Behavioral and Neural Mechanisms of Social Attention

Behaviorale und Neuronale Mechanismen der Sozialen Aufmerksamkeit

Doctoral thesis for a doctoral degree

at the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Neuroscience

submitted by

Lara Rösler

Born in Gießen

Würzburg, 2020

Members of the Thesis Committee

Chairperson	Prof. Dr. Keram Pfeiffer
Primary Supervisor	Prof. Dr. Matthias Gamer
Supervisor (Second)	Prof. Dr. Esther Asan
Supervisor (Third)	Prof. Dr. Johannes Hewig
Supervisor (Fourth)	PD. Dr. Angelika Schmitt-Böhrer

Date of Public Defence: ....

Date of Receipt of Certificates: ....

Submitted on: ....

Office stamp

To Marie Rösler

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Prof. Dr. Matthias Gamer for his constant support. Thank you for sharing your methodological and statistical expertise with me – I learned so much in the past years. More importantly, thank you for creating a warm and calm atmosphere in our lab. Many early career researchers struggle with unhealthy productivity expectations, often fueled by their superiors. You set a different example and taught me how to balance work and the life outside of it. Thank you also for your trust and tolerance, which allowed me to explore other lines of research and to find my own path in academia.

I would further like to thank my supervisory committee – Prof. Dr. Esther Asan, Prof. Dr. Johannes Hewig and PD. Dr. Angelika Schmitt-Böhrer – for helpful discussions and guidance throughout my PhD.

To my colleagues at the Lehrstuhl für Psychologie 1 – thank you for making the institute much more than just a workplace. It’s hard to come by another department with so many coffee breaks, lengthy lunches and after-work wines. I could list a million memories that I cherish, from advent calendar challenges to vivid discussions of Kellogg’s Smacks. I have had a lot fun at work in the past years and you will all be missed.

In particular, I would like to thank Kristina and Lea for forcing me to go swimming once a week and casually providing mentorship for my career and life in general. And finally, Aleya – thank you for always being there, listening and thinking along. My PhD experience would not have been the same without you.

My friends from home – Lucas, Carina, Melissa – you have been my side since the very beginning. Thank you for keeping me grounded and shifting my focus away from the university bubble whenever I needed it.

My friends in the Netherlands, thank you for great summer escapes and for making me return.

Fanny, thank you for your unfailing friendship and advice. You are extremely skilled in cheering me up from a distance and I am forever grateful for that.

My parents, Sabine and Klaus, thank you for being there for me unconditionally. I am, to this day, surprised how normal of a childhood I got to have because of you. I cannot imagine how difficult this must have been. I admire you for your unwavering love and ability to see the good in things. Thank you for teaching me how to cope when life seems insurmountably difficult.

Erik-Jan, thank you for making everything so much easier. I can't wait for what's to come.

In memory of my sister Marie,
who in her own way steered me towards neuroscience.

SUMMARY

Humans in our environment are of special importance to us. Even if our minds are fixated on tasks unrelated to their presence, our attention will likely be drawn towards other people's appearances and their actions. While we might remain unaware of this attentional bias at times, various studies have demonstrated the preferred visual scanning of other humans by recording eye movements in laboratory settings. The present thesis aims to investigate the circumstances under and the mechanisms by which this so-called *social attention* operates.

The first study demonstrates that social features in complex naturalistic scenes are prioritized in an automatic fashion. After 200 milliseconds of stimulus presentation, which is too brief for top-down processing to intervene, participants targeted image areas depicting humans significantly more often than would be expected from a chance distribution of saccades. Additionally, saccades towards these areas occurred earlier in time than saccades towards non-social image regions. In the second study, we show that human features receive most fixations even when bottom-up information is restricted; that is, even when only the fixated region was visible and the remaining parts of the image masked, participants still fixated on social image regions longer than on regions without social cues. The third study compares the influence of real and artificial faces on gaze patterns during the observation of dynamic naturalistic videos. Here we find that artificial faces, belonging to human-like statues or machines, significantly predicted gaze allocation but to a lesser extent than real faces. In the fourth study, we employed functional magnetic resonance imaging to investigate the neural correlates of reflexive social attention. Analyses of the evoked blood-oxygenation level dependent responses pointed to an involvement of striate and extrastriate visual cortices in the encoding of social feature space.

Collectively, these studies help to elucidate under which circumstances social features are prioritized in a laboratory setting and how this prioritization might be achieved on a neuronal level. The final experimental chapter addresses the question whether these laboratory findings can be generalized to the real world. In this study, participants were introduced to a waiting room scenario in which they interacted with a confederate. Eye movement analyses revealed that gaze behavior heavily depended on the social context and were influenced by whether an interaction is currently desired. We further did not find any evidence for altered gaze behavior in socially anxious participants. Alleged gaze avoidance or hypervigilance in social anxiety might thus represent a laboratory phenomenon that occurs only under very specific real-life conditions. Altogether the experiments described in the present thesis thus refine our understanding of social attention and simultaneously challenge the inferences we can draw from laboratory research.

ZUSAMMENFASSUNG

Menschen in unserer Umgebung sind von besonderer Bedeutung für uns. Selbst wenn wir uns gänzlich auf eine Aufgabe konzentrieren, wird unsere Aufmerksamkeit voraussichtlich immer wieder auf Menschen in unserer Nähe und ihre Handlungen gelenkt. Auch wenn wir uns dieser bevorzugten Betrachtung anderer nicht jederzeit bewusst sind, haben diverse Labortstudien anhand von Augenbewegungen diese Präferenz für andere Menschen aufgezeigt. Im Rahmen der vorliegenden Dissertation werden die Mechanismen dieser *sozialen Aufmerksamkeit* und die Bedingungen, unter denen sie auftritt, untersucht. So zeigt die erste Studie, dass die bevorzugte Betrachtung anderer Menschen in komplexen Szenen ein reflexiver Prozess ist. Nach einer Darbietungszeit von 200 Millisekunden, welche zu kurz für das Auftreten von zielgerichteten Prozessen ist, landen Sakkaden signifikant häufiger auf Bildregionen, die einen Menschen abbilden, als man basierend auf einer Zufallsverteilung der Sakkaden erwarten würde. Zusätzlich treten diese Sakkaden zeitlich früher auf als Sakkaden, die auf nicht-sozialen Bildelementen landen. In der zweiten Studie wird anhand eines blickkongruenten Paradigmas deutlich, dass Menschen auch dann am meisten betrachtet werden, wenn nur der aktuell fixierte Teil des Bildes sichtbar ist. Dies bedeutet, dass die Präferenz sozialer Information auch dann erfolgt, wenn Blickbewegung intentional gesteuert werden. In der dritten Studie wird geprüft, inwiefern sich soziale Aufmerksamkeit bei verschiedenen Gesichtstypen unterscheidet. Künstliche Gesichter, die zu menschenähnlichen Gestalten oder Maschinen gehören, haben zwar einen signifikanten Einfluss auf das Blickverhalten der Betrachter – dieser ist jedoch deutlich geringer als der Einfluss von Gesichtern, die zu echten Menschen in den gezeigten Videos gehören. In der vierten Studie werden die neuronalen Korrelate des reflexiven sozialen Aufmerksamkeitsprozesses untersucht. Analysen der funktionellen Bildgebung deuten auf eine Rolle der primären und assoziativen Sehrinde in der Lokalisierung sozialer Elemente hin.

Gemeinsam helfen diese Studien zu klären, unter welchen Umständen die präferierte Betrachtung sozialer Merkmale in Laborumgebungen stattfindet und wie diese Priorisierung auf neuronaler Ebene erreicht werden könnte. Das letzte experimentelle Kapitel befasst sich mit der Frage, inwiefern die berichteten Laborergebnisse auf die reale Welt übertragen werden können. In dieser Studie befanden sich die TeilnehmerInnen in einem Wartezimmer-Szenario, in welchem sie mit einem Komplizen interagierten. Die Analyse der Augenbewegungen ergab,

dass das Blickverhalten stark vom sozialen Kontext abhängt und davon beeinflusst wird, ob eine Interaktion aktuell gewünscht wird. Darüber hinaus fanden wir keine Hinweise auf ein verändertes Blickverhalten bei sozial ängstlichen TeilnehmerInnen. Die in der Literatur zuvor berichtete angebliche Blickvermeidung oder Hypervigilanz bei sozialer Angst könnte somit ein Laborphänomen darstellen, das nur unter ganz bestimmten Bedingungen im realen Leben auftritt. Insgesamt ermöglichen die in der vorliegenden Arbeit beschriebenen Experimente somit ein vollständigeres Verständnis von sozialer Aufmerksamkeit. Gleichzeitig wird im finalen Kapitel aber auch deutlich, wie essenziell Vergleiche mit dem wirklichen Leben für die Schlussfolgerungen in der Aufmerksamkeitsforschung sind.

CONTENTS

INTRODUCTION

A few weeks ago, I took the train from Würzburg to Utrecht to catch up with my university friends. I had taught a seminar earlier that afternoon, which is usually a wonderful but also draining experience, and so by the time I got to the platform at Würzburg Hauptbahnhof, my limbs were weary. As I saw the train approaching, I was desperate to swiftly find a space for my luggage and a seat to myself. I climbed up the three steps onto the first coach of my choice and, to my dismay, immediately realized that I was far from the only passenger with this mission. With resignation, I pulled my trolley through the narrow passageway between the stalls, scanning the available options to my left and to my right. There were none. Instead of empty seats I saw people – an elderly man looking out of the window, plenty of young kids on their electronic devices, businessmen hectically filling in excel sheets, a mother preparing slices of an apple for her little toddler. With several other passengers behind me, I continued walking down the corridor and moved on to the next coach. Again, I saw faces – many of them looked just as tired as I was feeling in that very moment, some were smiling at their phones or at each other and yet others were turned towards the outside so I could not attempt to decipher their inner states. Just when I felt the train moving to continue its journey west, I finally spotted the empty seat I had been looking for. I lifted my trolley up the luggage rack, smiled at who would be my neighbor for the next couple of hours and dropped onto my seat in exhaustion.

This little story does not contain any surprising events, nor can it be considered particularly entertaining – what it does do, however, is point to a phenomenon we experience daily. When embarking on the train, my actual and only task was to find a seat. Invariably, the accomplishment of this job entailed the visual exploration of other human beings. In this particular case, of course, the occupied seats were taken up by other people so the scanning of faces, inviting or discouraging me to pick their neighboring seat, was inherently related to the task. Over the past decades, however, researchers have accumulated evidence that we prioritize fellow humans in the exploration of visual scenes independent of the task at hand. What has come to be termed social attention has been demonstrated using simplistic and more complex visual stimuli inside the laboratory and, thanks to technological progress, has recently been investigated outside of laboratory settings. A few questions about the nature and implications of social attention, however, remain. Is the prioritization of human beings a rapid, automatic process or a goal-oriented mechanism taking place relatively late in the visual processing stages? Was my attention towards the other passengers on the train an immediate reflexive response or did I purposely look at their faces to evaluate the benefits and costs of sitting close to these strangers? What are the neural correlates of this attentional bias towards fellow humans? Is there one specific brain region responsible for my attentional shift towards the other passengers or are there multiple brain areas involved together orchestrating the responses to social information? How are these behavioral and neural mechanisms of social attention altered in psychiatric disorders in which social interactions are impaired? Which differential scanning of other people would, for instance, patients with social anxiety display, when stepping onto the train? My doctoral thesis aims to investigate all of these questions while bearing in mind that our world is a complex one.

Social Attention

How can we go about studying a mechanism that is fundamental to every single one of our interactions with others without neglecting the detail and diversity that social encounters bring about? For decades, researchers have opted for simplistic experimental designs to investigate the basic processes underlying attentional shifts to social stimuli (Kingstone, 2009). The pioneering work for the field of social attention was conducted by the physiologist Alfred Yarbus. Using suction caps attached to the sclera of the eye, Yarbus discovered that scenes containing people primarily trigger fixations of these and that photographs of faces predominantly lead to excessive scanning of the eyes of these faces (Yarbus, 1967). These findings inspired many subsequent researchers to investigate the prioritization of eyes in face processing and the visual exploration of social scenes more generally. Traditionally, the collective term *social attention* can either refer to attention allocation towards other social stimuli or towards objects or regions that receive the attention of another human being, usually signaled by their gaze (Capozzi & Ristic, 2018; Gobel et al., 2015; Risko et al., 2016). Accordingly, increased social attention can refer to increased fixations on specific body parts of humans displayed on a screen. Another instance of social attention could yet also be my increased focus on the magazine the person next to me in the train is reading. While I will briefly discuss the insights gained through investigations of gaze perception, the empirical studies reported in this thesis understand *social attention* merely as attention to (often reflected in fixations on) fellow human beings. This definition of social attention still gives way to a wide variety of research questions about the specific instances which trigger the attentional shifts to humans in our environment.

The Attentional Dynamics of Gaze-Following

One subfield within the social attention literature, which has greatly influenced our understanding of face processing, focuses on gaze direction as an attentional cue (Frischen et al., 2007). As the eyes are highly expressive and can communicate a person's inner states, gaze shifts to a specific spatial location might trigger shifts in an observers' attention to the gazed-at location. Indeed, this behavior is not only observed in adults but already infants seem to understand that the parental eyes convey crucial information as they show a preference for eye contact from birth onwards (Farroni et al., 2002). A preference for face-like stimuli has even been demonstrated in the human fetus (Reid et al., 2017), although methodological problems in this experiment have been identified elsewhere (Scheel et al., 2018). The prioritization of direct gaze is possibly the first step towards full-fledged gaze following behavior which is evidenced by infant orienting to head turns of adults (Scaife & Bruner, 1975) and increased saccades into the direction that schematic faces are looking at (Hood et al., 1998). While these gaze-following observations were made in almost one-year olds, the tendency to look at whichever object is visually attended by other people was even detected in infants as young as three months old (Striano & Stahl, 2005) and, at five months of age, infants were seen to react with decreased attention and smile to very slight aversions of gaze (Symons et al., 1998). The sensitivity towards the gaze of others is thus a core aspect of a child's development and is thought to be a driving force in the progression of its social cognitive skills (Striano & Reid, 2006). It has been suggested that gaze following aids children in associating the looked-at object with the name called out by the other person, usually the caretaker, thereby fostering its language understanding (Dare A. Baldwin, 1995).

Not surprisingly, gaze-following continues into adulthood and has become the subject of a vast collection of literature investigating the different stimuli that elicit gaze-cueing and the mechanisms underlying these attentional shifts (Frischen et al., 2007). A paradigm that has been commonly used to examine the workings of gaze as an attentional cue is the Posner-cueing paradigm. In its original version, participants are asked to fixate on a cross on the middle of the screen which is surrounded by two squares to its right and its left (Posner, 1980). A cue, either in form of an arrow on top of the fixation cross or the highlighting of one of the neighboring boxes, is used to indicate at which location a target will appear. Reaction times are faster, and responses are more accurate for targets at a validly cued location than for targets appearing at the non-cued location. If gaze is indeed an attention-capturing stimulus, the use of schematic or photographed faces looking at either the right or left should result in a similarly improved performance to targets appearing at the looked-at location. As expected, in the first study investigating this phenomenon participants responded faster to letters that appeared at the location that was looked at by a simplistic cartoon face (Friesen & Kingstone, 1998). These results were soon after replicated with the use of photographs of real faces (Driver et al., 1999) confirming that the direction of gaze triggers attentional shift independent of the type of social stimulus material.

The analysis of reaction times and response accuracies while the fixation cross remains fixated only allows implications about *covert* attentional orienting which refers to an internal shift of attention without any obvious external markers revealing a switch in attentional focus (Posner, 1980). *Overt* attention, on the other hand, refers to attentional orienting in which sensory receptors are openly orienting to the stimulus of interest. Turning one's head towards a person telling a story at a busy event to more clearly understand the narrative or openly exploring the dishes displayed at a buffet table to make an informed decision about what to eat are everyday examples of overt attention. Nowadays overt attention is commonly investigated using eye-trackers which record eye movements (although saccades can also be decoupled from attentional shifts after hours of training , see Reeves & McLellan, 2020). Eye recordings revealed that directional gaze cues also elicit automatic overt attention shifts as illustrated by saccades made towards the cued location (Mansfield et al., 2003). Additionally, when instructed by an initial cue to look at a specific direction, a subsequent incongruent gaze cue decreased saccadic performance (Ricciardelli et al., 2002). Taken together, the studies investigating gaze cueing in adults suggest that both overt and covert attentional orienting are affected by the eyes of others (Friesen & Kingstone, 2003).

The Competition Between Social Information and Physical Saliency

While the numerous studies providing evidence for the attention-steering effects of gaze also support the more general idea that social stimuli capture attention, a shortcoming of the majority of gaze-cueing studies is the use of overly simplistic stimuli in on-screen experiments (Risko et al., 2012). These impoverished experimental designs are unable to address the question whether our attention is just as easily grabbed by faces if other interesting stimuli are presented alongside the human features. Our visual surroundings are complex, and the intricacies of our visual input might distract from the impact faces presented in isolation have on us. Another line of research within the field of social attention therefore focuses on whether social stimuli, in the shape of human heads or body parts, are more attention-capturing than other physically salient areas within a scene.

Traditional approaches to attention in psychology or neuroscience do not only differentiate between overt and covert but also *top-down* and *bottom-up* attention (Knudsen, 2007). While top-down attentional processes describe situations in which attention is purposely directed towards a specific object or task, bottom-up attention refers to the quick and automatic orienting we experience, for instance, when hearing a car behind us honk. Saliency-based models of attention suggest that attention is primarily captured by those elements of a scene that pop out by virtue of their physical features (Itti et al., 1998; Itti & Koch, 2000). By reading out the orientation, luminance and color of each pixel and its surroundings in an image, computational algorithms can create so-called saliency maps which highlight the areas with the highest physical saliency. When participants were asked to memorize the content of an image, these saliency-maps served as good predictors of fixation densities (Foulsham & Underwood, 2008; Itti et al., 1998). As they performed best for early fixations, it has been suggested that saliency predominantly models bottom-up attentional processes (Parkhurst et al., 2002; but see Tatler et al., 2005). Indeed, the relevance of image regions drastically changes depending on task demands and saliency-based models offer little explanatory power when top-down demands are strong (Tatler et al., 2011). Additionally, when participants were asked to freely explore a real or virtual environment instead of merely looking at a screen image, feature-based salience poorly predicted fixations (Jovancevic-Misic & Hayhoe, 2009; Jovancevic et al., 2006; Turano et al., 2003), suggesting that physical saliency is of greater relevance inside rather than outside of the laboratory.

How does the processing of social features fit into saliency-models of attention? Various studies have shown that when social information is included in an image, physical saliency has a considerably weak influence on gaze allocation (Birmingham et al., 2009b; Sue Fletcher-Watson et al., 2008; Nyström & Holmqvist, 2008; Zwickel & Vo, 2010). The inclusion of a person in a complex visual scene accordingly elicits gaze-following behavior reflected in more frequent visiting of objects gazed-at by the depicted human (Zwickel & Vo, 2010). Independent of gaze cues, humans in a scene also trigger comparatively more saccades and fixations towards them than towards other highly physically salient areas (Birmingham et al., 2009b). From a more computational standpoint, models using both information of low-level saliency and face-detection within scenes significantly outperformed models relying only on low-level saliency information in fixation predictions (Cerf et al., 2008). This finding was corroborated by various other studies that showed that the inclusion of social information (e.g. the position of a face in the scene) in the computational model significantly improves fixation predictions (Kankanhalli et al., 2014; Marat et al., 2013; Parks et al., 2015; Tatler et al., 2011). While growing evidence suggests that social information overrides the influence of low-level physical features, it is crucial to properly disentangle the respective contributions to reach a sufficient understanding of the mechanisms underlying our gaze. If the humans depicted in the scenes happen to also consist of highly salient physical features, it is difficult to properly interpret the heightened fixations they receive. Previous studies either failed to report physical saliency of social features, used social features which had comparatively higher physical saliency than remaining image regions or did not evaluate differences between social and other non-social image regions statistically.

