statistics

Table S4a. Descriptive statistics for dwelling time to each AOI, in each condition and in each age-group.

Age-group	Type of scene	AOI	Mean	SD
Children	Interactive	Social	2101.30	1058.53
		Background	1558.98	1059.22
	Non-interactive	Social	1875.93	1125.79
		Background	1770.75	1092.72
Adults	Interactive	Social	1986.53	910.75
		Background	1622.93	893.27
	Non-interactive	Social	1745.58	939.49
		Background	1820.25	916.49

Table S4b. Descriptive statistics for time to first fixation to each AOI, in each condition and in each age-group.

Age-group	Type of scene	AOI	Mean	SD
Children	Interactive	Social	490.84	502.07
		Background	935.83	874.19
	Non-interactive	Social	547.66	592.68
		Background	870.38	847.26
Adults	Interactive	Social	440.54	462.51
		Background	910.34	836.14
	Non-interactive	Social	528.11	598.08
		Background	824.77	806.03

Appendix D – Supplementary materials for chapter 4

S1. Descriptive statistics across the main text

Table S1a. Descriptive statistics for dwelling time (ms) in each age-group, condition and AOI for Part 1.

Age-group	Scene	AOI	Mean	SD
Children (N = 54)	Interactive	Social	1962.21	1123.69
		Background	1722.64	1087.60
	Non-interactive	Social	2036.18	1113.80
		Background	1606.81	1092.14
Adults (N = 98)	Interactive	Social	2003.77	950.54
		Background	1607.67	926.69
	Non-interactive	Social	2101.20	901.19
		Background	1493.05	887.95

Table S1b. Descriptive statistics for dwelling time (ms) to each AOI, in the interactive three and four people pictures, in part 2a.

People	AOI - Human	Mean	SD
3	Interacting	663.31	409.83
	Not interacting	612.02	530.38
4	Interacting	542.50	360.42
	Not interacting	503.86	353.55

Table S1c. Descriptive of untransformed time to first fixation in the two age-groups, in three and four people pictures, to interacting and non-interacting individuals, for part 2c.

Age-group	People	AOI - Human	Mean	SD
Children (N = 54)	3	Interacting	801.71	834.77
		Not-interacting	1355.23	1183.20
	4	Interacting	1000.15	1002.07
		Not-interacting	1138.38	1035.49
Adults (N = 98)	3	Interacting	761.44	772.39
		Not-interacting	1312.54	1095.81
	4	Interacting	903.94	830.32
		Not-interacting	1127.11	965.27

S2. Other models

2a. Analysis of the dwelling time with a 2 (people) x 2 (type of scene) x 2 (AOI) structure in part 1.

The structure of the variance within the model after adding the number of people as a fixed factor did not change: random effects were participant, condition and AOI (*SD* = 323.86, $\chi^2(3) = 851.17$, $p < .001$). See Table S2a for main effects.

Table S2a. Main effects and interactions in the model with a 2 (number of people in the scene) * 2 (type of scene) * 2 (AOI) structure, for exploratory analyses of dwelling time in Part 1.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
people	1	14982	3.03	.08	< .001
type of scene	1	151	0.28	.60	.002
AOI	1	302	335.94	< .001	.53

scene * AOI	1	302	17.99	< .001	.06
people * AOI	1	14982	23.14	< .001	.002
scene * people	1	14982	0.14	.71	< .001
people * scene * AOI	1	14982	107.16	< .001	0.01

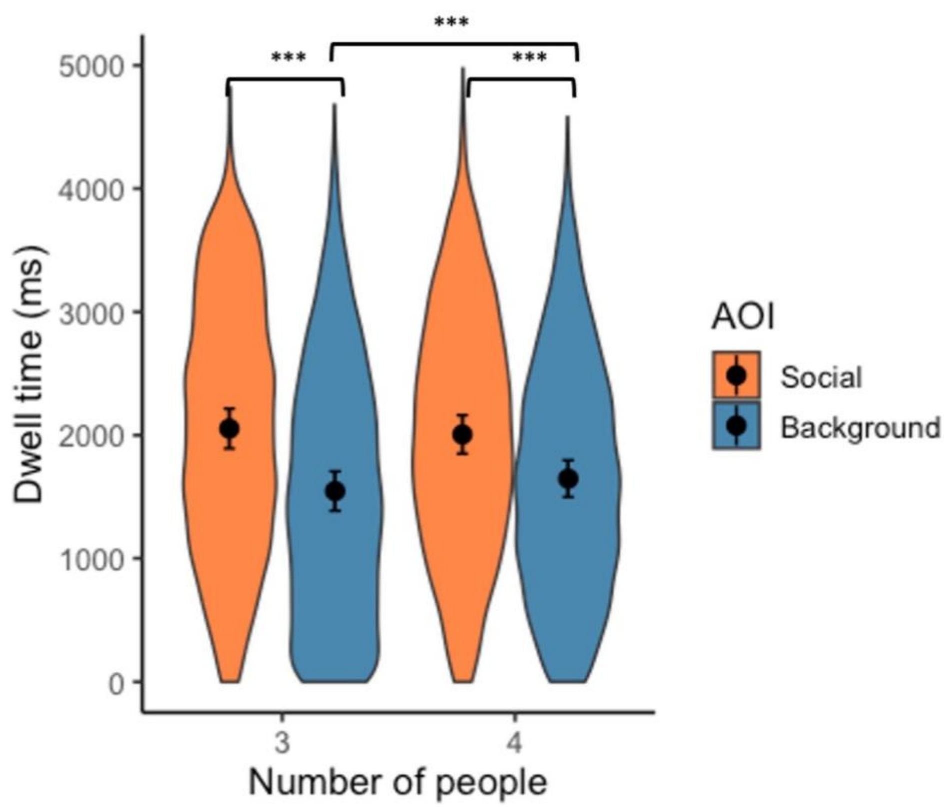

Figure S2a. Violin plot for mean dwell time for AOIs across types of scenes. Error bars represent 95% confidence intervals.

Table S2b. Descriptive statistics for dwelling time (ms) in the 3 and 4 people pictures, in each condition and AOI for Part 1 – exploratory analyses.

Number of people	Scene	AOI	Mean	SD
3	Interactive	Social	1938.91	1010.47
		Background	1670.51	997.72

	Non-interactive	Social	2169.10	1003.61
		Background	1424.11	998.06
4	Interactive	Social	2057.78	1015.76
		Background	1616.77	972.01
	Non-interactive	Social	1955.49	935.45
		Background	1680.25	898.67

2b. Analysis of the dwelling time with a 2(age-group) x 2 (people) x 2 (type of scene) x 2 (AOI) structure in part 1 – Exploratory analyses.

Table S2c. Main effects and interactions in the model with a 2 (age-group) * 2 (number of people in the scene) * 2 (type of scene) * 2 (AOI) structure.

Predictor	numDF	denDF	*F* - value	*p*- value	$\eta^2{}_p$
age-group	1	150	1.22	.27	0.01
type of scene	1	150	0.29	.59	< 0.001
AOI	1	300	342.81	< .001	0.53
people	1	14978	3.03	.08	< 0.001
age * scene	1	150	0.05	.82	< 0.001
age * AOI	1	300	10.87	.001	0.03
age * people	1	14978	1.56	.21	< 0.001
scene * AOI	1	300	18.36	< .001	0.06
scene * people	1	14978	0.14	.71	< 0.001
age * scene * AOI	1	300	0.05	.83	< 0.001
people * scene * AOI	2	14978	65.21	< .001	0.01
age * people * scene	1	14978	0.14	.71	< 0.001
age * people * scene * AOI	2	14978	6.47	.002	< 0.001

Table S2d. Descriptive statistics for dwelling time (ms) in each age-group, by scenes varying by number of people, social content and AOI.

Age-group	People	Scene	AOI	Mean	SD
Children (N = 54)	3	Interactive	Social	1939.39	1123.37
			Background	1702.39	110.33
		Non-interactive	Social	2083.02	1135.46
			Background	1512.32	1133.50
	4	Interactive	Social	1992.96	1124.35
			Background	1749.92	1056.56
		Non-interactive	Social	1972.40	1081.31
			Background	1735.49	1020.14
Adults (N = 98)	3	Interactive	Social	1938.65	945.36
			Background	1653.56	932.28
		Non-interactive	Social	2215.19	922.42
			Background	1376.87	914.27
	4	Interactive	Social	2092.72	950.82
			Background	1545.00	915.73
		Non-interactive	Social	1946.47	847.78
			Background	1650.76	825.55

2c. Analysis of the time to first fixation with a 2 (people) x 2 (type of scene) x 2 (AOI) structure in part 1 – Exploratory analyses.

Table S2e. Main effects and interactions in the model with a 2 (number of people in the scene) * 2 (type of scene) * 2 (AOI) structure for time to first fixation in part 1 – Exploratory analyses.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
people	1	14565	0.3	.59	< .001

type of scene	1	151	0.4	.55	.003
AOI	1	302	323.3	< .001	.52
scene * AOI	1	302	29.8	< .001	< .001
people * AOI	1	14565	26.7	< .001	.002
scene * people	1	14565	0.3	.62	.09
people * scene * AOI	1	14565	12.6	< .001	< .001

Table S2f. Descriptive statistics for transformed time to first fixation to each AOI, across types of scenes and different number of people.