To comprehensively assess the influence of low-level physical saliency and social features on gaze, End and Gamer presented complex naturalistic social or non-social scenes to participants while their eye movements were being recorded (End & Gamer, 2017). Importantly, stimulus selection ensured that social features were not confounded by high low-level physical saliency. An analysis of the first five fixations revealed that social features and heads in particular were heavily prioritized during image exploration from the first fixation onwards - even if more noticeable non-social elements were displayed in the scene. A linear mixed model analysis additionally confirmed significant influences of human heads and bodies, physical saliency and center bias on fixation densities. A similar approach was adopted by Rubo and Gamer to investigate the influences of physical saliency and social information in dynamic social scenes (Rubo & Gamer, 2018). They reported significant and similarly large influences of social features and physical saliency on gaze allocation, suggesting that social attention is a robust phenomenon generalizing from stationary to dynamic stimulation material.

We have thus come to understand that human features, when presented in isolation or embedded in complex scenes, readily grab our attention. As this prioritization is evident from the first fixation onwards (End & Gamer, 2017), it is likely that the attentional shift towards other humans is a reflexive phenomenon. However, top-down influences on these first fixations cannot be excluded as shifts from reflexive to voluntary attention presumably take place between 300 – 500 ms after stimulus onset (Müller & Rabbitt, 1989) which is also the time within which first fixations commonly occur. The first paper within the series of papers that make up this dissertation therefore systematically investigates whether social attention draws on reflexive mechanisms. By using an experimental design with a very brief stimulus presentation time (200 ms), a detailed visual exploration is made impossible and reflexive eye movements to social features can be properly assessed.

When we think back of my search for a seat on the packed train to Utrecht, it becomes obvious that social attention in daily life often encompasses more than a mere free exploration of our visual surroundings. An automatic attentional shift to the faces on the train might be overridden by my task-related focus on the available seating. The second paper of this dissertation therefore aims to unravel the interplay of bottom-up and top-down mechanisms on social attention by limiting the extent to which bottom-up information of the social stimuli is available. Lastly, the use of naturalistic static or dynamic scenes can inform us whether attentional prioritization of human features persists in competition with complex visual surroundings. It is, however, additionally interesting to investigate to what extent the nature of a social feature influences attention allocation. In the third paper of this dissertation, we presented dynamic scenes containing either exclusively real social features (i.e. actual humans moving through a video), exclusively artificial social features (i.e. a human-like statue or robot moving through the scene) or a combination of the two. This design allowed us to disentangle the respective influences of different types of faces in naturalistic scenes, adding to our general knowledge of the factors underlying gaze distribution in complex social scenes.

The Neural Mechanisms of Social Attention

While the delineation of the behavioral mechanisms of social attention is necessary to obtain a sufficient understanding of human gaze allocation, neuroscientists are particularly interested in the neuronal activation patterns that underlie or enable these shifts in attention. In the following section I set out to give an overview of the brain regions involved in the perception of social cues and the attentional shifts towards them. As electrophysiological recordings in non-human primates allow the examination of how single cells contribute to the processing of social information, I will start off by discussing the evidence provided by non-human primate studies for social attention-specific brain areas. The second part of this chapter will then address the results of human lesion and neuroimaging studies and we will come to see to what extent the human and the monkey social attention network can be considered homologous.

Evidence from Non-Human Primates

The macaque cortex comprises of more than 30 densely connected, hierarchically organized areas dedicated to visual processing including neurons which specifically respond to social clues (Felleman & Van Essen, 1991). Single-cell recordings in the anterior superior temporal sulcus (STS) of macaques showed that these cells respond selectively to specific orientations of gaze, heads or bodies (Perrett et al., 1992). While a small part of these cells reacted preferentially to specific faces, the majority responded to faces of all types but instead coded head orientation (front, back and profile face views) selectively, thereby contributing to what Perrett and colleagues call *perspective view* (Perrett et al., 1985). This perspective view, in turn, might enable the processing of social attention (Perrett et al., 1985). A small number of studies even pointed to neurons coding directions of gaze (De Souza et al., 2005; Perrett et al., 1985) and body orientation (Wachsmuth et al., 1994) providing further evidence that the STS is involved in the decoding of the attentional focus of conspecifics.

Another brain region assumed to be implicated in social attention and cognition is the amygdala (Rutishauser et al., 2015), two almond-shaped nuclei found in the medial temporal lobes of the brain and considered part of the limbic system. Evidence on the cytoarchitectonic organization within the amygdala suggests that the macaque amygdala can be considered a sensible proxy for the human amygdala (Freese & Amaral, 2009). In macaques, amygdala volume correlates with social network size (Sallet et al., 2011) and social status (Noonan et al., 2014). Additionally, amygdala volume is associated with the time macaques spend on fixating the eyes of others on a computer screen (Zhang et al., 2012) and amygdala lesions have been linked to reduced fixations of eyes (Dal Monte et al., 2015). Macaques with experimentally removed amygdalae also displayed a lack of appropriate emotional responses to novel or fear-inducing stimulus material (Capitanio et al., 2006). The amygdala’s contribution to the processing of novelty might also be related to its role in the brain’s reward system. Neurons in the macaque amygdala were seen to signal the anticipatory time course for both reward probability and magnitude confirming an involvement of the amygdala in timing processes of reward (Bermudez et al., 2012). Non-human primates rapidly learn to associate specific stimuli with either a positive or negative outcome and distinct cell populations within the amygdala responded exclusively to either positive or negative rewards (Belova et al., 2008; Paton et al., 2006; Sanghera et al., 1979). These observations thus support the idea of representations of value within the amygdala.

In recent years, an increasing amount of evidence suggested that many of the amygdala's functions might stem from its role in attentional processing. It is long established that the amygdala contributes to emotional arousal and vigilance (Davis & Whalen, 2001). Within the amygdala, the basolateral amygdala receives sensory input from the thalamus and cortex and then increases activation in those target areas which foster the association between stimulus and reward (Davis & Shi, 2000). Outputs to the central nucleus of the amygdala play a crucial role in heightened attention to stimuli paired with food (Holland & Gallagher, 1999). While it is generally accepted that the amygdala thus contributes to changes in vigilance and arousal, it was assumed until recently that this is accomplished in a non-spatial manner. Indeed, the amygdala is densely connected with the orbitofrontal cortex (OBF) which does not respond differentially to varying spatial input (Padoa-Schioppa & Assad, 2006) suggesting that the amygdala is primarily involved in non-spatial processing of stimuli. Evidence from human lesion studies, however, points to a potential role of the amygdala in the allocation of spatial attention (Phelps & LeDoux, 2005). Patient S.M., who suffered from complete amygdala calcifications due to Urbach-Wiethe disease, is unable to recognize fearful faces but overcomes this deficit when explicitly instructed to fixate the eyes of the displayed faces (Adolphs et al., 2005). This observation suggests that the amygdala might not code fear recognition *per se* but instead regulates attentional shifts towards highly aversive or appetitive stimuli (Peck et al., 2013).

To investigate whether the amygdala is truly involved in the processing of motivational relevance and spatial information of stimuli, Peck and colleagues trained monkeys in two tasks in which reward-predictive visual cues were presented at different locations while recording single-cell activations in the amygdala (Peck et al., 2013). In both tasks two differently colored, briefly presented cues indicated at which location a higher liquid reward could be expected. Subsequently, in the first task, the monkey had to make a saccade to a single Gabor patch appearing at one of the two initially cued locations. In the second task, two Gabor patches appeared at both cued locations and, after a brief delay, changed in orientation. A subsequent final cue quizzed the monkey on the change of orientation of one of the by then vanished patches. The electrophysiological recordings revealed that amygdala responses encoded both value and spatial information of the presented stimuli. Neurons that responded more strongly to higher rewards also reacted most pronounced when this high value reward-predictive cue was presented contralaterally. This finding suggests that amygdala activity contributes to spatial cognitive processes potentially by enhancing the attentional resources that are directed towards a certain stimulus.

At the same time, however, these observations beg the question of how the encodings of value and space within the amygdala are created. As the amygdala receives direct information from the ventral visual stream (Freese & Amaral, 2009), it could be a suitable candidate region for the convergence of “what” and “where” information of the processed stimuli. It is, however, less clear from where the amygdala receives spatial information. The traditional spatially selective regions that make up the dorsal processing pathway (the frontal eye fields (FEF), dorsal striatum or parietal cortex) have little to no direct projections to the amygdala (Freese & Amaral, 2009). It has been suggested that the amygdala receives subcortical input from the superior colliculus, carrying low spatial frequency information, via the pulvinar nucleus of the thalamus which might contribute to rapid processing of emotional information (Tamietto & de Gelder, 2010). This fast but coarse processing pathway circumvents the slower but more detailed cortical processing route. It remains unclear whether the amygdala response to visual information with a latency of more than 100 ms as observed in the monkeys studied by Peck and colleagues comes about through subcortical or cortical inputs. Generally, responses to visual stimuli that surpass 100 ms are more in line with a cortical processing pathway (Pessoa & Adolphs, 2010).

Peck and colleagues instead suggest that inferotemporal cortical neurons might project sufficient spatial information to the amygdala to account for its observed spatial selectivity (DiCarlo & Maunsell, 2003; Peck et al., 2013). However, as the timing of visual and value information differs in the amygdala, it is more likely that different sources contribute to the two types of processing. Another possible candidate for the projection of spatial information are areas within the frontal cortex which also show anatomical connections with the amygdala although with varying densities (Freese & Amaral, 2009; Ghashghaei et al., 2007). This idea is corroborated by studies showing that spatial information encoding in dorsolateral (DLPFC) and ventrolateral prefrontal cortex (VLPFC) is enhanced by reward (Kennerley & Wallis, 2009). The VLPFC has been implicated in the attentional reorienting response to novel relevant objects (Corbetta et al., 2008) and, although DLPFC and VLPFC show only weak connections to the amygdala (Ghashghaei et al., 2007), signaling between these prefrontal areas and the amygdala might enable reorienting in response to emotional relevance. In terms of connectivity, the amygdala is more strongly interconnected with the ventral anterior cingulate cortex (ACC) (Freese & Amaral, 2009). A small group of neurons within the dorsal ACC were recently seen to code both space and value of objects (Kaping et al., 2011). The connections between amygdala and dACC are, however, comparatively stronger from amygdala to dACC suggesting that the combined representation of value and space within the dACC might rely on input from the amygdala rather than the other way round (Ghashghaei et al., 2007; Peck et al., 2013). As the amygdala also shows reciprocal connections with the basal forebrain (Russchen et al., 1985) and the latter widely connects to vast parts of the cortex (Mesulam et al., 1983), it has recently been suggested that the amygdala and basal forebrain constitute a common pathway for reward oriented attention (Peck & Salzman, 2014). Here, the basal forebrain might receive amygdala input which it then uses to modulate attention (Broussard et al., 2009). How exactly the amygdala arrives at the convergence of value and space information yet remains to be elucidated.

The evidence described above nonetheless suggests that the amygdala might constitute a core component of the redirection or enhancement of social attention. When several stimuli compete for attention, the amygdala might contribute to the prioritization of social stimuli by pairing the encoding of their location with their motivational significance (Maunsell, 2004). Alternatively, the amygdala might contribute to attentional reorienting by means of increasing vigilance promoting arousal (Davis & Whalen, 2001) which is the more traditional understanding of the amygdala's working (Belova et al., 2007; Kapp et al., 1994). As the activity of reward-predictive neurons differed depending on where stimuli were presented (Peck et al., 2013), these results yet speak in favor of the amygdala's involvement in the allocation of attention rather than a general increase in vigilance which should in theory yield equal activation independent of stimulus location. While the study of non-human primates is certainly very informative regarding anatomical connections of neural areas and functional activity within cell groups or even single cells, it is also vital to assess to which extent the observations made in monkeys can be translated to the human brain. What evidence can be found in the human brain to support the notion of amygdala activity as the source of attention allocation to social stimuli? Do humans also possess neuronal areas that respond specifically to different types of gaze or head directions?

Gaze Perception in the Human Brain

When humans are presented with photographs of faces, functional brain imaging studies have consistently demonstrated the activation of the lateral fusiform gyrus (Halgren et al., 1999; Haxby et al., 1999; Kanwisher et al., 1997; McCarthy et al., 1997). This region, commonly termed fusiform face area (FFA), demonstrates strong face-selectivity and thus seems to play a key role in the processing and perception of faces. Even face-related gaze patterns, occurring without the presentation of an actual face, were seen to yield reliable FFA activations (Wang et al., 2019). The idea that face perception involves specialized neural structures garnered further support from the neurological condition of prosopagnosia. Prosopagnosic patients can no longer successfully distinguish faces from one another even though they can use other sensory input, such as voices, for person identification. Interestingly, the critical lesion site of prosopagnosia is close in space to the FFA (Kanwisher et al., 1997). While the lateral fusiform gyrus is thus assumed to code identity-invariant facial features crucial to person recognition, the superior temporal sulcus mediates the representation of changeable aspects of faces, such as eye gaze and lip movement (Haxby et al., 2000).

The distinction between two separate neural mechanisms, one engaged in face recognition itself and the other in variant facial features crucial to social communication, is supported by studies investigating adaptation effects in gaze. Calder and colleagues showed that adaptations to direct gaze yield significant aftereffects which argue for a multi-channel coding of gaze perception (Calder et al., 2008). Multi-channel coding models of gaze suggest that gaze is coded by separate channels for leftward, rightward and direct gaze, integrating the single-cell findings in primates which showed that different cells respond to different gaze orientations (Perrett et al., 1985). The effortless distinction between different facial identities had previously been shown to be mediated by an opponent-coding model (Leopold et al., 2001; Tsao & Freiwald, 2006) in which differential features are represented along one axis. This model can best be illustrated by the idea of each face containing an anti-face where all features are maximally different and intermediate features are plotted along one feature-axis. If gaze was also coded by an opponent model, direct gaze should represent the equilibrium between left and right or upward and downward gaze and yield no aftereffects in itself. Calder and colleagues yet showed that direct gaze also elicits biases in judging leftward and rightward gaze thus providing further support for the distinction between facial identity and gaze processing in the human brain (Calder et al., 2008).

Electroencephalographic (EEG) recordings have also contributed to the debate on the processing of invariant and variant facial features. Accordingly, scalp electrodes have measured a face-selective electrophysiological response to faces occurring already 170 ms after stimulus onset (Bentin et al., 1996), whereas event-related responses to gaze adaptation take place considerably later in time (~250-350 ms) (Jacques et al., 2007). Recently, the amplitude of the N170 response was yet seen to depend on fixations of the eyes of the presented faces (Itier et al., 2007; Parkington & Itier, 2018). Moreover, it has been argued that the N170 does not selectively respond to faces but instead signals general object expertise and that the human species is simply extraordinarily good in differentiating faces (Rossion et al., 2002; Tanaka & Curran, 2001). However, concrete evidence substantiating such a claim by showing N170 responses to both faces and objects of expertise is currently lacking leading researchers to believe that the N170 remains a face-selective response (Schultz et al., 2003).

As already illustrated by the case of prosopagnosia, a great deal about the functioning of the human brain can be learned from lesion studies. While the complete removal of the STS in monkeys caused impaired gaze but intact face perception (Heywood & Cowey, 2006), a human patient (MJ), missing parts of the right superior temporal gyrus (STG) displayed altered gaze discrimination as a result of a rightward bias (Akiyama et al., 2006). She was consequently more likely to categorize a leftward-looking face as rightward-looking. Considering that damage in the right STG is also linked to spatial neglect (Karnath et al., 2001), the observed difficulties in the detection of gaze direction might stem from a more general impairment in spatial attention. Consequently, gaze processing might depend on the interplay between the broader attention network and the STS, responsible for coding the representations of gaze (Haxby et al., 2000).

While research in primates and lesion studies point to a contribution of the STS in social attention, human neuroimaging studies have demonstrated the involvement of a much vaster network of neural regions. The STS was again identified as processing gaze direction but most fMRI studies found significant activation in the posterior instead of the anterior parts of the human STS (Allison et al., 2000). However, the results of these investigations have been inconsistent – one study found increased activation in the posterior STS for mutual gaze (Pelphrey et al., 2004), whereas others reported increased activity for gaze aversion (Engell & Haxby, 2007; Hoffman & Haxby, 2000) and yet others did not detect differences between gaze types (George et al., 2001; Kawashima et al., 1999; Wicker et al., 1998).

To get a more informative overview of the neural regions contributing to gaze processing, Nummenmaa and Calder used activation likelihood estimation on fMRI and PET studies to investigate the reliability of brain areas activated by gaze across studies (Nummenmaa & Calder, 2009). As expected, the meta-analysis demonstrated not only the contribution of the posterior STS to gaze processing but also of parts of the dorsal attention network which is associated with goal-directed, top-down attention (Corbetta et al., 2008). Involvement of the human MT/V5 was also statistically reliable which is most known for its role in motion processing but also contributes to the perception of facial characteristics (O'Toole et al., 2002). The authors furthermore suggest that the anterior STS might contribute to the processing of individual gaze directions (Rhodes et al., 2005), whereas subcortical structures, namely the amygdala and the hippocampus, might be responsible for monitoring gaze contact (George et al., 2001; Hooker et al., 2003; Kawashima et al., 1999). Gaze also reliably activated medial prefrontal regions (Calder et al., 2002; Williams et al., 2005) which have previously been implicated in theory of mind processes, such as deciphering the inner states of others (Gallagher & Frith, 2003). These activation patterns could be a result of analyzing why another person is looking at a specific object or what the potential intention of the gaze direction might be (Nummenmaa & Calder, 2009). Taken together, the processing of gaze recruits a vast network involving areas implicated in emotional and social cognition (mPFC and amygdala), temporal areas involved in face perception (superior parietal lobule, frontal eye fields) and regions within the dorsal attention network.

The Putative Neural Mechanisms Enabling Social Attention

There is a myriad of information on the brain regions implicated in the processing of eye gaze and the vast neural network recruited by gaze suggests that differential aspects of the processing of social information are distributed distinctly across the brain. More pertinent to the question of how social attention emerges in the brain is the examination of those areas that generally respond to social versus non-social stimuli – independent of gaze directions. When participants are presented with an ambiguous stimulus that can either be interpreted as social or a geometric directional cue (eyes with a hat or a car facing to one side), the right STS responded solely in those trials in which the stimulus was perceived as social (Kingstone et al., 2004). A similar design employed by Tipper and colleagues found selective activation of occipital and frontal areas for the perception of the cue as social (Tipper et al., 2008). How do these selective activations come about? Which brain regions contribute to the recognition of social elements in a scene?