People	Scene	AOI	Mean	SD
3	Interactive	Social	2.63	0.27
		Background	2.72	0.35
	Non-interactive	Social	2.59	0.25
		Background	2.77	0.38
4	Interactive	Social	2.64	0.27
		Background	2.71	0.35
	Non-interactive	Social	2.63	0.26
		Background	2.72	0.35

2d. Models of dwelling time without the pictures with extremely large interacting humans in the Exploratory analyses in part 2c.

Table S2g. Main effects and interactions in the model with a 2 (age-group) * 2 (number of people in the scene) * 2 (AOI) structure for square root transformed dwelling time in part 2c – Exploratory analyses.

Predictor	numDF	denDF	F - value	p- value	η^2_p

age	1	150	20.02	< .001	< .001
people	1	150	5.69	.02	< .001
AOI	1	6592	68.64	< .001	.32
age*people	1	150	0.05	.82	< .001
people*AOI	1	14565	14.14	< .001	.002
age*AOI	1	6592	7.85	.01	.09
age*people*AOI	1	6592	3.76	.05	< .001

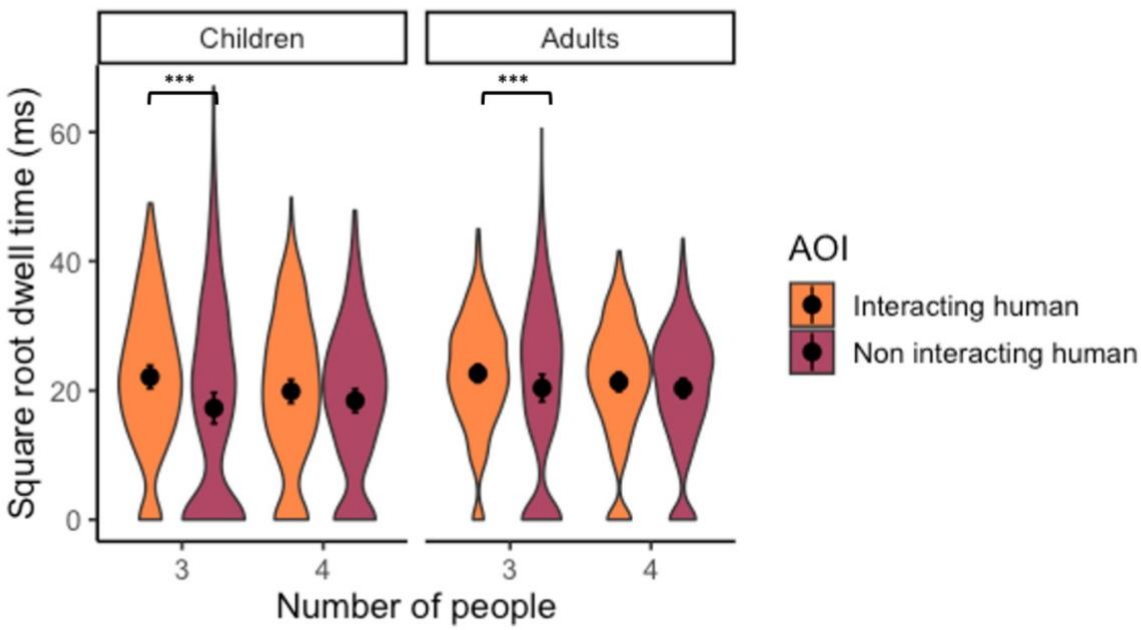

Figure S2b. Violin plot for mean square root dwell time for AOIs across types of scenes for part 2c – Exploratory analyses. Error bars represent 95% confidence intervals.

2e. Models of time to first fixation without the pictures with extremely large interacting humans in the Exploratory analyses in part 2c.

Table S2h. Main effects and interactions in the model with a 2 (age-group) * 2 (number of people in the scene) * 2 (AOI) structure for time to first fixation in part 2c – Exploratory analyses.

Predictor	numDF	denDF	*F* - value	*p*- value	η^2_p

age	1	150	2.15	.15	.01
people	1	150	0.38	.54	< .001
AOI	1	300	184.89	< .001	.38
age*people	1	150	0.15	.70	< .001
people*AOI	1	300	32.78	< .001	< .001
age*AOI	1	300	1.94	.17	< .001
age*people*AOI	1	300	0.08	.78	< .001

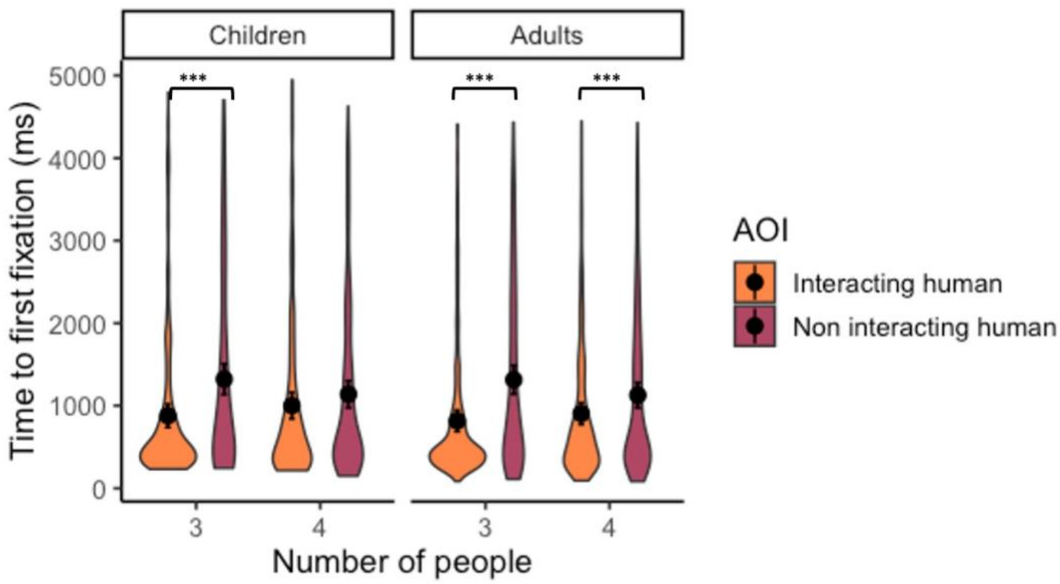

Figure S2c. Violin plot for mean time to first fixation for AOIs across types of scenes for part 2c – Exploratory analyses. Error bars represent 95% confidence intervals.

S3. AOI size information in pixels

3a. Size in pixels of the AOIs for the part 1 of the manuscript. The human AOI size for this part was calculated by summing the area of all the humans in the scene.

Table S3a. Descriptive statistics of AOI size across types of scenes (interactive or non-interactive), in 3 and 4 people scenes.

Number of people	Type of scene	AOI	Mean area (px)	SD
3	Interacting	Social	236829.35	149505.40
		Background	546053.76	148704.87
	Non-interacting	Social	190466.39	98728.34
		Background	524726.00	129910.31
4	Interacting	Social	191828.78	133692.74
		Background	592568.08	132905.95
	Non-interacting	Social	258886.95	129208.32
		Background	599237.02	112235.09

Table S3b. Main effects in an ANOVA with a 2 (number of people) x 2 (type of scene) x 2 (type of AOI) structure, on the AOI size in pixels.

Predictor	Df	F - value	p- value	$\eta^2{}_p$
people	1, 48	0.001	.98	< .001
type of scene	1, 48	0.00	.99	.00
AOI	1, 48	167.94	< .001	.78
scene * AOI	1, 96	0.14	.71	< .001
people*condition	1.96	0.00	.99	.00
people*AOI	1,96	5.14	.03	.05
people*scene*AOI	1,96	0.29	.60	< .001

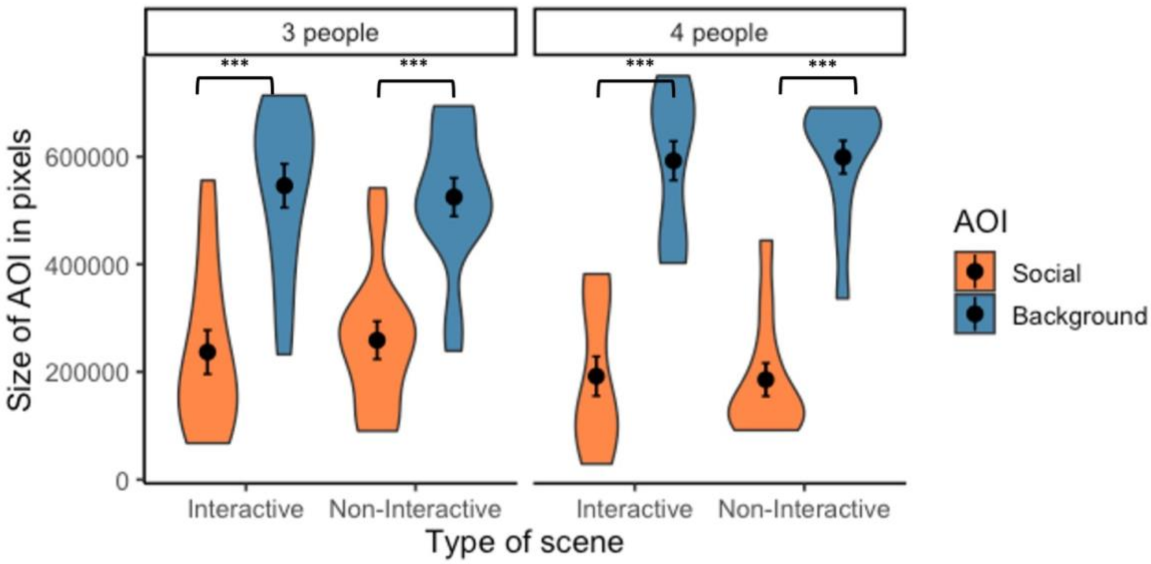

Figure S3a. Violin plots for mean size of AOIs (px) across 3 and 4 people scene, in the interactive and non-interactive conditions. Error bars represent 95% confidence intervals.