Ever since Corbetta and Shulman published the influential research article proposing two anatomically and functionally distinct pathways of attention more than a decade ago (Corbetta & Shulman, 2002), we have come to think of attention as mediated in either a bottom-up or top-down fashion. Within this distinction, the ventral pathway – spanning from the primary visual cortex over the temporoparietal junction (TPJ) to the ventral frontal cortex – might be of particular relevance to the automatic reorienting to social elements in a scene. The dorsal pathway, on the other hand, is most likely recruited when attention is deliberately navigated towards social features. The intraparietal sulcus (IPS) and frontal eye fields (FEFs) are crucial nodes of the dorsal attention network and both contain retinotopic spatial maps (Silver & Kastner, 2009) therefore rendering them likely candidate regions for saccade planning and the maintenance of spatial relevance maps (Jerde et al., 2012). Numerous electrophysiological studies confirmed that attended stimuli trigger larger responses in visual cortices than unattended stimulation material (e.g. Reynolds et al., 2000; for a review see Ungerleider, 2000). It has therefore been suggested that attention-capturing stimuli are more likely to receive cortical representation (Desimone & Duncan, 1995). The STS activation observed when stimuli are recognized as social is hence in line with the goal-oriented attentional processing mediated by the dorsal stream. However, visual representations travel along the cortical visual pathway comparatively late (~145–170 ms) (Silverstein & Ingvar, 2015) and it therefore stands to question whether the rapid prioritization of social stimuli is not achieved elsewhere.

A traditional view of emotion encoding in the brain suggests that highly relevant fear cues are rapidly processed by a subcortical pathway from the superior colliculi and pulvinar nuclei of the thalamus to the amygdala (Day-Brown et al., 2010). This so-called 'low-road' model of fear processing therefore offers an alternative to the more detailed but slower visual representations mediated by the cortical visual pathways (Tamietto & de Gelder, 2010). Evidence for a subcortical processing route yet largely stems from research in rodents (LeDoux, 1998), whereas neuroimaging studies in humans have only provided indirect support for the low-road model by assessing amygdala activation in response to subconsciously processed emotional stimuli (Garrido et al., 2012; Johnson, 2005; Morris et al., 1998; Whalen et al., 1998). Patients with cortical blindness who exhibit sustained processing of fear-relevant stimuli provide equally indirect but compelling evidence for a subcortical alternative processing route. These unseen fear cues lack representation in the visual cortices and therefore do not reach conscious awareness but are processed sufficiently enough elsewhere for cortically blind patients to be able to discriminate them from non-fear cues (Morris et al., 2001; Pegna et al., 2005). This rapid and subconscious processing might be mediated by connections between pulvinar and amygdala as indicated by quantitatively and qualitatively altered connectivity between these regions after damage to primary visual cortices (Tamietto et al., 2012). Various electrophysiological studies yet failed to provide direct support for early fear responses in the human amygdala (Krolak-Salmon et al., 2004; Oya et al., 2002). Méndéz-Bértolo and colleagues therefore examined intracranial electrophysiological amygdala responses to low versus high-spatial frequency emotional faces (Méndez-Bértolo et al., 2016). They observed short-latency (~ 74 ms) reactions to fearful low-spatial frequency stimuli in the amygdala but not to stimuli displaying other emotions, other spatial frequencies or photographs of arousing scenes. This unique finding suggests that the fast amygdala response is selective to coarse socially relevant stimulation. A magnetoencephalography study capitalized on these results and used dynamic causal modelling to demonstrate that the pulvinar-amygdala route is the most likely network underlying face and emotion discrimination – independent of emotion type and spatial frequency (McFadyen et al., 2017).

In line with the idea of the amygdala's involvement in the detection of socially relevant information, patient SM, suffering from complete bilateral amygdala lesions, fails to successfully detect fearful faces or threatening stimuli (Adolphs et al., 1994). This shortcoming was later understood to stem from a more general failure to fixate the eyes of faces (Adolphs et al., 2005) leading to particularly poor performance in fear discrimination as eyes are the most informative features of a fearful faces (Smith et al., 2005). When specifically instructed to fixate the eyes, patient SM was consequently able to successfully discriminate fear from other emotions (Adolphs et al., 2005). The amygdala might therefore represent a driving force in the allocation of attention to socially relevant elements of a scene. Indeed, an fMRI study by Gamer and Büchel showed that amygdala activity is correlated with saccades made towards the eye region of fearful faces (Gamer & Büchel, 2009).

Taken together, the evidence of both non-human and human primate data suggests that a vast neural network, comprising subcortical and cortical nodes, is engaged in the processing of social stimuli. Within this network and along the traditional line of thought of attention as divided into bottom-up and top-down processes, the dorsal visual stream likely contributes to task-relevant, goal-directed attention to social elements. The rapid automatic orienting towards social features might instead be mediated by amygdala activity receiving input from subcortical superior colliculi to pulvinar pathway. The fourth experimental chapter of this thesis aimed to investigate whether reflexive social attention is modulated by amygdala activity. We therefore recorded fMRI data while showing social and non-social images for a very brief presentation time (200 ms) to participants. The resulting images of brain activation were then examined by performing a multi-voxel pattern analysis. As most neuroimaging studies to date have used general linear models to evaluate social attention, the refined multivariate approach allows the detailed examination of brain activity patterns (Haynes & Rees, 2006; Kriegeskorte & Kievit, 2013) rendering it possible to notice not only changes in net activation but in the differential activation patterns enabling social attention.

Social Attention in Psychiatry

Fundamental research, such as the investigation of the neural mechanisms underlying social attention, often triggers questions about its potential applications and its actual value to society. When we turn our focus towards psychiatric disorders in which social interactions are impaired, the relevance of the fundamental research on social attention becomes evident. Among all DSM-5 classifications, social attention is arguably most prominently affected in autism spectrum disorder, characterized by difficulties in interpersonal communications and interactions, and social anxiety, marked by a strong fear of social encounters or evaluations by others. The present chapter will describe the evidence for altered social attention in both disorders, which will emphasize the clinical significance of the more fundamental articles presented in this thesis.

Altered Social Attention in Autism Spectrum Disorders

Empirical as well as anecdotal evidence suggests that eye contact is heavily reduced in autism (Dawson et al., 1998, 2004) and in many patients to such a pervasive degree that the integration into social settings is impeded (Mundy & Newell, 2007). Autism is officially recognized as a psychiatric disorder since the 1940s (Kanner, 1943). It is nowadays considered part of a spectrum of chronic, highly heritable disorders which have an early-life onset. Typical symptoms include but are not limited to deviations in attentional allocation and impairments in communication and social interactions (Wing & Gould, 1979). While diagnosis rarely occurs before three years of age, differential behavioral measures can already predict a later development of autism spectrum disorder (ASD) in infants (Dietz et al., 2006; Osterling & Dawson, 1994). Its familial occurrence clearly points to a biological fundament of the disorder and, in the past decade, the advance of the field of genetics has also garnered support for various copy number and single nucleotide variants implicated in ASD (Woodbury-Smith & Scherer, 2018). Nevertheless, given the vast heterogeneity of ASD, it is difficult to characterize its behavioral symptoms and the underlying neurobiology.

One predominant theory which aims to explain social deficits in ASD is the social motivation hypothesis (Chevallier et al., 2012; Dawson, 2008; Grelotti et al., 2002;

Stavropoulos & Carver, 2013). Here, the underlying idea is that patients with ASD consider social interactions as less rewarding and therefore lack the motivation to socially engage. When presented with social stimuli, brain regions typically associated with reward processing are less activated in ASD compared to healthy controls (Kohls et al., 2013; Scott-Van Zeeland et al., 2010). While it is unclear whether the altered reward processing in ASD uniquely concerns social rewards or is part of a more general reward processing deficit (Dichter et al., 2012), it is assumed to be the cause of later emerging symptoms typically observed in ASD including communication difficulties (Tager-Flusberg et al., 2005), abnormal neural processing of faces (Kleinhans et al., 2008; McPartland et al., 2004) and impairments in joint attention (Loveland & Landry, 1986; Mundy et al., 1986). Among the intricate neuromodulatory system of the human brain, oxytocin is the neuropeptide which is most prominently implicated in inter-human relationships (Baskerville & Douglas, 2010). A systematic literature search highlighted that oxytocin administration can lead to improved performance in social cognition tasks in ASD patients (Stavropoulos & Carver, 2013) including recognition of affective speech (Hollander et al., 2007) and emotion recognition (Guastella et al., 2010). Stavropoulos and Carver have therefore suggested combining behavioral interventions with intranasal oxytocin administration to increase the intrinsic motivation of patients with ASD to engage in social interactions (Stavropoulos & Carver, 2013).

A different account of the development of autism focuses on attentional rather than motivational deficits which purportedly deprive the autistic child of social input early in life further impeding normal brain development (Mundy & Neal, 2000) but see (Johnson, 2014). Neonates typically exhibit preferences for social stimuli as early as 13 to 168 hours (Farroni et al., 2005) or six weeks (Striano & Reid, 2006) after birth and more complex social responses, such as head-turning when one's name is called, usually emerge at 5-7 months of age (Morales et al., 1998). Retrospective analyses of home videotapes of 1st birthdays yet revealed that infants with ASD show deficits in social orienting; they orient less to other humans in the surrounding and fail to respond to their names (Osterling & Dawson, 1994). Decreased attending to others was also observed in 20-months old children with ASD - instead of fixating on social information, these children spend more time on fixating surrounding objects (Swettenham et al., 1998). Proponents of attentional origins of ASD assume that if these deficits were promptly diagnosed and targeted by therapeutic interventions, the progression of the disorder could potentially be slowed or stopped (Dawson et al., 2004; Jones & Klin, 2013). Whether this social orienting impairment is a symptom of a more general attentional processing deficit in ASD yet remains debated (Dawson et al., 2004).

It is hence of great interest to researchers to assess the attentional deficits exhibited in ASD in detail to investigate to which contextual settings and stimuli they apply. One caveat of social attention research in ASD is that many researchers claim to investigate the same question while using completely different experimental design and studying different age groups of patients. While some researchers focus on how patients with ASD react to their mothers in real life by conducting live observations or retrospective video analyses, others study how adults diagnosed with ASD react to photographs of faces on a laboratory PC screen. A systematic review of the literature on social attention in autism has therefore focused on three different aspects of alleged social attention deviations separately for children and adults with ASD (Guillon et al., 2014) – social orienting, the exploration of faces and gaze or head following. The hypothesis that patients with ASD are less likely to orient toward relevant social information in a scene was largely confirmed in young children. Accordingly, infants who are later in life diagnosed with ASD attend less to a social scene than their peers (Chawarska et al., 2013), young children with ASD do not show a preferential processing of biological motion (Falck-Ytter et al., 2013) and overall children with ASD look less at faces than control groups (Hosozawa et al., 2012; Nakano et al., 2010; Shic et al., 2011; von Hofsten et al., 2009). Only one study reported an equal pop-out effect of faces in ASD and control infants suggesting that attentional differences might be contingent on stimulus material (Elsabbagh et al., 2013).

In infants later diagnosed with autism results are mixed which, again, is probably caused by the different types of stimulus material used in the different studies. While Elsabbagh and colleagues failed to observe group differences in the viewing of static faces (Elsabbagh et al., 2013), Chawarska and colleagues reported decreased face processing in the ASD group while watching a video of an actress who also occasionally addressed the observers directly (Chawarska et al., 2013). Possibly the incommunicative nature of the static stimuli led the control group to respond relatively little to them rendering group comparisons insignificant (Guillon et al., 2014). In adults or older children with ASD, two out of five studies using static stimuli involving usually only one person, also did not report any differences in the time spent on attending faces (Fletcher-Watson et al., 2009; Freeth et al., 2010). When researchers compare the target of the first fixation, social orienting seems to be preserved but slightly delayed in ASD (Elsabbagh et al., 2013; Fletcher-Watson et al., 2009; Wilson et al., 2010). Overall, the elementary capability of orienting to social features does not seem to be impaired in ASD as is also evidenced by 2-6 months old babies who do not show any differences in viewing preferences of dynamic faces but develop ASD later in life (Jones & Klin, 2013). Interestingly, the same babies do show decreased face processing 2 months after the initial data acquisition (Jones & Klin, 2013). These results suggest that social orienting is not initially perturbed in ASD but differences in processing of social information emerge later in life (Johnson, 2014). It is therefore unlikely that deviations in social orienting stem from alterations in a subcortical processing pathway assumed to mediate rapid processing of faces (Johnson, 2005) but differences might arise in the later developing connectivity patterns between subcortical and cortical regions (Senju & Johnson, 2009).

Guillon and colleagues further found little evidence that ASD is associated with excessive mouth and reduced eye gaze. The ratio of eye to mouth gaze during the observation of an engaging partner does not predict clinical outcome of children with ASD (Chawarska et al., 2013; Elsabbagh et al., 2014; Merin et al., 2007; Shic et al., 2014; Young et al., 2009). Rather than being predictive of poor clinical development, increased mouth fixations are associated with better language outcome at 24 and 36 months irrespective of later clinical diagnosis (Elsabbagh et al., 2014; Young et al., 2009). In contrast, reduced gaze-following in ASD is largely supported by the literature. While no distinguishable differences in referential gaze are discernible in 6-10 months olds either at low or high-risk for ASD, 11-18 months old who are later diagnosed with ASD look significantly less at the gazed-at object than their peers who do not develop the disorder (Bedford et al., 2012). This finding suggests that infants prone to ASD are either unable to properly infer the meaning of another person's gaze or simply do not consider it as relevant in the second year of their lives. Freeth and colleagues, however, found increased fixation durations on gazed-at objects in adolescents with ASD compared to their typically developing peers (Freeth et al., 2011). They used naturalistic photographs in which a model either looked straight into the camera or at another object in the scene. Adolescents with ASD displayed more saccades navigating between the model and the gazed-at object on which they also fixated longer. An investigation of the temporal distribution of fixations yet revealed that typically developing adolescents immediately focused on the gazed-at object reflecting its social salience whereas the ASD group did not exhibit any early preferential viewing of the gazed-at object. Supporting the notion of reduced gaze-following in ASD, studies examining adults instead of children or teenagers also reported reduced gaze-following in ASD (Fletcher-Watson et al., 2009; Riby et al., 2013).

Although observations slightly differ regarding the details of gaze alterations between ASD and control groups, the majority of studies investigating gaze in ASD provides support for altered social attention as a crucial marker of the disorder. Overall, it becomes evident that researchers need to pay careful attention to choices of stimulation material and analysis types. As illustrated by the findings of Freeth and colleagues (Freeth et al., 2011), microstructural analyses focusing on temporal gaze dynamics might be more informative than macrostructural examinations such as total fixation duration on socially relevant elements of a scene. Additionally, the distinctive reactions to different types of stimuli in ASD and healthy control groups emphasize the need to properly characterize the behavioral mechanisms of social attention in the healthy population to, in turn, be able to assess deviations in clinical populations.

Avoidance of or Hypervigilance Towards Social Information in Social Anxiety

While there is still much research to be done to determine the extent to which social attention is altered in ASD, social interactions are also noticeably impaired in other psychiatric disorders – one of the most prominently affected one being social anxiety. Social anxiety disorder is marked by a persistent fear of social situations which could potentially have embarrassing or humiliating consequences (American Psychiatric Association, 2013). As a result, affected individuals often avoid social contexts entirely which, in turn, leads to a deterioration of their symptoms (Antony et al., 2008). It is the most common type of anxiety disorder and comes third in lifetime prevalence among all psychiatric disorders (APA, 2013). Overall, the prevalence of social anxiety is estimated to be as high as 10 – 15 % in the United States (Merikangas et al., 2010) and Europe (Essau et al., 1999). In the absence of treatment, social anxiety often turns chronic (Ginsburg et al., 2014; Wittchen et al., 1999), impeding the development of social networks and relationships which constitute a crucial protective factor against other mental health issues (Deater-Deckard, 2001).

Considering the prevalence and poor long-term outcome of social anxiety, a large body of research has provided different theoretical models attempting to explain the symptoms of the disorder. One influential model developed by Clark and Wells suggests that individuals suffering from social anxiety use attentional avoidance as a safety mechanism to prevent emotional distress or humiliation (Clark & Wells, 1995). Social interactions of socially anxious patients are therefore accompanied by gaze avoidance or complete avoidance of other individuals as an attempt to reduce the chance of negative evaluative feedback. Safety behaviors of this type yet commonly leave the affected to appear disinterested or aloof and prevent them from experiencing disconfirmation of their fears thereby further perpetuating anxiety symptoms (Hofmann, 2007). The strategic evasion of social interaction partners is therefore assumed to be acute during social-evaluative situation and serves to maintain the social anxiety disorder (Clark, 2005; Clark & Wells, 1995).

Another influential model of social anxiety, largely paralleling the model of Clark & Wells, is the *cognitive behavioral model* (Heimberg et al., 2010; Rapee & Heimberg, 1997). This model presupposes that a combination of predisposing factors, including traumatizing life events and behavior, lead the affected individuals to believe that their social environment is hostile and uncontrollable. In social-evaluative situations, patients with social anxiety disorder therefore excessively attend to socially threatening information. This selective attention towards external markers of negative evaluation then brings about various adverse effects which perpetuate the anxiety symptoms, including negative impressions of the social environment and negative updating of one's self-image.

Since both of the most influential models of social anxiety suggest selective attention as a key mechanism of the disorder, a myriad of experimental work has been dedicated to carving out the specific instances and patterns under which selective attention arises and functions. Using reaction time-based dot-probe tasks, studies have provided conflicting results. Some have found evidence for attentional avoidance of emotionally relevant stimuli (Chen et al., 2002; Mansell et al., 1999), whereas others reported increased attention to threatening social stimuli (Asmundson & Stein, 1994; Gilboa-Schechtman et al., 1999), particularly in the early stages of attentional processing (Mogg et al., 1997; Mogg et al., 2004). It has therefore been suggested that selective attention in social anxiety might be best described by an early hypervigilance towards social stimuli followed by a strategic avoidance (Chen & Clarke, 2017; Mogg et al., 2004).

As manual reaction time measures are unable to completely capture the time course of attentional bias, it is vital to consider evidence from eye-tracking studies when evaluating a hypothesis concerning attention. Eye-tracking measures allow for a continuous monitoring of overt attention, ranging from 60 up to 2000 Hz of sampling frequency depending on the device, and thereby feature a closer relationship to attention than manual button presses. At first glance, various eye-tracking studies support the notion of hypervigilance in social anxiety. When presented with photographs of faces, socially anxious participants correspondingly showed increased initial dwell times on threatening (Boll et al., 2016; Buckner et al., 2010; Lazarov et al., 2016; Seefeldt et al., 2014) or all social stimuli (Garner et al., 2006; Wieser et al., 2009). Other studies have yet reported reduced initial orienting to social stimuli (Byrow et al., 2016; Byrow et al., 2016b; Gamble & Rapee, 2009), whereas others did not observe any differences in orienting between socially anxious and control participants (Chen et al., 2012; Schofield et al., 2013; Singh et al., 2015). Similarly, when considering later stages of attentional processing results remain mixed with studies reporting both hypervigilance (Liang et al., 2017; Schofield et al., 2013) and avoidance towards social stimuli (Byrow et al., 2016; Shechner et al., 2017). Current eye-tracking studies consequently provide little support for the vigilance avoidance model of social anxiety (Mogg et al., 2004).