3b. Model on the pictures depicting three people, with participant, type of scene and AOI as random effects, and type of scene, AOI and size of the AOI as fixed effects, to investigate the role of the size of the AOI in moderating social orienting in interactive and non-interactive scenes. Intercepts were therefore, similarly to the models used for the main analysis, allowed to vary at participant, condition and AOI level (*SD* = 340.63, $\chi^2(3) = 398.41$, $p < .001$). See table S3c for main effects.

Table S3c. Main effects and interactions in the model with a 2 (type of scene) * 2 (AOI) structure, and area of AOI as continuous predictor.

Predictor	numDF	denDF	*F* - value	*p*- value	$\eta^2{}_p$
type of scene	1	151	0.09	.76	< .001
AOI	1	302	380.51	< .001	.56
area	1	8372	2969.39	< .001	.26
scene * AOI	1	302	37.60	< .001	.11
area * scene	1	8372	0.25	.61	< .001

area * AOI	1	8372	0.41	.52	< .001
scene * AOI * area	1	8372	0.94	.33	< .001

3c. Model on the pictures depicting three people, with participant, type of scene and AOI as random effects, and type of scene, AOI and size of the AOI as fixed effects, to investigate the role of the size of the AOI in moderating time to first fixation in interactive and non-interactive scenes. Intercepts were therefore, similarly to the models used for the main analysis, allowed to vary at participant, condition and AOI level ($SD = 0.80$, $\chi^2(3) = 143.96$, $p < .001$). See table S3d for main effects.

Table S3d. Main effects and interactions in the model with a 2 (type of scene) * 2 (AOI) structure, and area of AOI as continuous predictor for time to first fixation in part 1 – Exploratory analyses.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
type of scene	1	151	0.6	.45	.004
AOI	1	302	335.6	< .001	.53
area	1	8085	640.3	< .001	.07
scene * AOI	1	302	25.7	< .001	.08
area * scene	1	8085	3.5	.06	< .001
area * AOI	1	8085	21.3	< .001	.003
scene * AOI * area	1	8085	0.4	.50	< .001

3d. Size in pixels of the AOIs for the part 2 of the manuscript. The human AOI size for this part was calculated by averaging the area of all the humans in the scene per condition.

Table S3e. Descriptive statistics of human AOI size in the interactive 3 and 4 people scenes (part 2 of the main text).

Number of people	Type of human	Mean area (px)	SD
3	Interacting	90447.79	67234.26
	Non-interacting	55933.77	30619.67
4	Interacting	43494.58	34747.44
	Non-interacting	52419.81	44462.59

Table S3f. Main effects in an ANOVA with a 2 (number of people) x 2 (type of human AOI) structure, on the AOI size in pixels, for part 2.

Predictor	Df	F - value	p- value	$\eta^2{}_p$
people	1, 48	3.58	.06	.07
type of AOI	1, 48	1.5	.23	.03
people*AOI	1,96	2.66	.11	.05

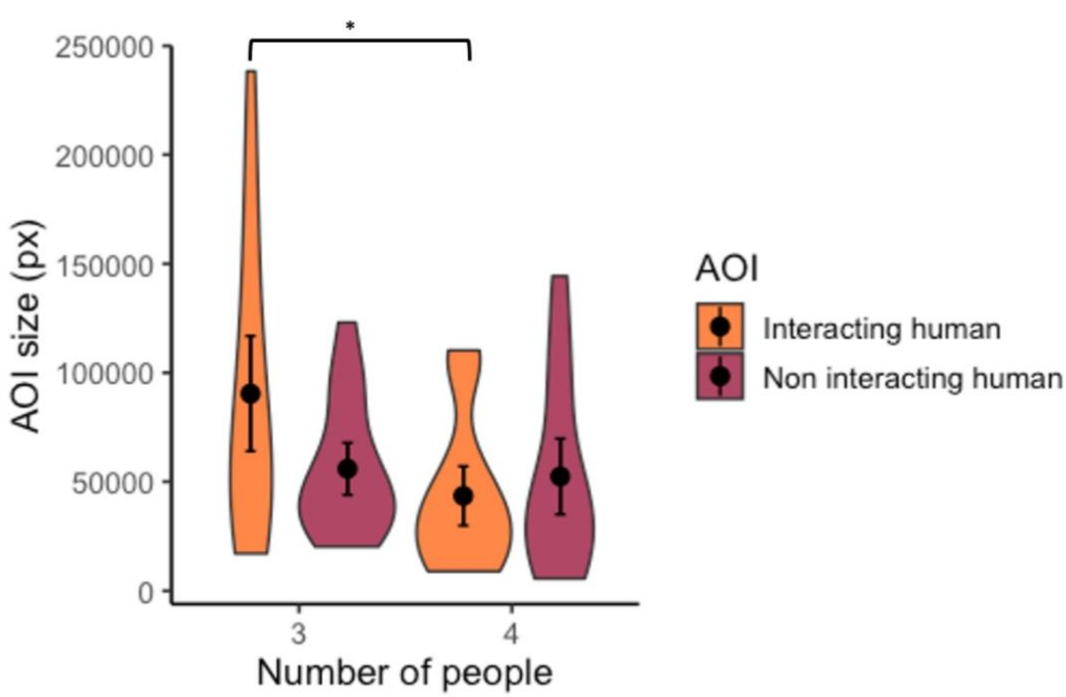

Figure S3a. Violin plots for mean size of AOIs (px) across 3 and 4 people scene, in the interactive and non-interactive conditions. Error bars represent 95% confidence intervals.

S4. Interpersonal spaces analysis for exploratory analyses in part 1

4a. Space analysis in part 1.

For each 3-person picture that contained a social interaction (number of scenes) we created hand-drawn AOIs of the space between the two interacting humans, between the interacting human and the non-interacting human in the interactive scenes, and the space between each pair of non-interacting humans in the non-interactive scenes. For each interactive scene we had, therefore, a dwell time value for the interactive space and one for the mixed space, and for each non-interactive scene a dwell time value for the non-interactive space, given by the average dwell time to all the spaces in the scene.

The same cleaning procedures were followed as for the main analysis. Additionally, because an interactive space was necessary in this analysis, we excluded 5 pictures from the interactive category either because the interactive space was interrupted by a human or there was no interactive space at all (give an example of why this might be).

Table S4a. Descriptive statistics for dwelling time (ms) in each age-group, and type of interpersonal space.

Age-group	AOI – spaces	Mean	SD
Children (N = 54)	Interactive	380.06	545.72
	Mixed	256.14	510.50
	Non-interactive	300.18	412.55
Adults (N = 98)	Interactive	384.38	472.47
	Mixed	211.41	358.13
	Non-interactive	255.07	313.39

Table S4b. Descriptive statistics for time to first fixation in each age-group, and type of interpersonal space.

Age-group	AOI – spaces	Mean	SD
Children (N = 54)	Interactive	1310.10	1192.94
	Mixed	1670.41	1444.46
	Non-interactive	1051.85	1096.12
Adults (N = 98)	Interactive	1368.12	1177.34
	Mixed	1550.11	1328.16
	Non-interactive	1111.55	1123.42

S5. Data transformations

5a. Details on time to first fixation log-transformation in part 1.

The Anderson-Darling test for normality (A = 1420.7, p < .001) and visual inspection of residual values showed that the time to first fixation data is positively skewed, therefore a logarithm in base 10 was applied to the data to improve skewness. After transformation, the Anderson test for normality was still significant (A = 436.26, p < .001) although the graphical shape of the distribution improved.

Table S5a. Descriptive statistics for log-transformed time to first fixation in each age-group, condition and AOI, for part 1 analyses.

Age-group	Scene	AOI	Mean	SD
Children (N = 54)	Interactive	Social	2.66	0.28
		Background	2.73	0.33
	Non-interactive	Social	2.65	0.27
		Background	2.77	0.35
Adults (N = 98)	Interactive	Social	2.62	0.27

		Background	2.70	0.36
	Non-interactive	Social	2.59	0.24
		Background	2.74	0.37

Table S5b. Descriptive statistics for untransformed time to first fixation in each age-group, condition and AOI.