Which factors could potentially account for these diverging results? It is plausible that symptom heterogeneity, in particular the use of non-clinical samples of social anxiety, have contributed to the mixed observations. However, even when we exclusively consider studies examining clinical patients, no clear trends in gaze patterns emerge. Initial orienting is still associated with vigilant (Seefeldt et al., 2014; Shechner et al., 2013; Stevens et al., 2011) as well as avoidant gaze behavior (Byrow et al., 2016b; Gamble & Rapee, 2009). Similarly, clinical social anxiety patients again demonstrate mixed gaze patterns featuring both vigilance (Lazarov et al., 2016) and avoidance (Byrow et al 2016b; Chen et al., 2012; Schofield et al., 2013; Shechner et al., 2017). Differences in gaze might instead be explained by the heterogeneity of the stimulus materials. Accordingly, initial orienting differences predominantly occurred in studies which used competing stimulus pair presentations (Chen & Clarke, 2017) and were not detected in experiments which used vaster stimulus arrays (Lange et al., 2011; Liang et al., 2017; Shechner et al., 2017). This suggests that attentional biases in social anxiety might heavily depend on the displayed stimuli and reported differences in initial orienting might be eliminated when various stimuli compete with each other (Armstrong & Olatunji, 2012). At the same time, different results hinging on choices of experimental design raise the question whether attentional biases also occur in complex everyday contexts. To date, the majority of studies have either used stimuli that are static (Horley et al., 2003; Moukheiber et al., 2010) or dynamic videos that are simulated by actors (Weeks et al., 2013) or virtual agents (Wieser et al., 2009). These experimental designs carry important methodological constraints as they lack many components a real social encounter encompasses. When photographs of faces or virtual agents are observed on a screen, there is no room for a genuine interaction between participant and the presented person. The participants are therefore likely aware that their behavioral responses are not registered by their opponents and that their choices will have little effect on the given situation. Social anxiety-related responses might therefore be diminished as no evaluation of others will follow (Heimberg et al., 2010).

While the examination of gaze in social anxiety is essential to clarify in which ways altered gaze patterns might contribute to the emergence and maintenance of phobic behavior, social anxiety is also characterized by strong physiological reactions to social situations. A meta-analysis revealed a moderate effect of increased cardiovascular activity in socially anxious individuals (Patterson & Ritts, 1997). These increases in heart rate were triggered by a variety of different experimental provocations including public speaking prompts (Boone et al., 1999), low evaluative threat (Gramer & Saria, 2007), video clips of people seeking direct eye contact (Wieser et al., 2009) but also an interview comprising only neutral questions (Shimizu et al., 2011). Heart-rate variability, a marker of autonomic adaptability, was also shown to be reduced in social anxiety (Chalmers et al., 2014). The body of research examining cardiovascular responses in social anxiety does, however, not consistently report increases in heart rate but some studies did not observe elevated heart rates in socially anxious individuals (Anderson et al., 2010; Jamieson et al., 2013; Mauss et al., 2003). To complement the investigation of physiological arousal in social anxiety, examinations of electrodermal responses can shed further light on the activation of the sympathetic nervous system in different contexts. Elevated skin conductance levels have accordingly been reported in socially anxious individuals while giving a public speech (Deiters et al., 2013), listening to threatening stories (Lang & McTeague, 2009) and watching videos of anxiety-provoking scenes (Panayiotou et al., 2017).

However, again, electrodermal responses in social anxiety are not uniform (see Panayiotou et al., 2017; Puigcerver et al., 1989) and it is therefore important to consider the various aspects that could potentially contribute to the mixed evidence. Genuine social encounters are complex and already subtle gaze differences in the presented social stimuli significantly influence physiological arousal (Helminen et al., 2011; Hietanen et al., 2008; Myllyneva et al., 2015). Up until now, very few studies investigated gaze behavior in social anxiety during a real interaction and yet again yielded mixed results (Farabee et al., 1993; Gilbert, 2001; Howell et al., 2016; Langer et al., 2017; McManus et al., 2008). As gaze patterns change from laboratory to real-world environments (Foulsham et al., 2011; Hayward et al., 2017; Laidlaw et al., 2011) and as increased electrodermal responses are predominantly associated with real eye contact rather than the observation of photographed faces (Myllyneva et al., 2015), further research is needed to clarify to what extent different degrees of social interactions impact gaze and physiological responses in social anxiety. For this reason, we decided to conduct an experiment in which participants with differing degrees of social anxiety symptoms took part in a social interaction while their eye movements, cardiovascular and electrodermal activity were being measured. This study makes up the fifth and last chapter of this thesis and will illustrate the importance of examining the phenomenon of social attention with high ecological validity.

Objectives

While the progress of social neuroscience has critically contributed to our understanding of higher order social functions such as empathy or theory of mind, less is known about the attentional processes underlying such complex social behavior. Numerous eye-tracking studies have confirmed that target objects or areas of people's gaze attract attention of observers. Following another person's gaze is believed to constitute an essential mechanism by which humans infer the mental states of others (Shepherd, 2010). Additionally, it has been established that humans allocate substantial attention to fellow humans within a scene. When freely observing naturalistic images, social information (i.e. human heads or bodies) is prioritized independent of the physical saliency of competing non-social stimuli. In the following research articles, which make up my doctoral thesis, I investigated both the behavioral and neural processes underlying these attentional shifts to human beings. On the behavioral level, I aimed to examine whether social attention takes place reflexively or in a goal-directed manner. To this end, we presented naturalistic scenes containing humans somewhere in the image to participants for a very brief presentation time and investigated whether first saccades targeted humans significantly more frequently than a chance distribution of saccades would suggest. The aim of the second paper was to test whether the established orienting to social information and its temporal dynamics remain unchanged even when bottom-up information is limited. We therefore constructed a gaze-contingent paradigm in which only a small part of the image was visible to the observer and gaze shifts were necessary to deliberately explore other parts of the image. In a third paper, we decided to investigate whether the attentional bias towards social information is stronger for real faces belonging to actual people than for cartoon or figurine faces found on, for instance, advertisement posters. In an fMRI study reported in the fourth paper of this thesis, we investigated which neural activation patterns are associated with reflexive shifts to social information. Lastly, we examined to what extent differing degrees of social anxiety modulate social attention during a real interaction with another person. In all of these studies, we opted for naturalistic stimulation material to be able to approximate the complexity of the world surrounding us.

ORIENTING TOWARDS SOCIAL FEATURES IN NATURALISTIC SCENES IS REFLEXIVE

Lara Rösler[1]*, Albert End*[2] *& Matthias Gamer*[1,2]

[1]Department of Psychology, Julius Maximilians University of Würzburg, Würzburg, Germany
[2]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Published 2017 in PLoS ONE, 12(7): e0182037.

Abstract

Saliency-based models of visual attention postulate that, when a scene is freely viewed, attention is predominantly allocated to those elements that stand out in terms of their physical properties. However, eye-tracking studies have shown that saliency models fail to predict gaze behavior accurately when social information is included in an image. Notably, gaze pattern analyses revealed that depictions of human beings are heavily prioritized independent of their low-level physical saliency. What remains unknown, however, is whether the prioritization of such social features is a reflexive or a voluntary process. To investigate the early stages of social attention in more detail, participants viewed photographs of naturalistic scenes with and without social features (i.e., human heads or bodies) for 200 ms while their eye movements were being recorded. We observed significantly more first eye movements to regions containing social features than would be expected from a chance level distribution of saccades. Additionally, a generalized linear mixed model analysis revealed that the social content of a region better predicted first saccade direction than its saliency suggesting that social features partially override the impact of low-level physical saliency on gaze patterns. Given the brief image presentation time that precluded visual exploration, our results provide compelling evidence for a reflexive component in social attention. Moreover, the present study emphasizes the importance of considering social influences for a more coherent understanding of human attentional selection.

Introduction

In order to successfully navigate in our social environment, it is essential for us to be able to correctly identify and interpret social cues. In a heated debate, one does not only need to recognize the emotion displayed on an opponent's face to prepare an appropriate response, but prior to doing so, one needs to rapidly allocate attention to the respective face. Although such social attention is crucial to every social skill and interaction, little is known about the neurobehavioral mechanisms enabling it. While numerous studies (e.g. Bindemann et al., 2005; Gamer & Büchel, 2009; Mack et al., 2002; Ro et al., 2001; Shelley-Tremblay & Mack, 1999; Theeuwes & Van der Stigchel, 2006; Vuilleumier, 2000) have shown that humans display an attentional bias towards faces or other human features, these studies typically employ a highly controlled design consisting of simplified social stimuli (e.g., schematic or isolated real faces). In the past decade, an increasing number of researchers began questioning the assumption that we can generalize findings from such controlled settings to gaze behavior in real social situations (Kingstone, 2009; Risko et al., 2012). After all, the presentation of an isolated face neglects various challenges entailed in real-life social settings as, for instance, the competition between different social features or with other relevant non-social items in the scene. The studies that have examined gaze patterns in naturalistic scenes have yet shown that humans, and in particular human faces, are still prioritized even when there is competition with other salient objects in the scene (e.g. Birmingham et al., 2008b, 2009a, 2009b; End & Gamer, 2017; Fletcher-Watson et al., 2008; Freeth et al., 2010; Suda & Kitazawa, 2015). A question that remains unanswered is whether this prioritization of human features is a reflexive response to the relevant social information, or a voluntary reaction possibly driven by the motivational goal of social conformance.

This dichotomy between automatic, bottom-up and controlled, top-down attention has shaped psychological research for decades (as reviewed in Knudsen, 2007). Traditionally, bottom-up processing is believed to be automatically driven by salient stimulus characteristics, which pop out of a scene, whereas top-down processing follows higher cognitive and motivational goals, and is considered a neuroanatomically separate component of attention (as reviewed in Corbetta et al., 2008). To our knowledge, all studies examining social attention in naturalistic scenes so far have used relatively long presentation times (i.e., several seconds), which does not allow for disentangling these two components. Various attention tasks using simplified social stimuli as well as studies examining primate responses to gaze cues suggest that social attention entails a reflexive component (Bindemann et al., 2005, 2007; Deaner & Platt, 2003; Langton et al., 2000; Ristic & Kingstone, 2005). Accordingly, if an isolated face is presented next to an inanimate object for varying stimulus onset asynchronies (SOAs) as brief as 100 ms, observers will respond faster to a cue if it appears on the side where the face had previously been shown (Bindemann et al., 2007). Similarly, rhesus macaques were reported to reflexively orient their attention according to the gaze direction of the isolated image of another conspecific's face (Deaner & Platt, 2003). While these findings advocate that social attention is indeed reflexive, it remains to be seen whether this rapid prioritization can also be observed with brief presentations of naturalistic scenes in which social features compete with highly salient non-social regions of an image.

Saliency-based models of attention postulate that attention is automatically oriented to those elements of a picture which stand out in terms of their low-level physical properties (e.g.. Itti et al., 1998; Itti & Koch, 2000; for reviews see Borji & Itti, 2013, Itti & Koch, 2001; Judd et al., 2012). Computational algorithms, for example taking into account color, intensity and orientation contrast in an image, can create so-called saliency maps which, in turn, can be used to predict gaze behavior. In previous studies in which participants were asked to freely view or memorize an image, these saliency maps were validated (e.g. Foulsham & Underwood, 2008; Itti et al., 1998) and were seen to work particularly well for early fixations, suggesting a predominant modelling of bottom-up attentional processes (Parkhurst et al., 2002; but see Tatler et al., 2005). Indeed, recent research showed that saliency-based models fail to predict gaze when top-down influences are strong (as reviewed by Tatler et al., 2011). Importantly, although prior studies have investigated the impact of social features on the prediction accuracy of these saliency maps, again, the stimuli chosen for the investigation were often not truly representative of naturalistic scenes. As faces frequently presented the focus of the research question, the social features were often found in the foreground of the image rendering a true comparison of gaze behavior towards social and nonsocial features difficult (e.g. Hall et al., 2011; Nyström & Holmqvist, 2008). Moreover, some studies did not report the saliency values of the social features in their analyses which complicates an interpretation of the separate influences of saliency and social information (e.g. Fletcher-Watson et al., 2009). However, overall, current evidence still suggests saliency maps perform worse when social features are present in the visual field (e.g. Birmingham et al., 2009b; End & Gamer, 2017; Fletcher-Watson et al., 2009). It yet remains to be seen whether saliency maps perform more accurately when image presentation time is too brief for goal-driven attention to occur.

To investigate whether social attention is truly a reflexive process, we conducted an eye-tracking experiment in which participants viewed complex naturalistic images with and without social features for a brief time period (200 ms) which precluded a detailed visual exploration of the scene. The images were chosen such that social features were always restricted to one quadrant of an image. To avoid that the quadrants containing social features were also the ones with highest low-level saliency in the image, images were carefully selected resulting in balanced physical saliency across all quadrants. We analyzed the direction of the first saccade of each trial and examined whether the presence of social features outperformed physical saliency in predicting saccade direction. We found that observers made significantly more first eye movements towards image quadrants containing social features than a distribution of eye movements at chance level would suggest. In addition, the social content of the quadrant contributed significantly more to saccade direction than its saliency. Considering the brief presentation time of the stimuli, these results support the hypothesis that social attention entails a reflexive component.

Methods

Participants

Thirty-nine participants were recruited via an online recruiting system hosted by the University of Würzburg between July and September 2016. Inclusion (age between 18 and 60 years, normal or corrected to normal vision with contact lenses) and exclusion criteria (history of psychiatric or neurological illness) were described on the website allowing participants to self-verify whether they were suitable candidates which was subsequently reconfirmed by the experimenter on the day of the experiment.

A prior power analysis revealed that a sample size of 36 participants was necessary to detect medium effects (d = 0.5) in paired comparisons (one-tailed) with a power of 0.9. All participants reported normal or corrected to normal vision. Of the 39 participants, 2 participants were excluded from further analyses because filled-in questionnaires revealed a history of psychiatric or neurological illness. One further participant had to be excluded because of missing data. The final sample consisted of 36 participants (20 males, mean age: M = 26.64 years, range: 19-42 years, SD = 4.76 years). Ethical approval was obtained by the ethics committee of German Psychological Society (DGPs) and performed in compliance with Declaration of Helsinki guidelines. All participants provided written informed consent and received monetary compensation.

Stimuli

The stimulus set consisted of 100 color photographs of complex naturalistic scenes and 5 color target images displaying fractals which we obtained from various image databases (NAPS: Marchewka et al., 2014), Spanky fractal database: http://www.nahee.com/spanky) and the Internet (e.g., Google picture search, flickr). All stimuli were cropped to have the same size of 800 x 600 pixels. The complex naturalistic scenes depicted various indoor and outdoor scenarios. Among these 100 naturalistic scenes, 80 images included parts of one or multiple human beings (social scenes), whereas the remaining 20 images did not contain any human features but instead depicted landscapes or objects (non-social scenes). The social scenes were chosen such that the social features in the scene were largely restricted to one of four quadrants of the image. By mirroring the image and using different cutouts, we were able to create four different versions of the same image displaying the social feature once in each quadrant (see Fig 1B). Written text was removed from the images using the software GIMP (Version 2.8.10, GNU Image Manipulation Program, The GIMP Team) because it would have appeared unusual in the mirrored images. Based on the saliency-algorithm developed by Koch and Ullman (1987) and first implemented to gaze behavior in naturalistic scenes by Itti, Koch and Niebur (Itti et al., 1998; Itti & Koch, 2000), we calculated and ranked the relative mean saliency of each quadrant per image. Subsequently, for each participant, one version of each image was pseudo-randomly chosen while ensuring that social elements appeared equally often in one of the four quadrants (20 trials each) and that saliency ranks across all social quadrants were balanced and did not differ systematically from non-social quadrants within each subject (i.e., the average saliency rank of social quadrants amounted to 2.5 at each of the four social feature positions for each subject).

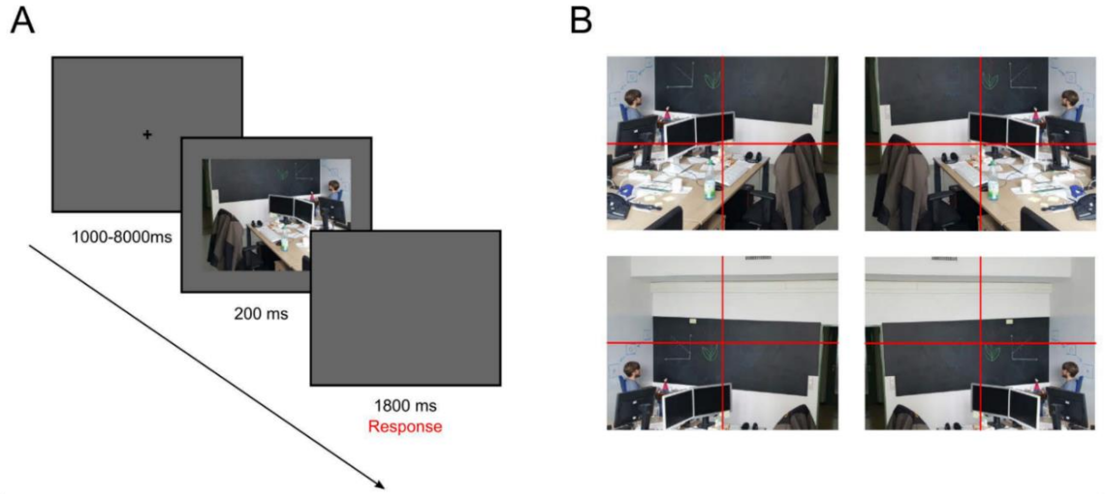

Figure 1. Trial procedure and example stimuli. A) Trial procedure. B) Example image, not included in the original dataset but taken post-hoc, illustrating how stimuli were cropped and mirrored such that social features were being restricted to a different quadrant in each version. The individual depicted in this figure has given written informed consent (as outlined in PLOS consent form) to publish these case details.

Apparatus

Participants were seated in a dimly lit room with their head stabilized at 50 cm distance to the computer screen. Stimulus presentation and response collection was controlled by the software Presentation 17.0 (Neurobehavioral Systems, Inc., Berkeley, CA, USA). Each stimulus was displayed centrally on a grey background of a 24" LG 24MB 65PY-B screen (516.5 x 323.1 mm; 1920 x 1200 pixels, 54.63° x 35.81° visual angle, 60 Hz). With this setup, the visual angle of the images amounted to 24.29° x 18.35°. Eye movements were recorded at a sampling rate of 1000 Hz of the right eye of each participant (EyeLink 1000 Plus, SR Research, Ontario, Canada).

Design and Procedure

In order to ensure that attention was paid to stimulus presentation at all times, participants were instructed to press a button upon the appearance of a fractal image. Prior to the experiment, participants underwent a training sequence consisting of twelve practice trials with different stimuli including one fractal image. The actual experiment was performed in two subsequent blocks to avoid fatigue. Each block consisted of 40 social, 10 non-social and 3 fractal images (one was used twice per participant) in pseudorandomized order. The entire experiment thus entailed a total of 106 trials. Because of practical circumstances, one participant performed the experiment in one go and viewed one fractal less (105 trials in total). Each trial started with a fixation cross shown for a random period between 1 and 8 seconds, followed by the stimulus image displayed for 200 ms and a subsequent blank screen shown for 1800 ms during which a response could be made (see Figure 1A).

Following the eye tracking experiment, participants filled in various psychometric tests and questionnaires which will be pooled across several studies and are not part of this manuscript.

Data Processing

Saccades and fixations were detected on a trial-by-trial basis. Accordingly, saccades were defined as eye movements surpassing a velocity threshold of 30°/s or an acceleration threshold of 8,000°/s^2. These eye movement data were then processed in R (Version 3.3.2, R Core Team, 2016). To facilitate drift correction and to ensure baseline stability in all trials entering the analysis, we considered the last 300 ms before stimulus onset (where a central fixation cross was shown) as baseline. For each participant, we examined baseline stability by conducting an iterative outlier removal procedure separately for x- and y-baseline coordinates (for a similar procedure see End & Gamer, 2017). Specifically, the smallest and the largest values were temporarily removed from the distribution. If any of these extreme values was more than three standard deviations away from the mean baseline position of the remaining data, it was permanently excluded from the analysis and this procedure was repeated until no more exclusion had to be performed. Saccade x and y coordinates were then corrected for gaze drift by subtracting the baseline from the actual x and y coordinate values. To determine reflexive reactions, we extracted the first saccade after stimulus onset of each trial in which a stable baseline was present. Finally, only those trials with a first saccade with an amplitude of at least 0.5° of visual angle occurring between 150 and 1000 ms after stimulus onset were considered for further analyses. Non-social images and fractals were excluded since they were only used to obscure the aim of this study and to ensure active processing of all images. Of all social trials, 35.39 trials per participant (*SD* = 23.70) were excluded on average because no or only very small saccades (i.e., below an amplitude of 0.5°) were made within that timeframe. Furthermore, an average of 2.47 social trials (*SD* = 2.99) per participant had to be excluded because of missing baseline values or outliers. The average amplitude of saccades remaining in the analysis amounted to 3.10 ° (*SD* = 1.65°).

For each trial, we computed saccade direction and latency to evaluate to which of the four image quadrants the first saccade went and how long it took to initiate it. A saccade was considered successful if the end position of the saccade was in the quadrant containing the social element.

Statistical Analyses

In a first step, we analyzed saccadic latencies as a function of saccade target. Therefore, we counted the number of saccades in 50 ms bins ranging from 150 ms to 1000 ms separately for saccades targeting social and non-social quadrants. These data were analyzed with a 2 (saccade target) x 17 (50 ms saccade latency bins) ANOVA on saccade frequency to investigate whether frequencies of saccadic latencies differed between successful and non-successful saccades.

Next, for each social feature position (left upper, left lower, right upper and right lower quadrant), percentage scores of successful saccades were calculated per participant. If saccades were not influenced by quadrant content but distributed randomly, one would expect a successful saccade percentage close to chance level (25%). In order to investigate whether saccades landed significantly more often in quadrants containing social elements than chance level would suggest, we subtracted 25% from the four percentage scores of each participant. We then submitted these values to a 2 x 2 repeated measures ANOVA with the factors horizontal (left versus right) plane and vertical (upper versus lower) plane of the saccade to investigate whether the distribution differed between quadrants (main and interaction effects) as well as from chance level (intercept of the ANOVA). We repeated the same procedure for saccade targets of all social images, independent of social feature location, to test whether a chance level distribution of saccades pertained when the influence of social information was not taken into consideration.

Finally, in order to investigate whether physical saliency drove saccade direction despite our initial balancing of saliency across quadrants, we also computed a generalized linear mixed model (GLMM) using the R package lme4 (Bates et al., 2014). Mixed-effect models are a powerful and flexible tool for statistical analysis as they contain both fixed and random effects allowing the modelling of correlated and potentially non-normal data (McCulloch & Neuhaus, 2011). Our response variable described whether a quadrant of the presented image was looked at or not. Since this is a binary event, we chose a model with a binomial error distribution and the probit-link function. After being transformed to have a mean of 0 and a standard deviation of 1, the binary variable ‘social content of quadrant’ (social content or non-social content) and the numeric variable ‘saliency of quadrant’, together with their interaction term, were included as fixed predictors into the model. As other algorithms have proven more successful in the prediction of visual exploration patterns than the one developed by Itti and Koch (Itti & Koch, 2000), we decided to compute saliency scores of each quadrant using the Graph-Based Visual Saliency algorithm (Harel et al., 2007) which performed very well in the prediction of human gaze in a recent study comparing ten computational models of saliency (Judd et al., 2012). The pattern of results remains similar when relying on the Itti and Koch algorithm (Itti & Koch, 2000). The relative saliency of each quadrant was calculated by dividing the summed saliency score of each quadrant by the summed saliency score of the entire image. To account for variability between subjects and scenes, participant ID and image ID were entered as random intercepts. The size of the beta coefficients was considered to evaluate which predictor influenced saccade direction more prominently. From each trial, information of the looked-at quadrant entered the model and, additionally, one non-looked-at quadrant was chosen randomly to prevent model bias. As the random process of choosing the non-looked-at quadrant could potentially affect the significance of the results, we decided to use a bootstrapping procedure to test the validity of our model. Correspondingly, the process of randomly choosing quadrants to enter the model was repeated 2000 times and 2000 respective GLMMs were computed. We could then calculate an empirical 95% confidence interval for the beta coefficient of each predictor based on the 2000 results of this procedure. Beta coefficients were considered as significantly different from another if the confidence intervals did not overlap.

Results

Task performance

In order to investigate whether participants paid full attention during the experiment, we calculated task performance by dividing the number of successful responses by the number of presented test stimuli. All participants had 100% accuracy in responding to test stimuli and a low false alarm rate (i.e., behavioral responses to non-fractal images, $M = 0.3\%$, $SD = 0.8\%$).

Eye movement data

Overall, subjects responded swiftly to the appearing stimuli as reflected by a mean saccade latency of 467.13 ms ($SD = 224.92$ ms). A 2 (saccade target) x 17 (50 ms saccade latency bins) ANOVA on saccade frequency revealed a significant interaction between saccade target and saccade latency bin ($F(16,560) = 11.17$, $p < .001$, $\eta_p^2 = .24$). Figure 2 suggests a bimodal distribution of saccade latencies with saccades towards social quadrants occurring earlier as compared to saccades targeting non-social quadrants. The analysis also revealed a main effect of saccade latency bin on saccade frequency ($F(16,560) = 9.27$, $p < .001$, $\eta_p^2 = .21$) but no main effect of saccade target ($F(1,35) = 0.00$, $p = .98$, $\eta_p^2 = .00$).

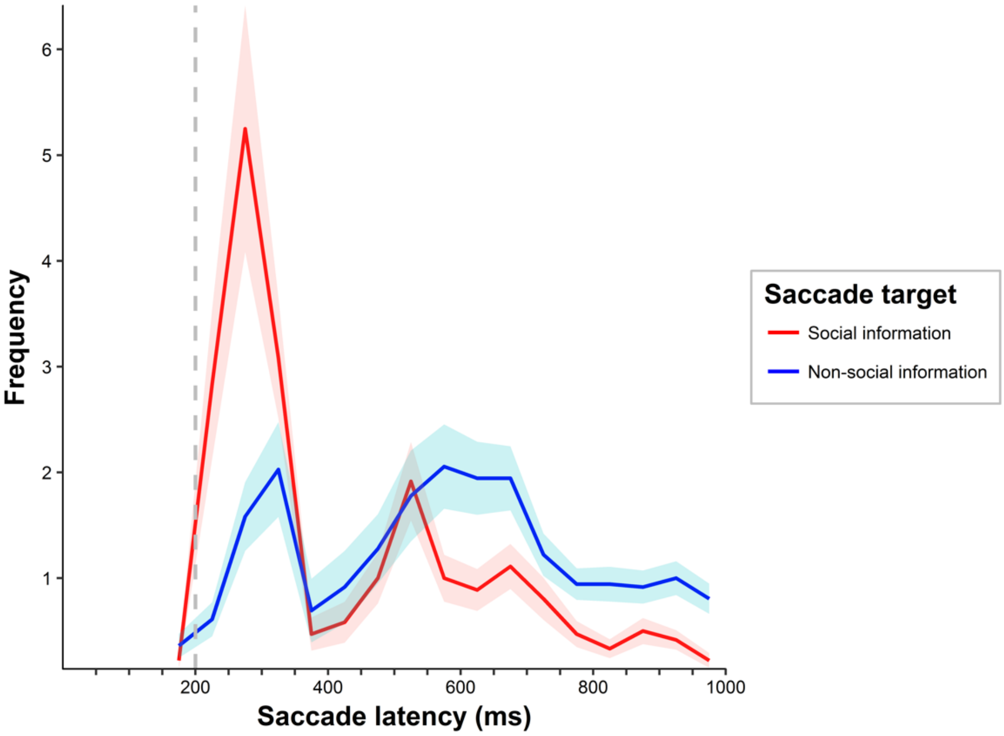

Figure 2. Latencies of successful and non-successful saccades. Distribution of the mean frequency of successful saccades (towards social information, in red) and non-successful saccades (towards non-social information, in blue) per 50 ms latency bin. Shaded areas are defined by the standard errors of the means. The dashed grey line indicates stimulus offset.

Although saccade frequency varied substantially across individuals, the 2 x 2 ANOVA, investigating the percentage scores of saccades to quadrants with social information, revealed an intercept significantly different from 0 ($F(1,33) = 66.39$, $p < .001$, $\eta_p^2 = .67$) indicating that the preference for quadrants containing social information was above chance level (see Figure 3A). We also observed a significant interaction effect between vertical and horizontal planes ($F(1,33) = 4.16$, $p = .049$, $\eta_p^2 = .11$) which suggests that a certain quadrant was preferred more than others. Indeed, Tukey post-hoc tests revealed that, when the social feature appeared on the right side of the image, participants looked significantly ($p < .05$) more often at the upper than at the lower quadrant. We did, however, not observe a main effect of horizontality ($F(1,33) = 0.001$, $p = .97$, $\eta_p^2 < .001$) and only a trend-level main effect of verticality ($F(1,33) = 3.46$, $p = .07$, $\eta_p^2 = .09$).

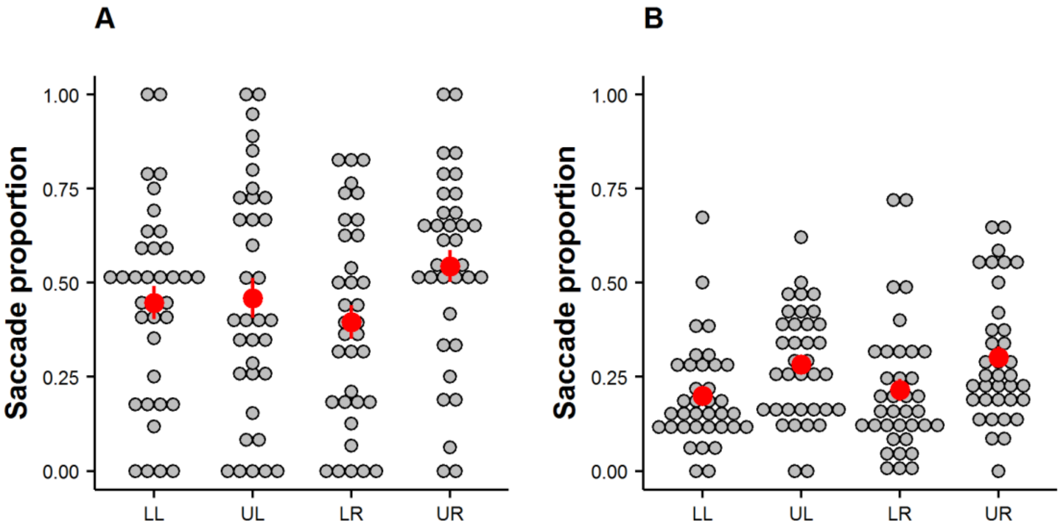

Figure 3. Saccade proportions. (A) Proportion of successful saccades (terminating in the quadrant in which social features were displayed). Here, proportions were calculated for all trials in which social features appeared in one of the four quadrants (LL = lower left, UL = upper left, LR = lower right, UR = upper right). Each circled dot represents one participant. Dark red dots denote the mean proportion of all participants and error bars depict the standard error of the mean. (B) Proportion of saccades terminating in one of the four quadrants (LL = lower left, UL = upper left, LR = lower right, UR = upper right) for all social scenes.

In contrast, a 2 x 2 ANOVA taking into account the general direction of saccades independent of social element location did not reveal an intercept significantly different from 0 ($F(1,35) = 1$, $p = .32$, $\eta_p^2 = .03$). There was, however, a significant effect of verticality ($F(1,35) = 6.86$, $p = .01$, $\eta_p^2 = .16$) as observers generally tended to look up more frequently than down (see Figure 3B). We did not observe an effect of horizontality ($F(1,35) = 0.26$, $p = .61$, $\eta_p^2 = .007$) nor an interaction effect between horizontality and verticality ($F(1,35) = 0.01$, $p = .93$, $\eta_p^2 < .001$).

In addition, we examined whether the saliency of a quadrant predicted saccade direction despite our initial balancing of stimuli. The bootstrapping procedure, which we employed to validate our general linear mixed model, revealed that both social content (mean β = 0.45) as well as saliency of a quadrant (mean β = 0.18) significantly predicted whether a quadrant was looked at (see Table 1). Since the mean of the social content beta coefficient was almost three times as large as the respective coefficient of saliency and since both confidence intervals did not overlap, these results suggest that the social content of a quadrant had a significantly greater influence on saccade direction. Importantly, there was no interaction effect between these predictors (mean β = -0.02) suggesting that the influence of these predictors was additive.

Table 1. Bootstrapping results.

	Mean	2.5%	97.5%
Social Content	0.45	0.42	0.48
GBVS-Saliency	0.18	0.14	0.21
Interaction	-0.02	-0.06	0.02

Mean and the 95 % Confidence Intervals for the fixed effects 'social content' and 'saliency' and their interaction calculated on the basis of 2000 iterations.

Discussion

In the present study, we showed that our ability to swiftly attend to humans in our surroundings seems to be reflexive. Participants briefly viewed photographs of naturalistic scenes with and without social features for 200 ms while their eye movements were being recorded. Analyses revealed that participants made significantly more first eye movements to regions containing social features than a chance level distribution of first saccades would suggest. Additionally, although saliency also drove saccade direction, the social content of an image region better predicted the target of the first saccade than its saliency, confirming that social features partially override the influence of low-level physically saliency on visual orienting. As a presentation time of 200 ms is too brief for voluntary shifts of attention to occur (as reviewed by Carrasco, 2011) and since such brief stimulus duration did not allow for a detailed visual exploration of the scene, our results substantiate the notion of a reflexive component of social attention.

Our main finding, that quadrants containing social features are prioritized as early as the first saccade after stimulus presentation occurs, provides novel insight into the mechanisms of social attention. Participants made significantly more first saccades to quadrants containing social information than towards other quadrants even though saliency was balanced across quadrants. Importantly, participants received no specific instructions prompting saccades but were simply told to respond with a button press when rare test stimuli (i.e., fractals) were shown. Our results consequently emphasize the reflexive nature of social attention. Admittedly, we are not the first to address the time course of social attention. However, previous studies have frequently used impoverished stimuli and experimental designs from which inferences about field conditions were more difficult to draw. Typically, social features were taken out of their context and contrasted with other isolated social features or isolated inanimate objects. In a study by Theeuwes and van der Stigchel (2006), for example, participants viewed photographs of isolated faces next to photographs of appliances for 200 ms, after which an arrow, indicating the direction in which an eye movement should be made, appeared. Eye movement reactions were delayed when arrows pointed to the location where a face was previously shown. This finding was interpreted as an indicator of inhibition of return (IOR) which can be used as a diagnostic tool in visual attention to identify reflexively attended locations (Klein, 2000). As IOR was greater for arrows pointing to face locations, Theeuwes and van der Stigchel concluded that attention to faces entails a reflexive component. Similarly, in a dot-probe study, isolated real faces were presented next to inanimate objects for 100, 500 or 1000 ms and participants had to promptly respond to a subsequent target appearing either at the face or at the object location (Bindemann et al., 2007). Participants were quicker in detecting targets appearing at previous face locations, providing further evidence for exogenous social attention. While our results are generally in line with these findings, we were able to investigate social attention with stimuli of higher ecological validity. Simplified social stimuli neglect many aspects of a real social scene – first and foremost, the competition between different elements in a scene (Risko et al., 2012). In our experiment, all stimuli depicted complex naturalistic scenes in which social features competed with several low-level salient non-human features. Consequently, we were able to show that social features are attended reflexively even when being surrounded by physically salient information.

Further support for the reflexive nature of social attention was provided by the mere observation of saccades in our study. Besides detecting fractals by a button press, participants did not receive any additional instructions. Hence, the observed saccades that did not allow for further stimulus exploration served no particular purpose but were reflexively triggered by the appearing stimuli. Interestingly, an investigation of saccade latencies revealed that saccades towards social information were significantly faster than saccades to image regions without social information, thus corroborating the notion of a reflexive component in social attention. These results are also in line with previous studies suggesting a dichotomy between reactive short-latency saccades and higher order saccades which display relatively longer latencies (eg. Anderson et al., 2015; Mackay et al., 2012 in natural scenes; for a general review see Hopp & Fuchs, 2004). Accordingly, reactive saccades are believed to reflect bottom-up processes regulated by subcortical circuits, specifically the superior colliculi (Dorris et al., 2007; Johnson, 2005). However, there is currently no consensus in the literature as to which latencies reactive saccades typically exhibit (suggestions for humans vary from 60-100 ms after image onset (Mackay et al., 2012) to ~ 180 ms (Smit et al., 1987)). They are considered distinct from saccades with relatively longer latencies which, in turn, are thought to involve cortical top-down processing. Although successful saccades in our study, on average, took place slightly later than 180 ms, the difference observed between successful and non-successful saccadic latencies might be related to these different saccade types. Hence, earlier saccades which predominantly targeted social information in the current study might largely reflect reflexive, bottom-up processes, whereas later saccades mostly targeting non-social information might be further modulated by top-down, goal-driven mechanisms.

The present study revealed that social features influenced saccade direction significantly more than low-level salient features of the image. In agreement with saliency-based prediction models, saliency contributed significantly to saccade direction (mean ß = 0.18), yet social content had an even greater influence (mean ß = 0.45) thus partially overriding the influence of saliency. Importantly, no interaction between these two predictors could be observed in our model suggesting that these effects are truly additive. The studies that previously investigated social attention and saliency in complex naturalistic scenes (e.g. Birmingham et al., 2009b; Cerf et al., 2008; End & Gamer, 2017; Fletcher-Watson et al., 2009; Nyström & Holmqvist, 2008; Suda & Kitazawa, 2015) also found a prioritization of social features versus low-level salient objects in a scene. Notably, our study complements these observations in three important points: (1) our presentation time was considerably shorter ensuring that we can reliably test for a reflexive component of attention, (2) earlier studies focused on the investigation of human body parts preferences (e.g., eyes versus head) and therefore usually presented humans in the center and foreground of the image, and (3) previous studies frequently relied on older and less efficient saliency algorithms. Specifically, up until now, first fixations were frequently used as a measure of early attention (Birmingham et al., 2009a, 2009b; Cerf et al., 2008; End & Gamer, 2017; Freeth et al., 2010). As voluntary attention can occur as early as 300 ms after stimulus presentation (as reviewed by Carrasco, 2011), first fixations on specific image locations cannot be warranted as reflexive when using relatively long presentation times (≥ 2 s). We avoided this ambiguity by presenting the images for mere 200 ms which reduces the impact of higher-order processes on visual orienting. Moreover, in contrast to most previous studies (Birmingham et al., 2009a, 2009b; Fletcher-Watson et al., 2009), we carefully balanced physical saliency across social and non-social quadrants per participant prior to data collection to control for the relative saliency of social features in the currently used visual scenes. In previous studies, humans frequently took up a large and central part of the image, which can potentially explain why some failed to find any contribution of saliency to saccade direction (Birmingham et al., 2009b). Lastly, earlier studies frequently relied on the saliency algorithm by Itti and Koch (Itti & Koch, 2000) which has performed poorly in a recent review comparing different algorithms in their prediction strength (Judd et al., 2012).