Age-group	Scene	AOI	Mean	SD
Children (N = 54)	Interactive	Social	609.37	685.38
		Background	758.85	782.85
	Non-interactive	Social	579.31	631.77
		Background	850.34	848.58
Adults (N = 98)	Interactive	Social	536.73	540.94
		Background	750.82	810.77
	Non-interactive	Social	476.30	465.29
		Background	819.85	838.58

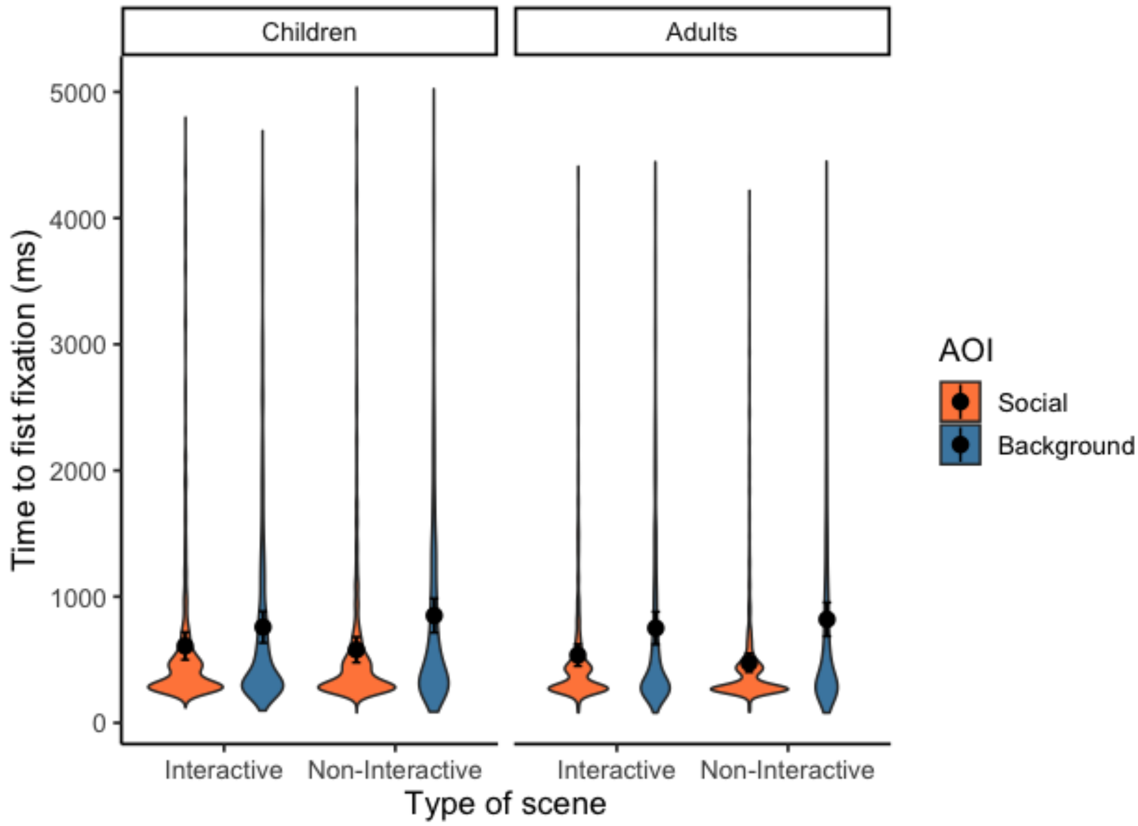

Figure S5a. Violin plot for mean untransformed time to first fixation to social and non-social AOI across types of scenes in the two age-groups. All contrasts between human and background are significant at p < .001. Error bars represent 95% confidence intervals.

5b. Square root transformation details of time to first fixation in part 2.a

The Anderson-Darling test for normality (A = 234.09, p < .001) and visual inspection of residual values showed that the time to first fixation data is positively skewed, therefore square root transformation was applied to the data to improve skewness. After transformation the Anderson test for normality was still significant (A = 117.76, p < .001) although the graphical shape of the distribution improved.

We used the same model reported in the main manuscript with the transformed data: participant, type of scene, and type of human were set as random factors, (SD = 2.94, $\chi^2(3)$ = 51.43, p < .001). The main effects using the transformed data are very similar to those reported in the main text, with a main effect of AOI-type and an interaction between the number of people in a scene and the AOI-type. See Table S5c for main effects.

Table S5c. Main effects and interactions in the model with a 2 (number of people) * 2 (AOI) structure for transformed time to first fixation in part 2.a

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
number of people	1	97	0.64	.43	.01
AOI (type of human)	1	194	234.624	< .001	.55
people * AOI	1	194	37.48	< .001	.16

Table S5d. Descriptive statistics for transformed time to first fixation to each AOI, in the three and four people pictures for analysis in part 2.a

People	AOI - Human	Mean	SD
3	Interacting	25.30	11.02
	Not interacting	33.12	14.70
4	Interacting	27.62	11.89
	Not interacting	30.75	13.48

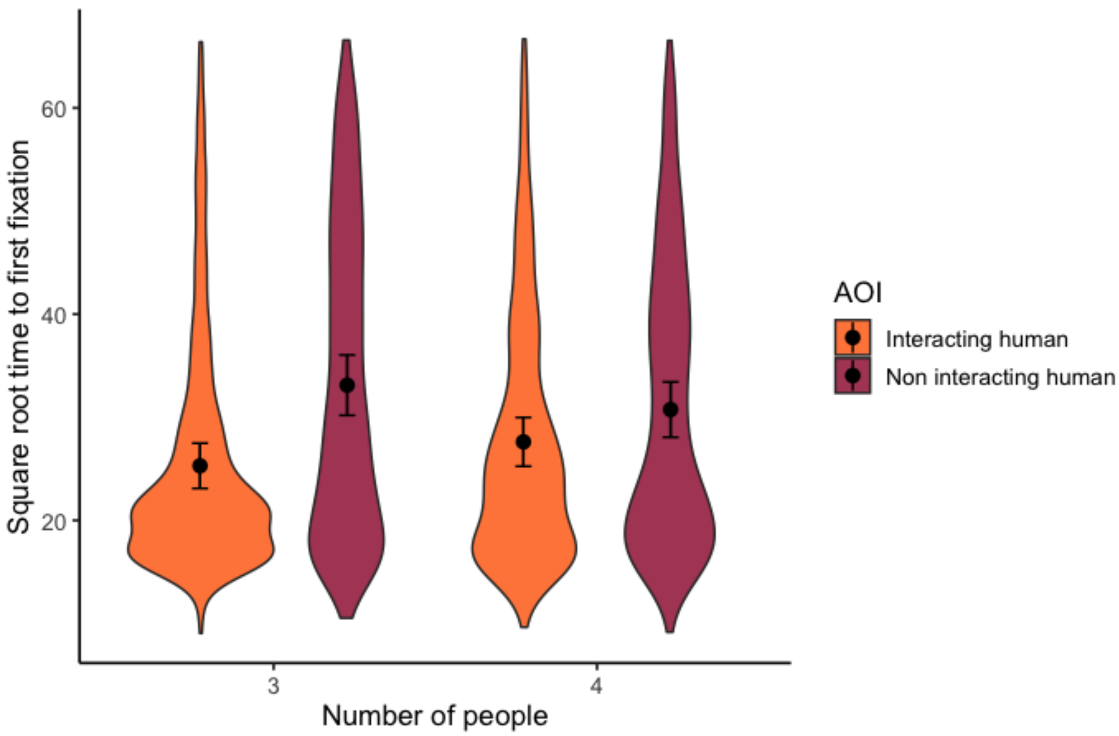

Figure S5b. Violin plot for transformed mean time to first fixation for interacting and non-interacting humans in the three and four people pictures, for analysis in part 2.a. Error bars represent 95% confidence intervals.

5c. Details on square root transformation in the attentional engagement in part 2b

The Anderson-Darling test for normality (A = 65.35, p < .001) and visual inspection of residual values showed that dwelling time data was positively skewed, therefore a square root transformation was applied to improve skewness. After transformation the Anderson test for normality was still significant (A = 12.33, p < .001) although the graphical shape of the distribution improved.

Similarly to the results in the main text, in this model, participant, type of scene and type of human were set as random factors, (SD = 2.43, $\chi^2(3)$ = 16.68, p < .001), and the main effects were number of people in the scene and type of human (interacting or not interacting).

See Table S5e for main effects.

Table S5e. Main effects and interactions in the model with a 2 (number of people) * 2 (AOI) structure for transformed dwelling time in part 2b.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
age	1	52	0.03	.86	< .001
number of people	1	52	11.64	.001	.18
AOI	1	104	83.68	< .001	.45
people * AOI	1	104	27.73	< .001	.21
age * people	1	52	0.40	.53	.01
age * AOI	1	104	0.13	.72	.001
age * people * AOI	1	104	4.40	0.04	.04

Table S5f. Descriptive statistics for transformed dwelling time to each AOI, in the three and four people pictures, for part 2b.