One shortcoming of the present study is that we used the Itti and Koch algorithm for our initial balancing of saliency across quadrants as we tried to conform to the most prominent model in the field. However, the use of the Itti and Koch algorithm was restricted to experiment preparation and for our later analysis, specifically the investigation of the contribution of saliency on eye movements, we opted for the better-performing GBVS algorithm (Harel et al., 2007). Furthermore, it needs to be noted that while we attempted to balance saliency across social quadrants and the remaining parts of the image, this balancing was based on ranks and we can thus not guarantee that social and non-social regions do not differ slightly with regard to relative saliency density. However, we have attempted to account for a potential mismatch by investigating the contributions of saliency with a generalized linear mixed model which showed that social content contributed significantly more to saccade direction. Another drawback of the current study is the large variance of valid trials between subjects. Particularly, as we did not explicitly instruct participants to make saccades, the number of saccades per participant varied greatly. Ideally, all participants would have contributed equally to our results, but our findings suggest that social information is prioritized even if some subjects react with only few saccades. Lastly, by using natural scenes, we presented ecologically valid stimuli which allowed us to shed some light on the potential mechanisms underlying gaze behavior in real-life social situations. However, we are still unable to draw strong inferences about real social interactions and it is therefore crucial that future investigations, possibly by means of mobile eye tracking or virtual reality, address gaze behavior in an interactive context to gain a more coherent understanding of social attention (see also e.g. Kingstone, 2009).

A great proportion of the existing literature on social attention focused on gaze following, primarily on the question whether another person’s focus of attention, indicated by their eyes being turned to one side or another, can exert a reflexive influence on visual orienting of the observer. To what extent are our findings conformant to existing gaze orienting literature? While initial studies provided evidence for a reflexive shift of attention following gaze cues (e.g. Friesen & Kingstone, 1998; Langton & Bruce, 1999), follow-up studies suggest that attention to gaze cues is not mediated by an exogenous mechanism (e.g. Itier et al., 2007; also see Nummenmaa & Calder, 2009). Various studies have shown that targets are detected much faster at locations looked at by others than at non-attended locations, which, given the brief presentation time of these stimuli, led credence to the idea that attentional shifts following gaze-cues could be considered reflexive. However, evidence from a study with patient EVR who suffered from frontal lobe damage and is able to exert exogenous but not endogenous attention undermines this assumption (Vecera & Rizzo, 2004). When presented with gaze cues, EVR fails to exhibit the typical faster response to validly cued targets suggesting that gaze following is not exogenous in nature. This finding is further corroborated by a study of Itier and colleagues in which participants either had to judge gaze direction of a presented head or the direction of the head itself (Itier et al., 2007). In the gaze task, participants directed ~90% of their first saccades to the eyes of the face, whereas only ~50% of initial saccades were directed to the eyes in the head task. If attention to gaze orienting was truly reflexive, gaze patterns should not differ between tasks.

With regard to the results of the current study, one could speculate that attention to social features is reflexive whereas attentional shifts following gaze cues might represent a secondary mechanism. Humans might reflexively attend the presented faces and employ additional resources to respond to gaze cues. Future research should therefore focus on the investigation of the neural mechanism underlying reflexive social attention and gaze orienting to clarify whether distinct neural substrates are recruited by these tasks. Finally, it needs to be noted that the validity of typical gaze cuing tasks has been recently put to question (Birmingham & Kingstone, 2009; Kingstone, 2009). While gaze cues were seen to elicit similar behavioral responses as arrows in the typical simplistic laboratory designs, effects of both stimulus types were not replicated to be similar using naturalistic stimuli (Birmingham et al., 2009a). Instead, depictions of humans were heavily prioritized over arrows in complex scenes, which is in line with the fast selection of human features observed in our study. This discrepancy between behavioral responses to complex versus simplistic stimuli again emphasizes the necessity to render psychological experiments more ecologically valid.

To conclude, in the present study we observed that social features in complex naturalistic scenes are attended reflexively. In addition, we were able to show that social features have a significantly greater impact on first saccade direction than low-level saliency. These results therefore argue against the generalizability of saliency-based models of attention and for a crucial impact of social information on early human visual attention.

Acknowledgements

The authors thank J. Noah Bremer for his help in stimulus collection.

ATTENTIONAL SELECTION OF SOCIAL FEATURES PERSISTS DESPITE RESTRICTED BOTTOM-UP INFORMATION AND AFFECTS TEMPORAL VIEWING DYNAMICS

Aleya Flechsenhar, Lara Rösler, Matthias Gamer

Department of Psychology, Julius Maximilian University of Würzburg, Würzburg, Germany
Aleya Flechsenhar and Lara Rösler contributed equally to this work.

Published 2018 in Scientific Reports, 8(1), 12555.

Abstract

Previous studies have shown an attentional bias towards social features during free-viewing of naturalistic scenes. This social attention seems to be reflexive and able to defy top-down demands in form of explicit search tasks. However, the question remains whether social features continue to be prioritized when peripheral information is limited, and bottom-up image information is consequently reduced. Therefore, we established a gaze-contingent viewing paradigm, in which the visual field was constrained and updated in response to the viewer's eye movements. In an eye-tracking experiment, participants viewed social and non-social images that were randomly allocated to a free and a gaze-contingent viewing condition. Our results revealed a strong attentional bias towards social features in both conditions. However, gaze-contingent viewing altered temporal and spatial dynamics of viewing behavior. Additionally, recurrent fixations were more frequent and closer together in time for social compared to non-social stimuli in both viewing conditions. Taken together, this study implies a predominant selection of social features when bottom-up influences are diminished and a general influence of social content on visual exploratory behavior, thus highlighting mechanisms of social attention.

Introduction

Amongst the variety of information in the environment, our visual system selects relevant aspects to attend in order to reduce the complexity of incoming input. This allocation of attention is commonly accomplished via eye movements and the method of eye tracking has therefore been used extensively as a straight-forward measure to investigate attentional exploration of naturalistic scenes. To predict gaze patterns and explain their underlying mechanisms, several algorithms have been implemented on the grounds of physical saliency (for a review see Borji & Itti, 2013). The majority of these approaches rest on the assumption that high local contrast in visual features (e.g., color, intensity, spatial frequency) should be conspicuous to the viewer and correspondingly attract attention. Indeed, such algorithms performed well in predicting human fixations for a multitude of stimuli under free-viewing conditions (Foulsham & Underwood, 2008; Peters et al., 2005).

While such saliency approaches particularly emphasize stimulus-driven, bottom-up attentional control, free-viewing entails the engagement of both bottom-up, as well as voluntary (top-down) attentional processes (Corbetta & Shulman, 2002; Theeuwes, 2010). In our daily lives, however, we do not only freely perceive our surroundings, but often have a certain question in mind – these task-related requirements are known to engage mainly top-down control6. Within top-down driven models, a different approach by Geissler & Najemnik (2005) has taken locations of maximum information gain into consideration, characterizing the ideal observer model and emphasizing the role of the resolution of the visual system, which is maximal at the point of the fovea and limited in the periphery. Similarly, Foulsham & Underwood (2008) and Tatler & Vincent (2009) have emphasized the importance of systematic tendencies of eye movements in scenes that may predict gaze behavior as well as saliency models, ensuing that eye movements and attention are associated, as they are driven by the same internal mechanism (see "pre-motor theory of attention", Rizzolatti et al., 1987). Importantly, while stimulus-driven and goal-driven attention are closely intertwined in free-viewing conditions, gaze-contingent viewing offers the possibility to effectively restrict pre-attentively available feature information (Loschky & McConkie, 2002). When only the currently fixated location is revealed to observers, low-level features of the image periphery cannot attract the observers' eyes in a bottom-up fashion as proposed by saliency models of attention. Indeed, search time, saccade length and fixation durations were found to be affected during gaze-contingent viewing, indicating that differential attentional mechanisms are employed during image exploration (Loschky &

McConkie, 2002). Previous studies have used gaze-contingent viewing windows to investigate how information is acquired during reading (Rayner, 1998) and which field of view optimizes picture memorization (Shioiri & Ikeda, 1989). Despite different tasks at hand, both studies rested on the assumption that gaze-contingent windows are moved in such a manner that task execution is optimized. Saccades, however, also tend to process information within the current viewing window as vertical windows trigger a higher number of vertical saccades while horizontal shapes yield more horizontal saccades (Foulsham et al., 2011). It could therefore be argued that gaze-contingent viewing reduces bottom-up processing of peripheral information but cannot entirely eliminate bottom-up processing of stimuli presented within the viewing window. While this might hold true, Kennedy and Adolphs (2010) showed that patient S.M., who suffers from a bilateral amygdala lesion, failed to fixate the eyes of faces when allowed to observe these freely but exhibited regular eye fixations when the stimuli were viewed through a gaze-contingent window. This result suggests that gaze-contingent viewing meaningfully eliminates competing bottom-up features of social information which drive gaze behavior. To what extent does gaze-contingent viewing alter gaze patterns when viewing complex naturalistic social scenes? Typically social features are prioritized over competing physically salient objects when viewing complex naturalistic scenes (Birmingham et al., 2008a; End & Gamer, 2017; Flechsenhar & Gamer, 2017; Rösler et al., 2017). Specifically, Rösler, End and Gamer (2017) have shown that attention to social features takes place reflexively as revealed by the direction of first saccades after a very brief stimulus presentation time of only 200 ms. While bottom-up processes thus seem to drive social attention, top-down processes, e.g. attempting to spot a friend in a crowded bar, are likely to additionally impact gaze behavior. Flechsenhar & Gamer (2017) showed that the implementation of tasks that specifically intended to drive attention away from social aspects of the scene still resulted in preferential allocation of attention onto depicted human beings. Collectively, these studies suggest that bottom-up mechanisms are essential in driving social attention. However, the precise role of top-down attentional control is less clear since the vast majority of studies in this domain used free-viewing conditions that do not permit a dissociation between bottom-up and top-down processes.

To investigate influences of bottom-up and top-down mechanisms in more detail, the current study contrasted a free-viewing and a gaze-contingent condition. In order to evaluate gaze pattern differences between these conditions more elaborately, we employed recurrent quantification analysis (RQA) which has been previously used to exhibit altered scan paths depending on stimulus type in a gaze-contingent compared to a free-viewing condition (Anderson et al., 2013). While Anderson and colleagues (2013) showed increased fixation recurrences in gaze-contingent viewing of naturalistic scenes, it remains unknown whether this increase persists using social stimuli. The aim of the current study was hence two-fold. Firstly, contrasting gaze-contingent with free-viewing conditions, we aimed to investigate top-down influences on social attention when bottom-up visual information is restricted. We expected these top-down mechanisms to manifest in a strong prioritization of social features within the gaze-contingent condition, which would suggest an additional importance of top-down mechanisms in regulating social attention. Secondly, we explored the temporal dynamics of social attention more generally using RQA. Here, we expected to find more recurrent and deterministic fixations for social features supporting the attentional bias towards social information in naturalistic scenes.

Results

Saliency-based prediction of fixations

As a difference measure between two probability distributions, we analyzed the Kullback-Leibler Divergence (D_{KL}) to examine how well physical saliency predicted the observed eye movements during free and gaze-contingent viewing. Herein, the distributions of saliency and fixations diverged significantly more for social stimuli as compared to non-social ones, as described by a significant main effect of stimulus content ($F_{(1,61)} = 152.64$, $p < .001$, $\eta^2 = .038$). Further, a significant main effect of viewing condition ($F_{(1,61)} = 70.78$, $p < .001$, $\eta^2 = .251$) generally describes lower predictability of fixations by saliency in free-viewing than in gaze-contingency. A significant interaction effect of both factors ($F_{(1,61)} = 31.48$, $p < .001$, $\eta^2 = .006$) refers to a smaller difference between stimulus categories within the gaze-contingent condition compared to free-viewing. Coherently, when regarding results for the area under the receiver-operating curve (*AUC*), we found an inverse relationship, namely a significant main effect of viewing condition ($F_{(1,61)} = 8.02$, $p = .006$, $\eta^2 = .045$) with worse saliency-based prediction of fixations for gaze-contingent displays than for free-viewing. A significant interaction between viewing condition and stimulus category ($F_{(1,61)} = 38.37$, $p < .001$, $\eta^2 = .019$) describes the observation that fixation predictions were worse for social stimuli in the free-viewing condition, yet better in the gaze-contingent condition. The main effect of stimulus category, however, was not statistically significant ($F_{(1,61)} = 0.56$, $p = .46$, $\eta^2 < .001$). Results of the Pearson product-moment correlation coefficient (r) showed worse saliency-based prediction of fixations for social as compared to non-social stimuli (main effect of stimulus category: $F_{(1,61)} = 28.26$, $p < .001$, $\eta^2 = .017$), and a significant difference between viewing conditions (main effect of viewing condition: $F_{(1,61)} = 43.25$, $p < .001$, $\eta^2 = .022$). Similar to the analysis of D_{KL}, the difference in predictability between stimulus categories was higher in the free-viewing than in the gaze-contingent presentation (interaction effect: $F_{(1,61)} = 21.63$, $p < .001$, $\eta^2 = .015$) (Figure 4).

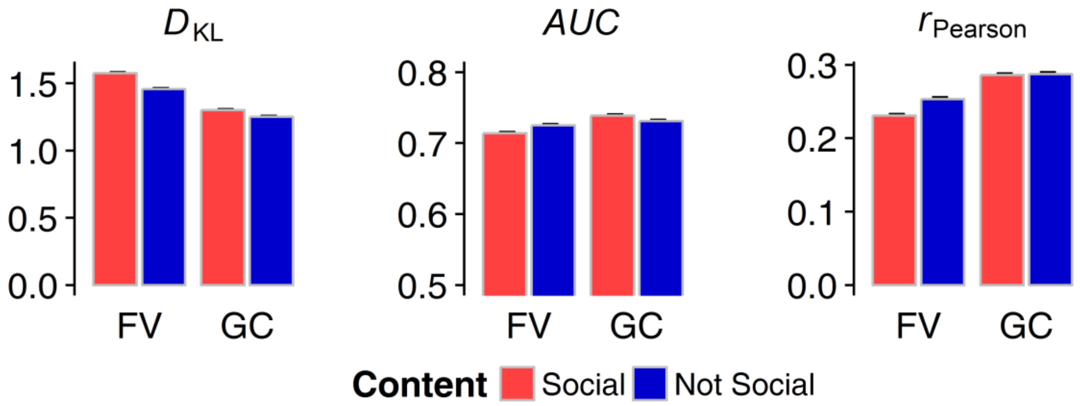

Figure 4. Divergence (Kullback-Leibler divergence, D_{KL}) and correspondence (area under the receiver-operating curve, AUC; Pearson product-moment correlation coefficient, r) between saliency and fixation density maps for social and non-social scenes in free-viewing (FV) and gaze-contingent (GC) conditions. Error bars represent standard errors of the mean.

ROI-Analysis

Considering fixation density on pre-defined ROIs, we found a significant main effect of ROI ($F_{(3,183)} = 1291.00$, $\varepsilon = .44$, $p < .001$, $\eta^2 = .899$) with social ROIs gaining most attention, especially heads, compared to all other regions in both conditions. A significant main effect of viewing condition ($F_{(1,61)} = 153.29$, $p < .001$, $\eta^2 = .239$) describes higher fixation densities in general for free-viewing. We also observed a significant interaction of viewing condition and ROI ($F_{(3,183)} = 121.52$, $\varepsilon = .46$, $p < .001$, $\eta^2 = .342$) depicting overall lower fixation densities for gaze-contingent displays than for free-viewing, which is especially the case for head and body ROIs. The interaction effect may therefore be driven mainly by the fact that exploration of social ROIs is reduced in gaze-contingent displays compared to the free-viewing condition (Figure 5, left panel). To test whether this may arise from the fact that social stimuli could not be immediately attended due to the masking, we reanalyzed the data starting from the time point at which the social aspect was first fixated.

Indeed, when comparing the time points of first fixations on social ROIs for both viewing conditions, we found that participants needed significantly less time until encountering a social ROI in free-viewing (*M* = 894.88 ms, *SD* = 268.55 ms) than in gaze-contingency (*M* = 2153.76 ms, *SD* = 461.66 ms; $t_{(74)}$ = 22.27, *p* < .001, *d* = 3.45). When further analyzing fixation data from the time point of this first social detection until the end of the presentation time, we again obtained significant main effects of ROI ($F_{(3,183)}$ = 1227.15, ε = .42, *p* < .001, η^2 = .874) depicting a fixation bias for social ROIs, and a main effect of viewing condition ($F_{(1,61)}$ = 5.61, *p* = .021, η^2 = .016) implying higher fixation densities for free-viewing as opposed to gaze-contingent viewing. A significant interaction of ROI by viewing condition ($F_{(3,183)}$ = 8.59, ε = .46, *p* < .001, η^2 = .042) emphasizes that fixation densities were different across ROIs and viewing conditions, showing slightly reduced viewing behavior for social ROIs in the gaze-contingent displays. However, compared to the pattern found previously, the difference between free-viewing and gaze-contingency regarding the fixation of social ROIs seems to be slightly smaller (Figure 5).

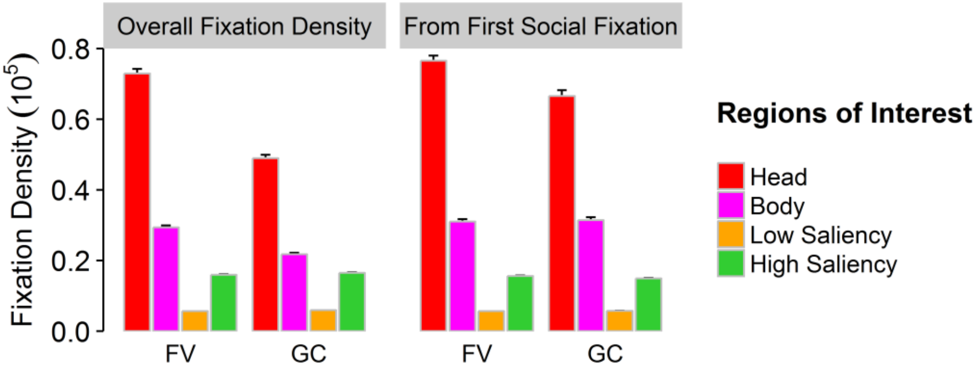

Figure 5. Relative area-normed fixation density on regions of interest (ROIs) for free-viewing (FV) and gaze-contingent (GC) viewing. The left panel depicts the overall fixation densities for the presentation duration of 10 s. The right panel shows fixation densities measured from the time point in which the participants first fixated a social feature until the end of the presentation time of 10 s. Error bars represent standard errors of the mean.

Recurrent Quantification Analyses

Recurrent quantification analyses (RQA) were suggested to complement analyses of fixation density since they provide additional information on the temporal dynamics of fixations. Figure 6 demonstrates that recurrent and deterministic fixations reveal discrepancies to fixation densities that might systematically differ between social and non-social stimulus content.