People	AOI - Human	Mean	SD
3	Interacting	24.05	11.30
	Not interacting	17.52	14.84
4	Interacting	19.87	11.58
	Not interacting	18.42	11.36

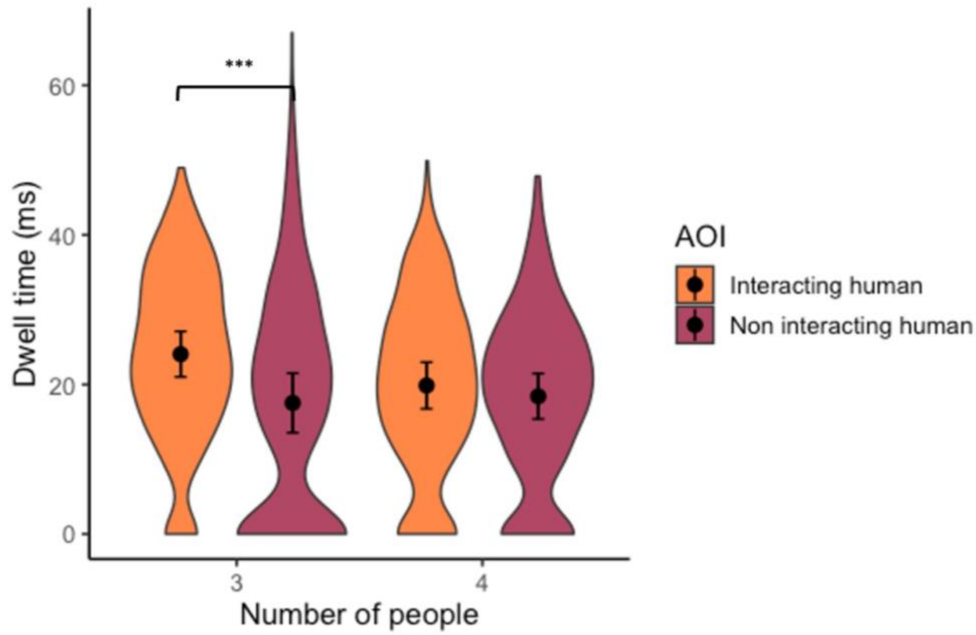

Figure S5c. Violin plot for transformed mean dwelling time for interacting and non-interacting humans in the three and four people pictures, for part 2b. Error bars represent 95% confidence intervals.

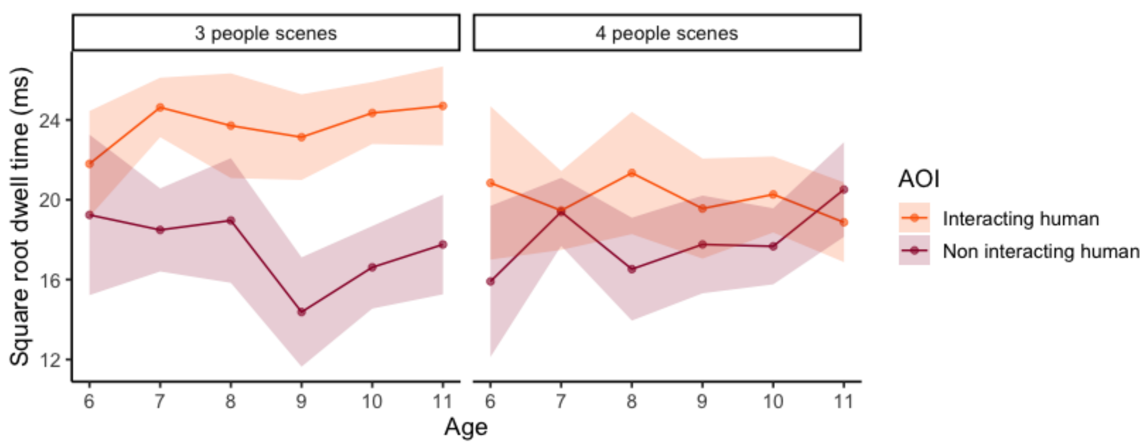

Figure S5d. Transformed mean dwelling time for interacting and non-interacting humans in the three and four people pictures across childhood. Width of bands represent 95% confidence intervals.

5d. Details on logarithm in base 10 transformation of time to first fixation data in part2b

The Anderson-Darling test for normality (A = 142.99, p < .001) and visual inspection of residual values showed that time to first fixation data were positively skewed, therefore a logarithm in base 10 transformation was applied to improve skewness. After transformation the Anderson test for normality was still significant (A = 30.27, p < .001), although the graphical shape of the distribution improved.

As reported in the main text using the untransformed data, in this model, participant, type of scene and type of human were set as random factors, (SD = 0.06, $\chi^2(3) = 8.57$, p = .003), and the main effects were number of people in the scene and type of human (interacting or not interacting). See Table S17 for main effects.

Table S5g. Main effects and interactions in the model with a 2 (number of people) * 2 (AOI) structure for transformed time to first fixation in part 2b.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
age	1	52	4.27	.04	.08
number of people	1	52	1.60	.21	.03
AOI	1	104	79.11	< .001	.43
people * AOI	1	104	20.74	< .001	.17
age * people	1	52	0.04	.84	< .001
age * AOI	1	104	0.03	.85	< .001
age * people * AOI	1	104	0.01	0.92	< .001

Table S5h. Descriptive statistics for transformed time to first fixation to each AOI, in the three and four people pictures for analysis in part 2b.

People	AOI - Human	Mean	SD
3	Interacting	2.76	0.32
	Not interacting	2.96	0.39
4	Interacting	2.84	0.36
	Not interacting	2.89	0.37

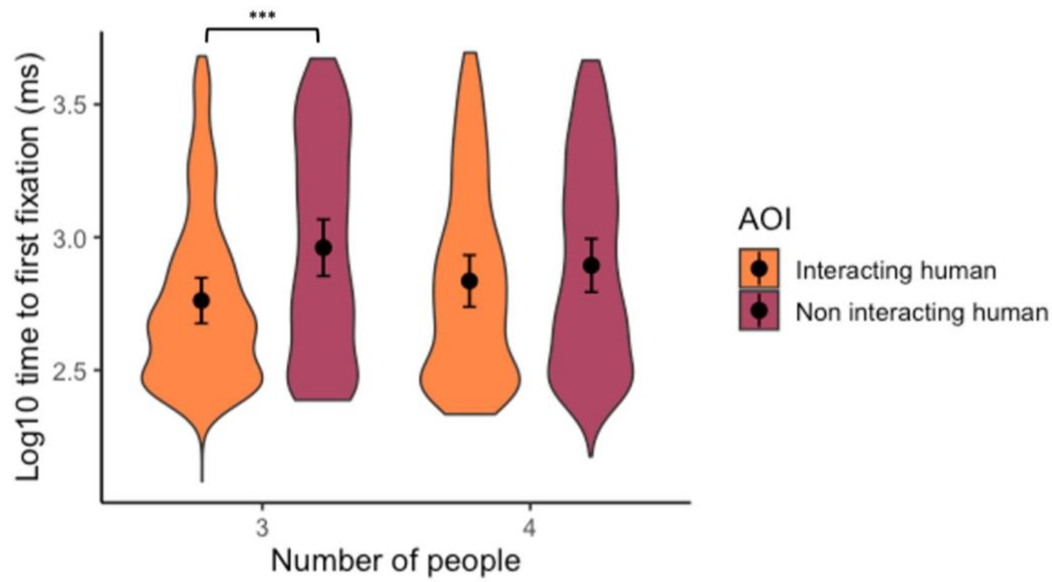

Figure S5e. Violin plot for log transformed mean time to first fixation for interacting and non-interacting humans in the three and four people pictures. Error bars represent 95% confidence intervals.

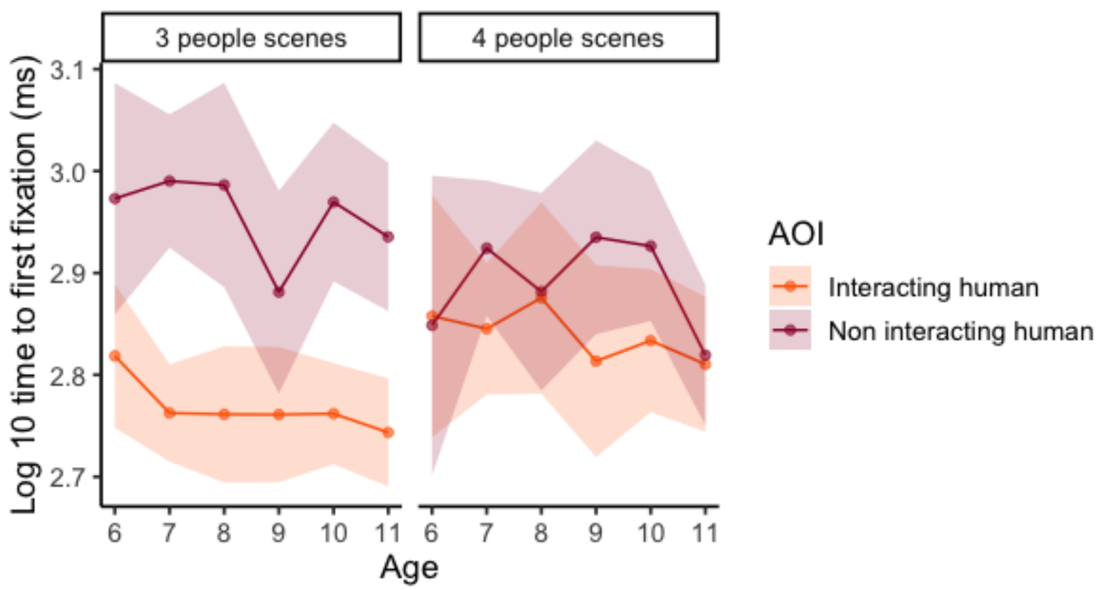

Figure S5f. Transformed mean time to first fixation for interacting and non-interacting humans in the three and four people pictures across childhood. Width of bands represent 95% confidence intervals.

5e. Details on square root transformation of dwelling time data in part2c

The Anderson-Darling test for normality ($A = 96.81$, $p < .001$) and visual inspection of residual values showed that dwelling time data were positively skewed, therefore a square root transformation was applied to improve skewness.

After transformation the Anderson test for normality was still significant ($A = 42.45$, $p < .001$) although the graphical shape of the distribution improved.

Table S5i. Descriptive statistics for transformed (square root) dwelling time (ms) in each age-group, condition and AOI for part 2c.

Age-group	People	AOI - Human	Mean	SD
Children (N = 54)	3	Interacting	24.05	11.30
		Not-interacting	17.52	14.84
	4	Interacting	19.87	11.58
		Not-interacting	18.42	11.36
Adults (N = 98)	3	Interacting	24.18	8.87
		Not-interacting	21.23	12.71
	4	Interacting	21.34	9.34
		Not-interacting	20.33	9.51

Table S5j. Descriptive statistics for untransformed dwelling time (ms) in each age-group, condition and AOI for analyses in part 2c.

Age-group	People	AOI - Human	Mean	SD
Children (N = 54)	3	Interacting	706.13	532.87
		Not-interacting	527.12	654.38
	4	Interacting	528.50	468.23
		Not-interacting	467.98	435.65
Adults (N = 98)	3	Interacting	663.31	409.83

		Not-interacting	612.02	530.38
	4	Interacting	542.50	360.42
		Not-interacting	503.86	353.55

5f. Details on logarithm in base 10 transformation of time to first fixation in part2c

The Anderson-Darling test for normality (A = 375.99, p < .001) and visual inspection of residual values showed that time to first fixation data were positively skewed, therefore a logarithm in base 10 transformation was applied to improve skewness. After transformation the Anderson test for normality was still significant (A = 74.56, p < .001) although the graphical shape of the distribution improved.

The same model was used as in the main text with participant, type of scene, and type of human set as random factors, (SD = 0.07, $\chi^2(3) = 53.26$, p < .001), and the main effects were age-group, number of people in the scene and type of human (interacting or not interacting). See Table S21 for main effects.

Table S5k. Main effects and interactions in the model with a 2 (age-group) * 2 (people in the scene) * 2 (AOI) structure for transformed time to first fixation for part 2c.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
age	1	150	1.3	.27	.009
number of people	1	150	3.3	.21	.02
AOI	1	300	295.5	< .001	.50
people * AOI	1	300	56.7	< .001	.16
age * people	1	150	0.0	.99	.00
age * AOI	1	300	1.2	.28	.004
age * people * AOI	1	300	0.1	.76	< .001

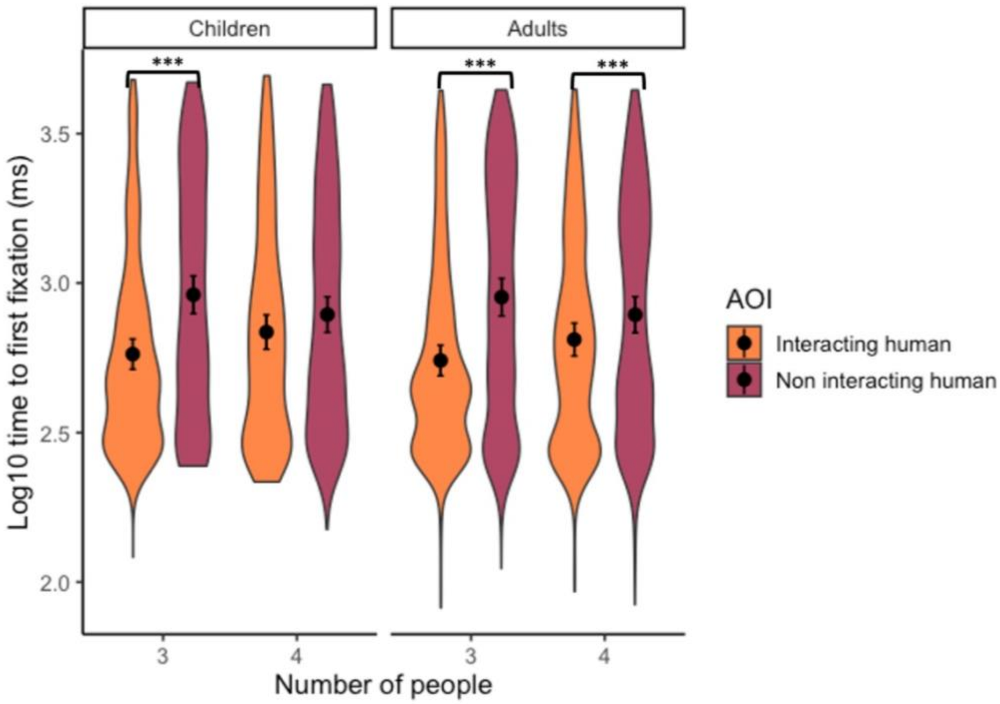

Figure S5g. Violin plot for transformed mean time to first fixation in the two groups for interacting and non-interacting humans in the three and four people pictures. Error bars represent 95% confidence intervals.

Table S5l. Descriptive statistics for transformed (logarithm in base 10) time to first fixation (ms) in each age-group, condition and AOI for part 2c.

Age-group	People	AOI - Human	Mean	SD
Children (N = 54)	3	Interacting	2.76	0.32
		Not-interacting	2.96	0.39
	4	Interacting	2.84	0.36
		Not-interacting	2.89	0.37
Adults (N = 98)	3	Interacting	2.74	0.32
		Not-interacting	2.95	0.39
	4	Interacting	2.81	0.34
		Not-interacting	2.89	0.38

Appendix E – Supplementary materials for chapter 5

S1: AOI size information in pixels

Table S1a. Descriptive statistics of AOI sizes across conditions for the adult sample.

Type of scene	AOI	Mean area (px)	SD
Interacting	Social	212286.13	105769.54
	Background	568078.75	105327.00
Non-interacting	Social	190466.39	98728.34
	Background	590320.98	98722.62

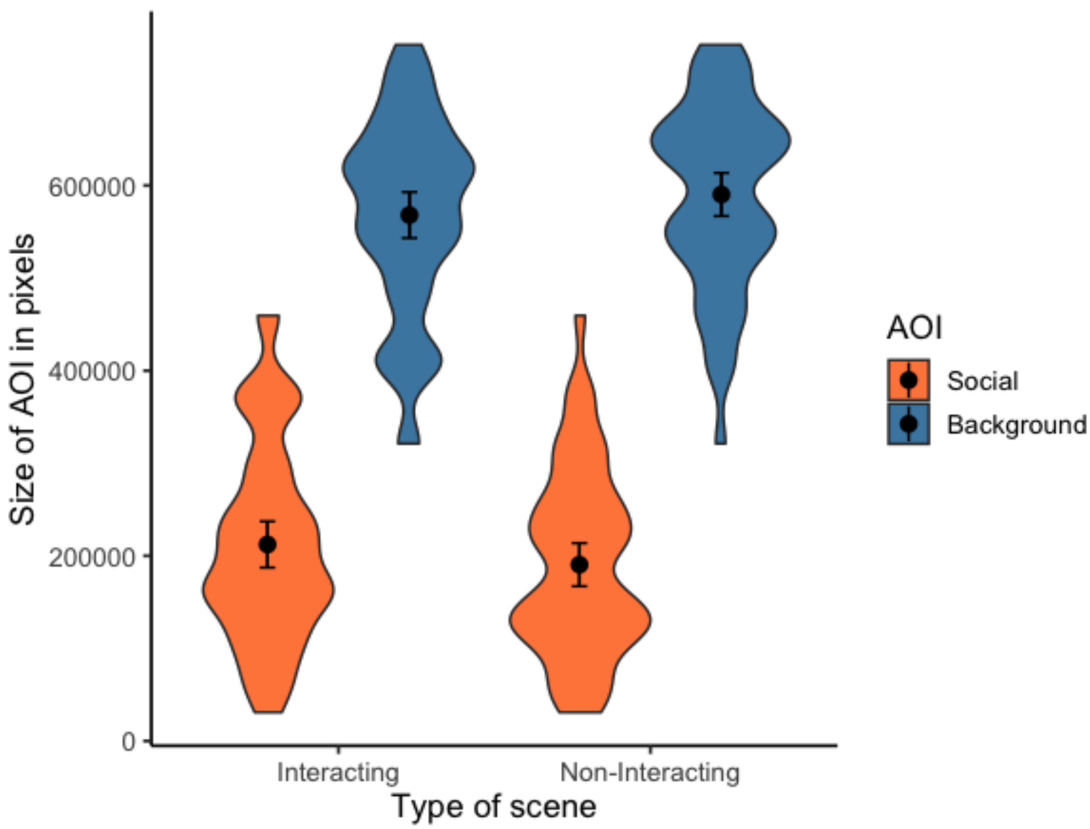

Figure S1a. Violin plot for mean size of AOIs (px) across scenes categorized as interacting and non-interacting in the adult sample. Error bars represent 95% confidence intervals.