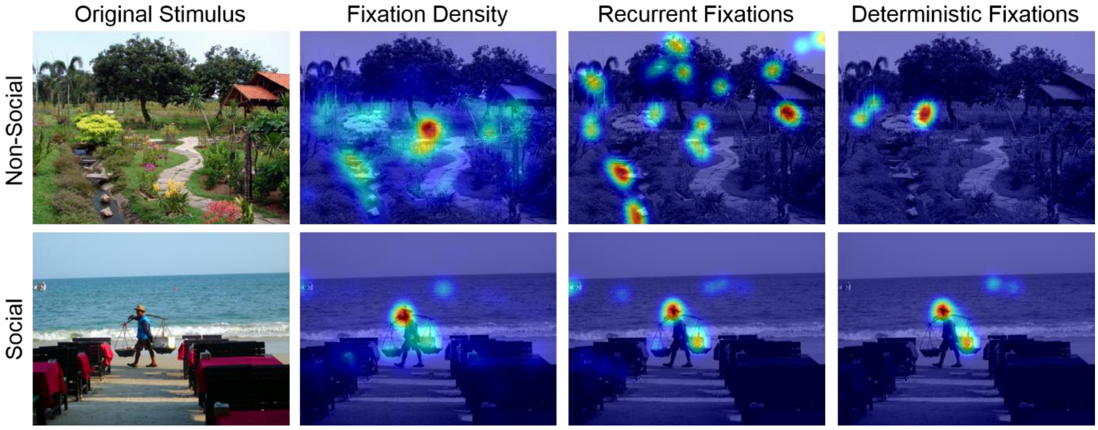

Figure 6. Example of a non-social (top) and a social (bottom) stimulus with respective heat maps for fixation densities, recurrent fixations and deterministic fixations of all participants across both viewing conditions. Warm colors represent areas with higher values of the respective measure, whereas cool colors indicate low values. Image taken with permission from the Nencki Affective Picture System (Marchewka et al., 2014).

In order to systematically quantify the influence of viewing conditions and stimulus category, we analyzed four different RQA measures: (1) for the sum of recurrent fixations, we obtained a significant main effect of condition ($F_{(1,74)}$ =31.96, p < .001, η^2 = .124), indicating higher mean recurrence for the free-viewing condition than for the gaze-contingent display. A significant main effect of content ($F_{(1,74)}$ = 104.40, p < .001, η^2 = .013) describes higher mean recurrence for social than for non-social stimuli. However, we did not find a statistically significant interaction of condition and content ($F_{(1,74)}$ = 2.88, p = .09, η^2 < .001), which signifies that there was no significant difference in the sum of recurring fixations between social and non-social stimuli across viewing conditions. (2) Deterministic fixations displayed a reversed pattern with higher means for gaze-contingent than for free-viewing ($F_{(1,74)}$ = 97.43, p < .001, η^2 = .275). Repeated subsequent fixations were also more frequent for social than for non-social stimuli ($F_{(1,74)}$ = 46.95, p < .001, η^2 = .025) but this difference between stimulus content was more pronounced for free-viewing as compared to gaze-contingent viewing as indicated by a significant interaction effect ($F_{(1,74)}$ = 10.47, p = .002, η^2 = .005). (3) Laminarity is another fixation repetition measure describing the tendency to attend certain locations multiple times (here more than twice). Our results showed a significant main effect of condition ($F_{(1,74)}$ = 121.25, p < .001, η^2 = .307) with higher laminarity for free-viewing than gaze-contingent viewing and a significant main effect of content ($F_{(1,74)}$ = 304.56, p < .001, η^2 = .111) depicting higher mean values for social stimuli. A significant interaction between condition and content ($F_{(1,74)}$ = 32.79, p < .001, η^2 = .014) suggests that in images with social content locations were revisited more often than in images with non-social content in free-viewing, but less so in gaze-contingent viewing. (4) The measure for center of recurrent mass (CORM) enabled us to examine the temporal distribution of recurrent fixations. A significant main effect of condition ($F_{(1,74)} = 270.02$, $p < .001$, $\eta^2 = .474$) describes that recurrent fixations were closer in time for gaze-contingent displays than for free-viewing. A significant main effect of stimulus content further shows that recurrent fixations were closer in time for social than for non-social stimuli ($F_{(1,74)} = 4.16$, $p = .04$, $\eta^2 = .003$). A significant interaction ($F_{(1,74)} = 14.00$, $p < .001$, $\eta^2 = .011$) between viewing condition and content suggests that recurrent fixations occurred closer in time for social stimuli than for non-social ones in free-viewing, but farther in time for the gaze-contingent condition (Figure 7).

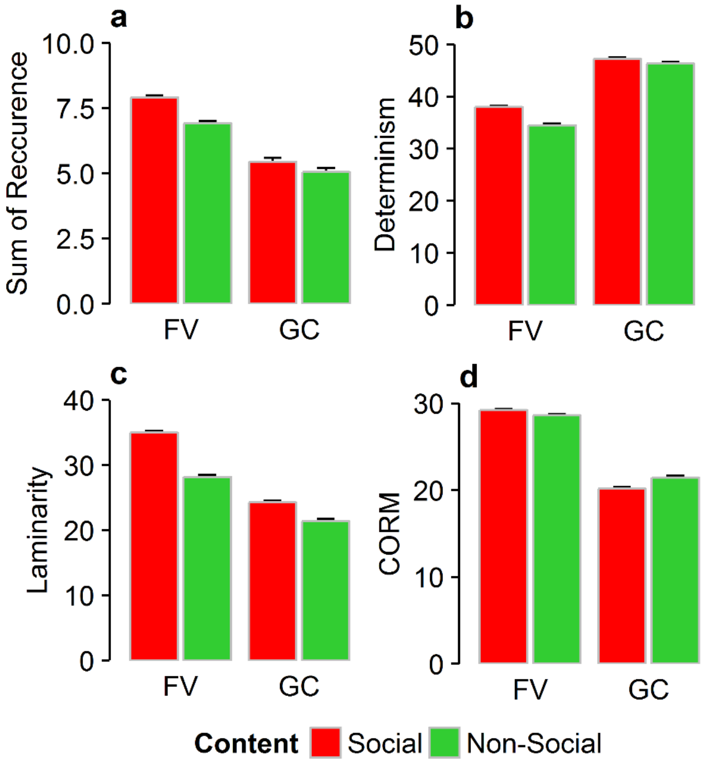

Figure 7. Averages of four recurrence quantification analysis measures: a) Recurrence, b) Determinism, c) Laminarity and d) Center of Recurrent Mass across free-viewing (FV) and gaze-contingent (GC) conditions for social and non-social stimulus content. Error bars represent standard errors of the mean.

Discussion

This study used a gaze-contingent display to investigate social attention when peripheral visual information is limited. The current results from a relatively large group of participants revealed a robust attentional exploration of social features even when reducing the influence of bottom-up mechanisms. Additional analyses of the temporal dynamics of fixation patterns demonstrated increased recurrences and deterministic fixations for social as compared to non-social images which suggests that social information might be special regarding its influence on the generation of priority maps for attentional selection.

In detail, our results showed that social features, especially faces, were preferentially fixated over physically salient areas independent of the viewing condition. Since gaze-contingent paradigms subdue bottom-up driven mechanisms and rely more heavily on voluntary control over gaze direction and allocation, the prevailed, yet somewhat diminished fixation density on social features in our study suggests that social attention involves voluntary attentional selection. When further comparing modalities from the time point at which the first fixation on social features was registered, this disparity across conditions decreased, yet remained significant. This proposes the possibility that the difference in fixation density is partly impacted by the time spent searching for a social element in the gaze contingent condition. Importantly, this attention bias for social features cannot be ascribed to the fact that these aspects were physically highly salient. Consequently, the power of saliency-based predictions was considerably reduced when social features were present in complex naturalistic visual input. This is in line with findings of End & Gamer (End & Gamer, 2017), who also observed that the influence of physical saliency on gaze behavior is weakened by social stimuli in free-viewing. This further implies that physical saliency is insufficient in predicting gaze behavior when the visual field contains social information (Birmingham et al., 2009b; Fletcher-Watson et al., 2008; Scheller et al., 2012). Similarly, saliency-based attention models tend to work less efficiently when faced with top-down implications (Einhäuser et al., 2008; Foulsham & Underwood, 2008). Coherently, a viewing condition in which bottom-up processes are largely reduced, may also alter gaze predictability through saliency. Indeed, the measures used in this study reflect worse predictions for gaze-contingent than free-viewing modalities. Moreover, the differences between fixation predictions for social and non-social stimuli were more apparent during free as compared to gaze-contingent viewing.

As an investigation of the temporal dynamics of fixation sequences complements the analysis of mere fixation densities, we also examined recurrence quantification measures in both viewing conditions. The characterization of viewing behavior concerning recurrent fixations aimed to find not only differences between viewing conditions, but also for stimulus content, most importantly increased recurrences for social features. More generally, our results replicated those of Anderson and colleagues (Anderson et al., 2013) who observed increased recurrences when natural scenes were viewed freely compared to when they were viewed gaze-contingently. We were further able to replicate the observation that deterministic fixations (i.e., one fixation repeatedly following another) occur more frequently in gaze-contingent viewing, likely due to the sequential targeting of features within the gaze-contingent window. Similarly, we also found laminarity and center of recurrence mass to be increased in free-viewing, suggesting that single fixations were repeated more often and that repetitions generally occurred further apart in the trial sequence in free-viewing than in the gaze-contingent condition. Importantly, although Anderson and colleagues (Anderson et al., 2013) did use different sets of stimuli (exteriors, interiors and landscapes), our stimuli allowed us to compare re-fixations in social versus non-social scenes to investigate the role of social content in attentional control. This revealed that recurrences were higher for social than for non-social images. Herein, all fixation repetition measures (sum of recurrence, determinism, laminarity) indicated greater recurrences for social than non-social image areas. Furthermore, recurrences were closer together in time for social than non-social image areas as measured by the center of recurrence mass. Conclusively, the results of the recurrence quantification analysis support preferential viewing behavior towards social information shown by fixation densities, by revealing that this prioritization manifests through multiple re-fixations throughout the viewing time.

The combination of viewing modalities allowed an additional examination of predominantly top-down influences (gaze-contingent viewing) and both, bottom-up and top-down influences (free-viewing) on social attention. While bottom-up processing mechanisms cannot be ruled out completely in the gaze-contingent viewing condition, fewer low-level salient information is available and most executed saccades will draw on top-down processes for the determination of saccade endpoints. Our current results therefore suggest that social attention is not merely reflexive but also relies on top-down attentional processes. . So how does social attention then fit into the traditional theory of bottom-up and top-down mechanisms? The recurrence quantification analysis used here further implicates that viewing behavior towards social stimuli is different than for non-social stimuli with regard to fixation sequence as well as temporal structure. Foulsham & Kingstone (2010) already showed that gaze patterns can change with image content in a scene, but our data presents explicit differences between social and non-social content, suggesting that social attention is inherently different from general attention mechanisms. This, in turn, raises the question whether a special neuro-cognitive system, distinct from the ventral or dorsal network suggested for bottom-up and top-down attention, mediates social attention and its rapid allocation. The study of Kennedy & Adolphs (2010), who showed that irregular bottom-up processing caused by amygdala lesions can be overcome using a gaze-contingent paradigm, indicates how important the disentanglement of these processes are. Furthermore, such patient studies can offer insight to underlying mechanisms and further our understanding of brain areas involved in social processing. Future neuroimaging studies investigating potential candidates for a social attention network are necessary to further elucidate this assumption.

Even though our findings depict robust and successfully replicated results, our study has a few limitations. First, we cannot control for certain influences arising from the use of naturalistic stimuli that are reduced in simplified laboratory stimuli. For instance, although the distribution of social dimensions was considered, such that they were not always presented centrally, in the foreground or depicted only single individuals, internal validity could not be completely ensured by our setup. Furthermore, even though spatial frequencies have repeatedly been reported to affect attentional capture (Gomes et al., 2018; Stein et al., 2014), we did not control for the spatial frequencies in our stimulus set. Furthermore, Nevertheless, we chose these complex scenes as they have comparatively high ecological validity and contain contextual information which plays an important role for the orientation in our environment (Torralba et al., 2006). Moreover, we deliberately wanted to defer from isolated or artificial setups, as they have yielded different results (Kingstone et al., 2003; Smilek et al., 2006). However, the use of photographs of naturalistic scenes, has also been put into question (Kingstone et al., 2003), as these are not equivalent to experiencing the real world and some recent studies have shown conflicting results comparing eye tracking in the laboratory compared to mobile eye tracking (Foulsham, Walker, et al., 2011). Second, our study included animal pictures and studies have shown that eye movements may be influenced by animacy of depicted features within a complex scene (e.g. Altman et al., 2016; New et al., 2007). However, this theory implies that gaze behavior for our non-social stimuli should increase in the presence of animate objects. Even so, our results still show better predictions through saliency measures for non-social stimuli and higher recurrent measures for social as compared to non-social stimuli within our recurrence quantification analysis. Therefore, social features may still be preferred over other animate aspects. Third, even though the previous study (End & Gamer, 2017) yielded no significant results concerning stimulus ratings, we did not assess personal relevance of the images in our study. Still, our results provide valuable information about the processing of social stimuli and endorse the notion of differential mechanisms than those traditionally suggested.

In summary, this study successfully replicated and extended previous results using recurrent quantification analysis, showing that gaze patterns were not only very different for free-viewing as opposed to gaze-contingent viewing, but also for social compared to non-social content. This attention bias was also evident for fixation densities and cannot be accounted for by physical saliency. Concluding, our results imply a social prioritization that appears to involve voluntary attentional selection and thereby substantiates the notion that social stimuli are exceptional concerning visual attention.

Methods

Participants

We used power analyses (Faul et al., 2007) to calculate the number of participants necessary for revealing medium-sized effects in paired t-tests (Cohen's $d = 0.50$) or repeated measures analyses of variance (ANOVAs, $f = 0.25$), respectively, at a significance level of $\alpha = .05$ and a power of .95. When assuming a correlation of $r = .50$ between factor levels in the ANOVA, these analyses revealed a required sample size of 54 participants. We thus aimed at recruiting a minimum of 60 participants in order to account for potential dropouts.

Since participant recruitment was more successful than anticipated, a total of 82 subjects (37 males) participated in this study. Of these 82 participants, 30 participants were recruited primarily from the University of Würzburg's Human Participant Pool and 52 from a database allowing pre-screening of social anxiety and the subsequent selection of a normal distribution of social anxiety (which is of no further relevance to the current study). Three participants were excluded because of current medication usage or a neurological illness. Participants with more than 30% missing baseline values or outliers (see below) were also not considered in the analysis resulting in the exclusion of four additional participants. The final sample thus consisted of 75 participants (30 males) with a mean age of 24.08 years (*SD* = 5.29 years). All participants had normal or corrected-to-normal vision.

The study was approved by the by the ethics committee of the German Psychological Society (DGPs) and conducted according to the principles expressed in the Declaration of Helsinki. Each participant provided written informed consent prior to the experiment and was awarded extra course credit or monetary compensation.

Apparatus

The experiment was programmed with MATLAB© 2011b (MathWorks, Inc., Natick, MA, USA) using the Psychophysics Toolbox (Version 3.0.12; Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and presented on an LG 24MB 65PY-B 24" monitor with a physical display size of 516.9 x 323.1 mm. The monitor had a resolution of 1920 x 1200 pixels and a refresh rate of 60 Hz. Eye movements were tracked using a mounted EyeLink 1000 Plus system (SR Research Ltd., Ottawa, Canada). The sampling rate was set to 1000 Hz and we tracked the right eye at a viewing distance of 50 cm.

Stimuli

The stimuli used consisted of 160 naturalistic images. Half of these images displayed scenes containing one or more human beings displayed anywhere within the image, which will be referred to as social images in the following. The other 80 images showed scenes containing non-social features, predominantly complex landscapes, including objects and on rare occasions animals. The stimulus set was taken from End and Gamer (End & Gamer, 2017) and created from various image databases including the Nencki Affective Picture System (Marchewka et al., 2014), EmoPics (Wessa et al., 2010), the International Affective Picture System (IAPS; Lang et al., 1997), McGill Calibrated Colour Image Database (Olmos & Kingdom, 2004), Object and Semantic Images and Eye tracking dataset (OISE; Xu et al., 2014) and websites such as Flickr and Google. Contrast and luminance were adjusted manually by visual judgement. Stimuli were presented in a resolution of 1200 x 900 pixels resulting in a visual angle of 35.81° x 27.24° within the current setup. The currently used social and non-social images were already employed in a previous study and were shown to be comparable regarding basic visual properties such as image complexity or clutter as well as affective quality and personal relevance (End & Gamer, 2017).

Design

The experiment consisted of two different types of viewing modalities for the stimuli: (1) free-viewing and (2) gaze-contingent viewing. For each participant, images were randomly associated to these viewing conditions while ensuring for an equal number of social and non-social images in each condition. In the free-viewing condition, the whole image was visible at a time and could be explored freely. The gaze-contingent display enabled the participant to only see the part of the stimulus that was centered at the current fixation location. The online tracking enabled real-time contingency on the display with the movement of the participant's eye. The visible area was defined by a Gaussian transparency mask with full-width half-maximum of 3° of visual angle around the center of the current fixation location (adapted from Kennedy & Adolphs (2010). The stimuli in the gaze-contingent condition were masked with a fixed grid of small dots located 2.2° from one another with a 3-pixel diameter to offer a sense of coordination during stimulus exploration (Figure 8). There was no postulated task for either the free-viewing nor the gaze-contingent condition, but participants were instructed that they could explore the stimuli freely if desired. Further, they were informed that the image would be masked in the gaze-contingent condition and that they would be able to uncover image areas by moving their eyes.

Procedure

Each trial began with a fixation cross presented on a grey background for 1 s. Stimuli were presented for 10 s in both viewing conditions. Afterwards, a fixation cross appeared again comprising an inter-trial-interval of 1-3 s. The experiment was divided into four different blocks, two of which were free-viewing, the other two were gaze-contingent. The blocks were alternated as such that a block of one condition would always follow a block of the other. Every second participant started with a gaze-contingent block to avoid sequence effects. A 9-point calibration was conducted at the beginning of each block and a drift correction after every 8 trials to ensure precise measurement and correct exposure of stimulus details in the gaze-contingent condition. Six training trials using a different set of pictures were included to enable participants to become acquainted with the paradigm.