Since each participant made their own personal categorization of the scenes as interacting or not, each participant had a different set of pictures in each category. Therefore the size of the AOI in the two types of scenes are different for each group, because individuals in the two groups categorized the scenes differently. In particular, adults categorised *more* scenes are interactive than did children. Here, for the adult and the child groups separately, we show the size of AOIs in the pictures as categorized by the participants.

Table S1b. Descriptive statistics of AOI sizes across conditions in the children's group.

Type of scene	AOI	Mean area (px)	SD
Interacting	Social	208046.64	96552.89
	Background	601640.51	95978.51
Non-interacting	Social	202050.75	111695.25
	Background	578653.87	111742.26

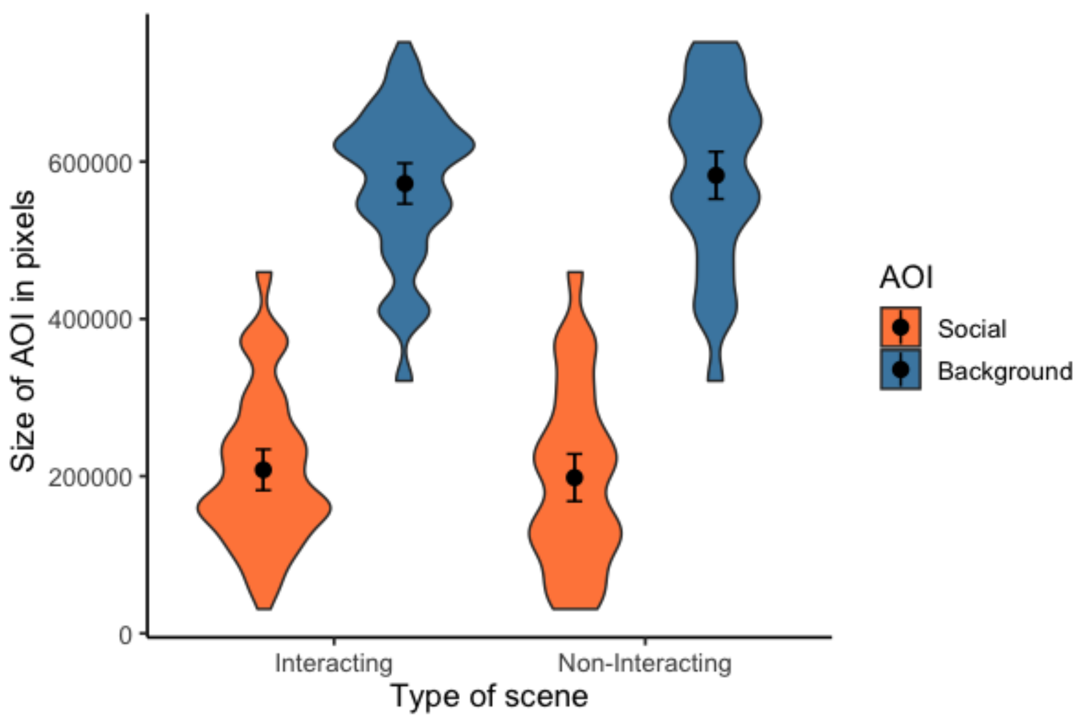

Figure S1b. Violin plot for mean size of AOIs (px) across scenes categorized as interacting and non-interacting. Error bars represent 95% confidence intervals.

S2. Data transformations

2a. Transformation of time to first fixation data, and analysis of transformed time to first fixation in the adult sample

The Anderson-Darling test for normality (A = 304.43, p < .001) and visual inspection of residual values showed that the time to first fixation data is positively skewed, therefore a logarithm in base 10 transformation was applied to the data to reduce skewness. After transformation, however, the Anderson test for normality was still significant (A = 78.10, p < .001) although the graphical shape of the distribution was much improved. When the transformed data was entered into the model, results were entirely similar to those found using the untransformed data, resulting in a main effect of AOI type but no main effect of categorization and no interaction between categorisation and AOI type. Participant and type of AOI were set as random factors, (SD = 0.1, $\chi^2(2) = 155.58$, p < .001), and the fixed effects were the categorization of the scene as interacting or non-interacting, and the AOI – human or background. When the size of the AOIs was added as a random effect, the model did not change significantly – SD = 0.31, $\chi^2(3) = 0.00$, p = .99.

See Table S2a for main effects, Table S2b for descriptive statistics, and Figure S2 for mean transformed time to first fixation across conditions.

Table S2a. Main effects and interactions in the model with a 2 (categorization) * 2 (AOI) structure in the adult sample.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
categorization	1	3958	0.43	.51	< .001
AOI	1	69	177.41	< .001	.72
categorization * AOI	1	3958	0.00	1.00	0.00

Table S2b. Descriptive statistics for transformed time to first fixation to each AOI, in the scenes categorized as interacting and non-interacting in the adult sample.

Categorization	AOI	Mean	SD
Interacting	Social	2.59	0.25
	Background	2.75	0.37
Non interacting	Social	2.60	0.25
	Background	2.75	0.36

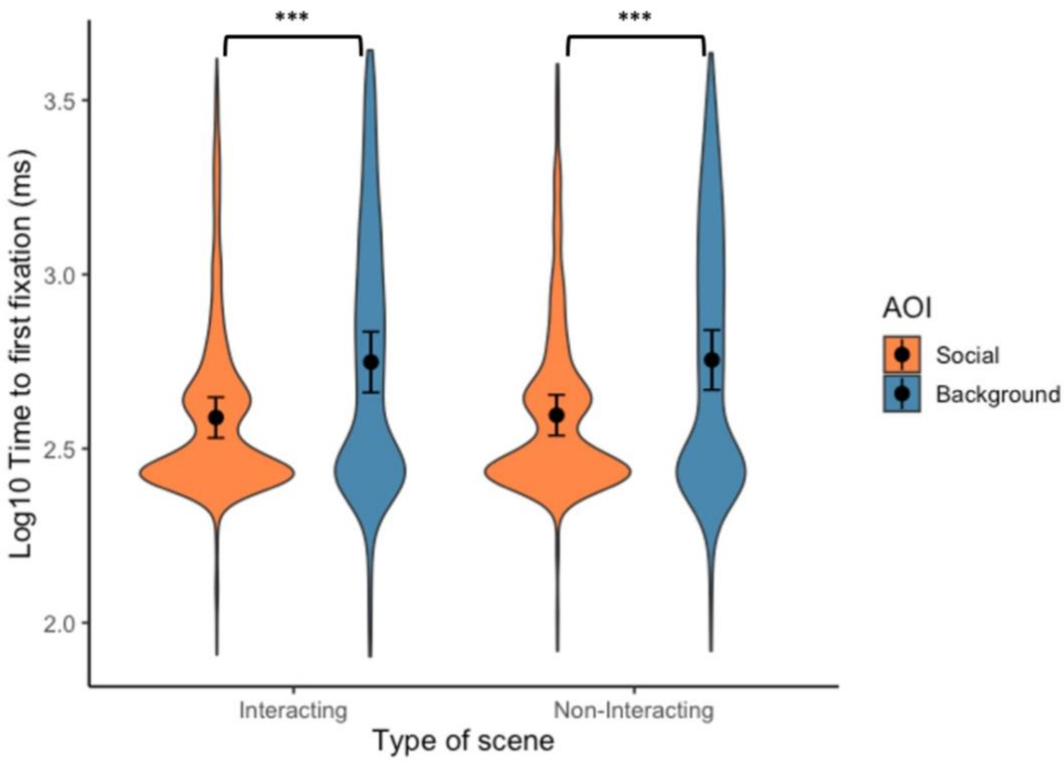

Figure S2a. Violin plot for transformed mean time to first fixation for human and background AOIs in scenes categorized as interacting and non-interacting in the adults sample. Error bars represent 95% confidence intervals.

2b. Transformation of time to first fixation data, and analysis of transformed time to first fixation in the developmental sample

The Anderson-Darling test for normality (A = 269.44, p < .001) and visual inspection of residual values showed that the time to first fixation data is positively skewed, therefore a logarithm in base 10 transformation was applied to the data to improve skewness. However, even after transformation, the Anderson test for normality was still significant (A = 77.31, p < .001) although the graphical shape of the distribution was improved. Participant and type of AOI were set as random factors, (SD = 0.1, $\chi^2(2)$ = 94.54, p < .001), and the fixed effects were the categorization of the scene as interacting or non-interacting and the AOI – human or background. The centred age was modelled as a continuous predictor. When the size of the AOIs was added as a random effect, the model did not change significantly – SD = 0.31, $\chi^2(3)$ = 0.00, p = .99. As in the data reported in the main manuscript text, the only significant effect shown was a main effect of AOI, where children were faster to look at human AOIs than at any other information in the scene.

See Table S2c for main effects, Table S2d for descriptive statistics, Figure S2b for mean transformed time to first fixation across conditions and Figure S2c for transformed time to first fixation across conditions in relation to age.

Table S2c. Main effects and interactions in the model with log10 transformed data.

Predictor	numDF	denDF	F - value	p- value	η^2_p
age	1	52	1.79	.19	.03
categorization	1	2896	0.13	.72	< .001
AOI	1	52	128.36	< .001	.71
age*categorization	1	2896	0.07	.79	< .001
categorization*AOI	1	2896	0.24	.63	< .001
age*AOI	1	52	1.09	.30	.02
age*categorization*AOI	1	2896	0.63	.43	< .001

Table S2d. Descriptive statistics for transformed time to first fixation to each AOI, in the scenes categorized as interacting and non-interacting.

Categorization	AOI	Mean	SD
Interacting	Social	2.63	0.27
	Background	2.79	0.34
Non interacting	Social	2.64	0.26
	Background	2.78	0.35

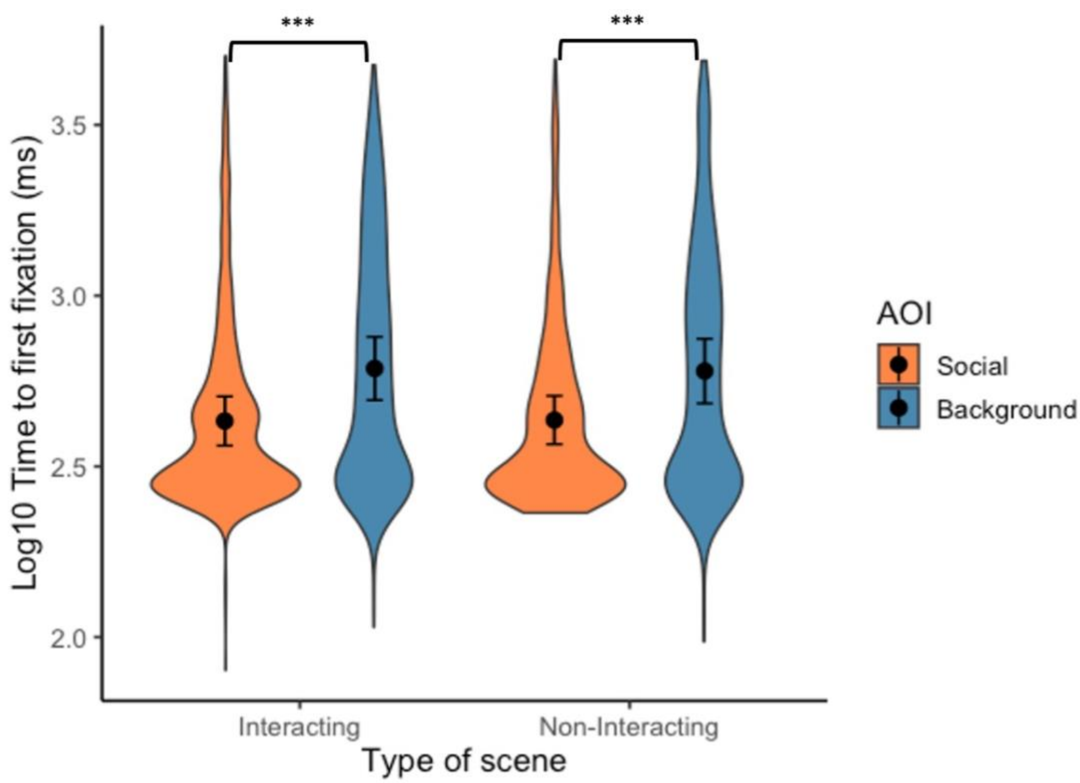

Figure S2b. Violin plot for transformed mean time to first fixation for human and background AOIs in scenes categorized as interacting and non-interacting in the developmental group. Error bars represent 95% confidence intervals.

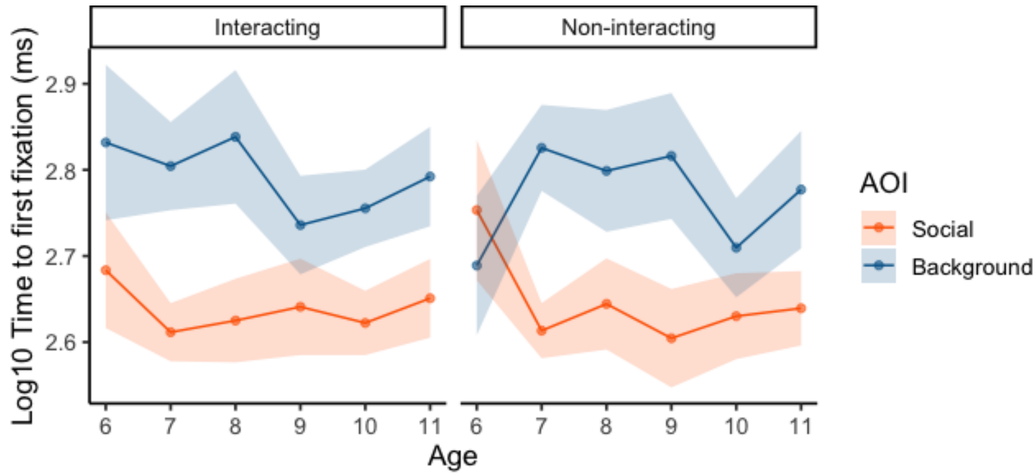

Figure S2c. Average log transformed time to first fixation (ms) to social and background AOIs in scenes categorized as interacting and non-interacting in relation to age in the developmental group. Width of the bands represent 95% confidence intervals.

2c. Transformation of time to first fixation data, and analysis of transformed time to first fixation in the developmental model in part 3

The Anderson-Darling test for normality (A = 572.05, p < .001) and visual inspection of residual values showed that the time to first fixation data is positively skewed, therefore a logarithm in base 10 transformation was applied to the data to improve skewness.
After transformation the Anderson test for normality was still significant (A = 153.4, p < .001) although the graphical shape of the distribution improved.

In this model participant and type of AOI were set as random factors, (SD = 0.1, $\chi^2(2)$ = 263.67, p < .001), and when size of the AOIs was added as a random effect, the model did not change significantly – SD = 0.31, $\chi^2(3)$ = 0.00, p = .99.See Table S8 for main effects, Table S9 for descriptive statistics, and Figure S6 for mean transformed time to first fixation across conditions in the two age-groups.

Table S2e. Main effects and interactions in the model with log10 transformed data for developmental model of time to first fixation in part 3.

Predictor	numDF	denDF	F - value	p- value	$\eta^2{}_p$
age	1	122	17.5	< .001	.13
categorization	1	6856	0.1	.76	< .001
AOI	1	122	302.9	< .001	.71
age*categorization	1	6856	0.4	.53	< .001
categorization*AOI	1	6856	0.1	.73	< .001
age*AOI	1	122	0.3	.59	.003
age*categorization*AOI	1	6856	0.1	.72	< .001

S3. Descriptive statistics

Table S3a. Descriptive statistics for average proportion of the scenes categorized as interacting and non-interacting, for each age in the developmental group.

Age	Scene	Mean	SD
6	Interacting	50.83	9.95
	Non interacting	47.50	10.32
7	Interacting	46.19	22.79
	Non interacting	47.86	23.12
8	Interacting	48.10	12.00
	Non interacting	50.95	12.13
9	Interacting	56.19	15.80
	Non interacting	42.38	14.62
10	Interacting	59.09	11.06
	Non interacting	40.00	13.17
11	Interacting	53.33	13.50
	Non interacting	42.12	16.14

Table S3b. Descriptive statistics for a 2 (age-group) * 2 (type of categorization) * 2 (AOI) interaction for dwelling time (part 3 in the main manuscript).

Age-group	Scene categorization	AOI	Mean	SD
Children	Interacting	Social	1927.85	1107.07
		Background	1777.00	1098.61
	Non interacting	Social	1927.36	1100.03
		Background	1751.89	1098.17
Adults	Interacting	Social	1869.89	895.70
		Background	1734.31	874.39
	Non interacting	Social	1854.93	882.72
		Background	1725.73	865.23

Table S3c. Descriptive statistics for a 2 (age-group) * 2 (type of categorization) * 2 (AOI) interaction for untransformed time to first fixation in part 3 of the main manuscript.

Age-group	Scene categorization	AOI	Mean	SD
Children	Interacting	Social	558.27	601.44
		Background	861.01	807.07
	Non interacting	Social	557.00	592.06
		Background	866.98	881.77
Adults	Interacting	Social	484.10	472.94
		Background	831.12	852.09
	Non interacting	Social	489.87	472.06
		Background	821.18	773.92