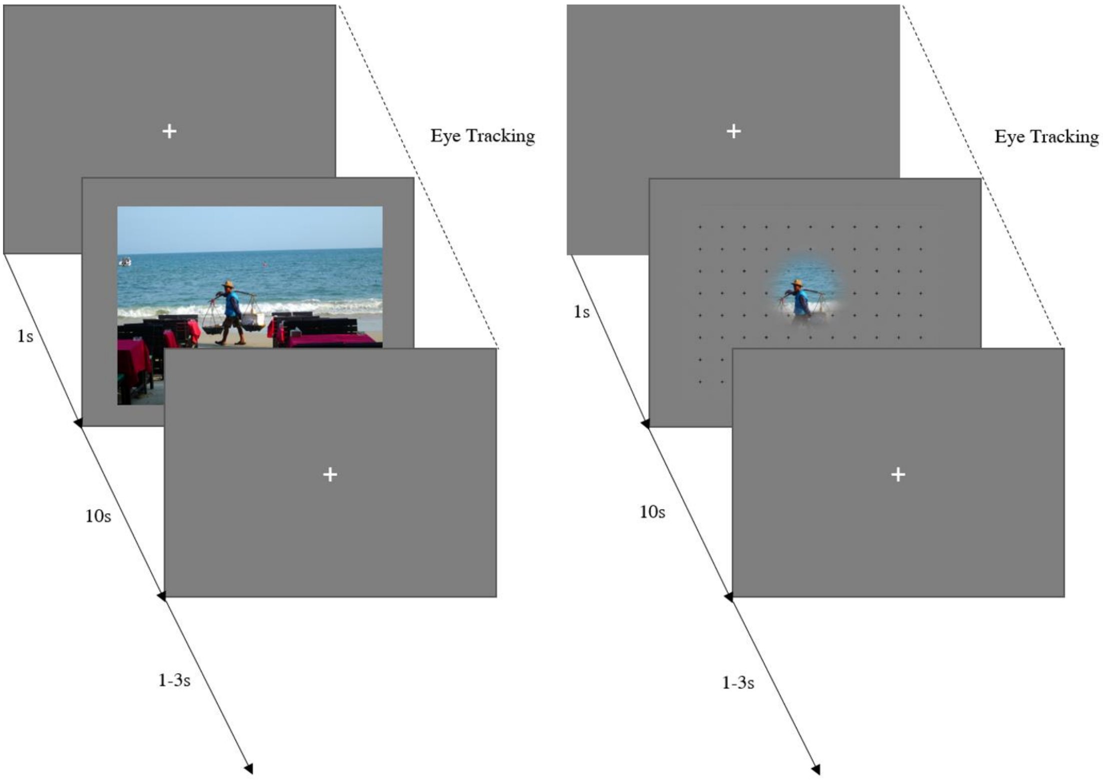

Figure 8. Example of an experimental trial for a free-viewing condition (left) and a gaze-contingent condition (right). The presentation time for both conditions was set to 10 s. Image taken with permission from the Nencki Affective Picture System (Marchewka et al., 2014).

Data Analysis

Data were analyzed using the open-source statistical programming language R (version 1.0.136) and MATLAB® R2011b. The R-package *ez* (version 4.3; Lawrence, 2016) was used for all repeated-measures analyses of variance (ANOVAs). An a priori significance level of $\alpha = .05$ was specified for all statistical tests. Generalized η^2 (Bakeman, 2005) and Cohen's *d* are reported as estimates of the effect size for ANOVAs and *t*-tests, respectively. The Huynh-Feldt procedure was used for all repeated-measures ANOVAs containing more than one degree of freedom in the enumerator to account for potential violations of the sphericity assumption.

Eye Tracking Preprocessing

Eye tracking data preprocessing was essentially identical to an earlier study17 including all commonly applied steps – drift correction, iterative baseline outlier removal and creation of fixation maps with Gaussian kernel smoothing of 2° of visual angle (see Supplementary Material S1 for full details).

General Influence of Saliency

In order to determine to what degree low-level visual saliency predicts fixations for social and non-social scenes, we compared similarities between fixation density and saliency maps. The latter were calculated for each image using the Graph-Based Visual Saliency (GBVS) algorithm (Harel et al., 2007) that was shown to be capable of predicting visual exploration with considerable accuracy (Borji & Itti, 2013; Judd et al., 2012) . Similar to fixation densities, saliency maps were normalized to range from 0 to 1. Both maps were compared using standard metrics (Wilming et al., 2011). These comprised of divergence of the distribution of physical saliency and fixation density (Kullback-Leibler divergence, $D_{KL;}$ Itti & Baldi, 2005; Kullback, 1960), the classification of saliency at fixated and non-fixated image locations (area under the receiver-operating characteristic curve; *AUC* ; Fawcett, 2006; Tatler et al., 2005) and the linear dependence between the two variables (Pearson product-moment correlation coefficient *r* ; Hwang et al., 2009; Kootstra et al., 2011). For AUC, fixation density maps were binarized using the mean fixation density as threshold. All metrics were calculated separately for social and non-social scenes and the two viewing conditions and compared using 2 x 2 repeated-measures ANOVAs with factors viewing condition (free-viewing, gaze-contingent) and stimulus category (social, non-social) on each measure.

Regions of Interest

To quantify the fixation density onto physically salient aspects and social features, we introduced regions of interest (ROIs). Similar to our previous studies (End & Gamer, 2017; Flechsenhar & Gamer, 2017), we differentiated between regions of high saliency, low saliency, head and body (see Supplementary Material S2 for full details). Area-normed fixation density scores for these ROIs were analyzed using a 2 x 4 repeated-measures ANOVA with factors viewing condition (free-viewing, gaze-contingent) and ROI (head, body, low saliency, high saliency).

To investigate potential influences on attention towards social features in the gaze-contingent viewing condition as compared to free-viewing, we conducted post hoc analyses to determine if the observed difference was due to a significant time difference in initial detection of the social feature. Hence, we compared viewing conditions anew, selecting fixations from the time point in which a social ROI (head or body) was first fixated. The time points of initial social fixations were compared for both viewing conditions in a paired *t*-test for the social stimulus set. Furthermore, we generated new fixation density maps for the time window after the social ROI was detected and analyzed area-normed fixation densities on ROIs using a 2 x 4 repeated-measures ANOVA with factors condition (free-viewing, gaze-contingent) and ROI (head, body, areas of low saliency and high saliency).

Recurrence Quantification Analysis

Another tool for describing complex dynamic systems and characterizing gaze patterns is recurrence quantification analysis (Anderson et al., 2013; Marwan et al., 2002; Webber & Zbilut, 2005). Herein, fixations which repeatedly occur at the same location can be identified, which offers additional information about gaze patterns in the presence of social features for different viewing conditions. The determination of whether a fixation was recurrent or not was accomplished by a fixed radius revolving around the previous fixations. The radius was chosen according to the size of the gaze-contingent window used in the experiment (adopted from Anderson et al. 2013) and thus amounted to 97 pixels, which is equivalent to 3° visual angle (Webber & Zbilut, 2005). To compare fixation sequences across experimental conditions, quantitative measures were extracted, namely, a recurrence measure (how often observers fixate previously viewed image locations), a determinism measure (describing fixation locations that likely follow one another), a laminarity measure (indicating that regions were fixated multiple times) and a center of recurrence mass (CORM; indicates where in time most of the recurrent fixations were located with small CORM values implying re-fixations that are closer in time than those with large CORM values) (for details see Anderson et al. (2013); the code was kindly made available by Nicola Anderson and implemented in MATLAB). The measures were computed separately for both viewing conditions and social and non-social images and subsequently analyzed in four 2 x 2 repeated-measures ANOVAs with factors viewing condition (free-viewing, gaze-contingent) and stimulus category (social, non-social).

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

This work was supported by the European Research Council (ERC- 2013-StG #336305). The authors thank Sandrine Michelberger, Michael Strunz and Jenni Rall for their assistance with participant recruitments and data acquisition.

Author Contributions

A.F., L.R. and M.G. designed the study. A.F. and L.R. analyzed the data. M.G. supervised data analysis. A.F., L.R. and M.G. wrote and reviewed the manuscript.

ARTIFICIAL FACES PREDICT GAZE ALLOCATION IN COMPLEX DYNAMIC SCENES

Lara Rösler, Marius Rubo, Matthias Gamer

Department of Psychology, Julius Maximilians University of Würzburg, Würzburg, Germany
Lara Rösler and Marius Rubo contributed equally to this work.

Published 2019 in Frontiers in Psychology, 10, 2877.

Abstract

Both low-level physical saliency and social information, as presented by human heads or bodies, are known to drive gaze behavior in free-viewing tasks. Researchers have previously made use of a great variety of face stimuli, ranging from photographs of real humans to schematic faces, frequently without systematically differentiating between the two. In the current study, we used a Generalized Linear Mixed Model (GLMM) approach to investigate to what extent schematic artificial faces can predict gaze when they are presented alone or in competition with real human faces. Relative differences in predictive power became apparent, while GLMMs suggest substantial effects for real and artificial faces in all conditions. Artificial faces were accordingly less predictive than real human faces but still contributed significantly to gaze allocation. These results help to further our understanding of how social information guides gaze in complex naturalistic scenes.

Introduction

When exploring our surroundings, we preferentially allocate attention to other human beings. Various eye-tracking studies have shown that our strong tendency to fixate others is apparent both when viewing images or videos in laboratory settings (Bindemann et al. 2010; Birmingham and Kingstone 2009; Cerf et al. 2009; Coutrot and Guyader 2014; Itier et al. 2007; Kingstone 2009; Nasiopoulos et al. 2015; Xu et al. 2014; Flechsenhar and Gamer 2017; Rösler et al. 2017; End and Gamer 2017) and, although to a slightly reduced extent, in real-life social interactions (Freeth et al. 2013; Laidlaw et al. 2011; Foulsham et al. 2011). Among these different viewing modalities, a strong preference for heads (Freeth et al. 2013) and, if stimulus resolution allows, eyes of others (Birmingham et al. 2008) can be discerned. It has been argued that this bias towards the eyes of conspecifics enables the deciphering of others' internal states and therefore represents an essential prerequisite for successful social interactions and integration in society (Shimojo et al. 2003; Ristic et al. 2005; Frischen et al. 2007).

We are sometimes, however, confronted with human-like features which do not give room for interaction. We here refer to any human-like face that has been produced by another human being as artificial. By this definition, an advertisement poster of a local politician but also a statue in a church or a humanoid robot are considered instances of artificial faces. How does the processing of these artificial faces differ from the processing of real faces? Mimicry and gesture of cartoon figures or statues also convey information about their alleged emotions or internal states and were even seen to yield higher accuracies in emotion detection than real faces (Kendall et al., 2016). Observers yet commonly know that these human representations are not real and therefore cannot be meaningfully interacted with. Previous studies have shown that gaze patterns are affected by social presence (Freeth et al. 2013) and the possibility of a social interaction (Laidlaw et al., 2011), leading to gaze behavior that is adapted to social norms (e.g., reduced fixations on strangers' heads). The attentional bias towards eyes was yet seen to persist even when these are part of very unhuman-like fictional monsters and located in surprising parts of their bodies (Levy et al. 2013). The similarity between artificial and real human face processing is further highlighted by a vast body of electrophysiological studies which reported neural face-processing signatures, e.g. the electrophysiological N170 response in the electroencephalogram, to schematic faces (Jeffreys, 1996), inverted schematic faces (Sagiv & Bentin, 2001) and even scrambled face features after face priming (Bentin et al., 2002; Bentin & Golland, 2002). However, effects of direct versus averted gaze in these electrophysiological responses could only be detected with photographic but not with schematic faces (Rossi et al., 2015) and overall the amplitude of the N170 largely seems contingent on the fixation of eyes (Itier et al., 2007; Parkington & Itier, 2018). Functional magnetic resonance imaging studies have further shown that similar brain regions are recruited when perceiving a performed action (Gazzola et al., 2007) or emotion of a robotic or human agent (Chaminade et al., 2010). Interestingly, a preference of face-like artificial stimuli could even be observed in the human fetus (Reid et al., 2017; but see Scheel et al, 2018), yielding initial evidence that our tendency to orient to artificial face-like structures is not contingent on postnatal experience. These findings suggest that social features attract attention even when they are not part of an actual fellow human being. How are fixations distributed, however, when both real and artificial faces directly compete for attentional resources?

For the further exploration of processing differences between real and artificial faces the choice of appropriate stimulus material is a challenging one. While static images can in theory display both artificial and real human faces, they will ultimately be an instance of artificial material (e.g., a picture of a person viewing a picture of a person). Videos, however, allow the possibility to display both real human and artificial faces while rendering the difference between the two more evident. Furthermore, videos are a better approximation of real-life dynamic situations than static stimuli potentially rendering the interpretation of results more meaningful (Risko et al., 2012). Accordingly, computational accounts of gaze allocation perform significantly better when motion, which is only available during dynamic and not static stimuli, is considered during face processing analyses (Curio et al. 2011). The superiority of dynamic stimuli in face processing research is further supported by clinical studies showing that certain differences in gaze allocation between patients with autism spectrum disorder and healthy controls only become apparent when using dynamic instead of static stimuli (Speer et al., 2007).

There is an on-going scientific debate to what extent low-level physical features of the stimulus material (so-called physical saliency) need to be considered when analyzing gaze patterns in static or dynamic scenes. While proponents of saliency approaches claim that bottom-up processing of scenes can be fully accounted for by low-level physical features such as luminance, color intensity and orientation (e.g. Itti et al. 1998; Itti and Koch 2000), various studies have shown that these algorithms do not work well when top-down influences are strong (as reviewed by Tatler et al. 2011). The use of dynamic stimuli, however, introduces additional temporal saliency features (e.g., flicker and motion) which were seen to predict viewing behavior during free-viewing (Mital et al., 2011) supporting the general notion of low-level physical saliency as a crucial predictor of gaze allocation.

To disentangle the influences of physical saliency and the appearance of human and artificial faces on gaze patterns, we presented videos including human faces only, artificial faces only and videos including both human and artificial faces to participants while recording their eye movements. Saliency maps were computed using the Graph Based Visual Saliency algorithm first introduced by Harel, Koch and Perona (Harel et al. 2018). Using a generalized linear mixed model, we were able to separately evaluate the impact physical saliency and human and artificial faces had on fixation probability. Since participants freely viewed the stimulus material, we expected human faces and low-level physical saliency to be most impactful on eye movements but assumed artificial faces to also attract attention although to a somewhat lesser extent.

Methods

Participants

A prior power analysis (Faul et al., 2007) showed that 34 participants were necessary for revealing medium-sized effects in paired t-tests (Cohen's *d* = 0.50) at a significance level of α = .05 and a power of .80. In order to take into account potential dropouts, we recruited thirty-six participants (15 males). Because of a too large variability of baseline coordinates (for calculations see below), one participant had to be excluded from our sample. Our final sample thus consisted of 35 participants with a mean age of 25.66 years (*SD* = 4.88 years) via the University of Würzburg's Human Participant Pool. All participants had normal or corrected-to-normal vision. Ethical approval was obtained by the ethics committee of German Psychological Society (DGPs). Each participant provided written informed consent and was awarded monetary compensation or course credit for participation.

Stimuli

The stimulus set consisted of a total of 60 videos varying between 18 and 20 seconds of length without any cut interruption. These 60 videos contained four subsets of 15 videos each displaying either only real human faces, only artificial faces, both human and artificial faces, or no faces at all. Artificial faces were categorized as such when they shared key features of a human face including round shape, nose and eyes but did not belong to an actual human being in the scene. Examples include posters of humans, statues or street art (for a detailed description of the video content see

Supplementary Table 1). The 30 videos including artificial faces were newly acquired via a free online streaming platform, while the remaining 30 videos were taken from an earlier study (Rubo & Gamer, 2018). In order to be included in our study, videos generally had to depict natural scenes, usually representing outdoor scenery, and had to be taken from a wide angle with a still or slowly moving camera. Additionally, the human beings displayed in the videos were not to perform any surprising actions. As text is known to greatly influence gaze allocation (Cerf et al. 2009), we further attempted to avoid the display of conspicuous text within our videos. All videos had a resolution of 1280 x 720 pixels and were converted from their original format to a 30 frame-per-second MPEG-4 video file resulting in a total of 35,041 frames across all videos.

Apparatus

Videos were presented centrally on 24" LG 24MB 65PY-B screen (516.9 x 323.1 mm; 1920 x 1200 pixels, 60 Hz). We used a chin and forehead rest to minimize head movements and to warrant a constant viewing distance of 50 cm, resulting in a viewing angle of 38.03° x 21.94° of the displayed videos. Eye movements of the right eye were tracked at a sampling rate of 1,000 Hz (EyeLink 1000 Plus, SR Research, Ontario, Canada). Stimuli were presented using MATLAB© 2011b (MathWorks, Inc., Natick, MA, USA) and the Psychophysics Toolbox (version 3.0.12; Brainard, 1997; Kleiner et al., 2007; Pelli, 1997)

Procedure

Prior to data acquisition, participants were instructed to watch the videos as if watching a TV-show. To avoid fatigue, the experiment was split into two blocks, each containing 30 videos. Each trial began with a fixation cross displayed centrally on a grey background for 5 s, followed by the onset of a video. Eye movements were recorded together with time stamps marking the beginning of each video frame. To avoid sequence effects, videos were displayed in random order to each participant. As a final part of the experiment, participants filled in various psychometric tests and questionnaires which will be pooled across several studies and are not analyzed as part of this manuscript.

Eye tracking preprocessing

Gaze data were analyzed using R (version 3.2; R Development Core Team, 2015). Any eye tracking data recorded up until 150 ms after stimulus onset were excluded from the analysis to account for lingering on the initial fixation cross position. Since the eye tracker sampled eye movements at 1,000 Hz and videos had a frame rate of 30 Hz, approximately 33 raw eye positions were recorded per frame. Eye data was consequently collapsed over each frame such that fixation coordinates refer to the mean of these 33 raw eye positions per frame. Baseline x and y coordinates were calculated as the mean fixation positions 300 ms before stimulus onset. Similar to our previous studies (e.g., Rubo and Gamer 2018; End and Gamer 2017), baseline outliers were identified by an iterative outlier removal procedure which was conducted separately for x- and y-coordinates. Specifically, the largest and smallest values were removed temporarily from the distribution. If any of these extreme values was more than three standard deviations from the mean of the remaining distribution, it was permanently excluded. Otherwise, the values were returned to the distribution. This procedure was then repeated until no more exclusions had to be performed. Subsequently, missing baselines (*M* = 9.55% of all trials across participants, *SD* = 9.16%) were replaced by the mean baseline of all valid trials and, to account for gaze drifts, baseline coordinates were then subtracted from the gaze data of each trial. Frames were excluded from analyses if the corneal reflection was lost during blinks or large eccentricity fixations and if gaze was directed towards a position outside of the video area (*M* = 2.12% of all data points for each participant, *SD* = 3.05%)

Influence of saliency, region of interests and distance to center

To investigate the influence of physical saliency on gaze allocation, we calculated saliency maps for each frame of each video. These maps were created using the Graph Based Visual Saliency (GVBS) algorithm (Harel et al. 2018) which takes luminance, color, orientation and flicker with equal weights into account and has been shown to have high prediction accuracy (Judd et al., 2012). We additionally applied Gaussian blurring along the temporal dimension of the video data to reduce the influence of strong changes in low-level saliency between successive video frames (Rubo & Gamer, 2018). These saliency values were normalized to have a mean of 1. Regions of interests (ROIs) for human and artificial faces were defined manually using circular masks. Video locations that included a face were coded as 1 whereas the remainder of the frame was coded as 0. Finally, we modelled a predictor for center bias by calculating the inverse Euclidean distance of scene locations to the center of the video.

In order to estimate the relative contribution of these predictors on gaze allocation, we aggregated data across 40 x 40 pixels patches that were arranged in a regular 32 x 18 grid. This grid size was already used in a previous study (Rubo & Gamer, 2018) and approximates the size of the functional field of the human fovea centralis at the current viewing distance. For each feature map (i.e., physical saliency, human and artificial faces, centrality), we calculated mean values for each of the 576 cells of the grid. Finally, values were z-standardized across each map to allow for comparison of the beta coefficients in the statistical analyses.

Statistical Analyses

As a first analysis, we calculated fixation durations per ROI (human and artificial faces) weighed by ROI size and ROI presentation duration per video category. To this end, we summed the number of frames per video in which the looked-at grid cell contained a human or artificial face separately for each ROI per video category and divided it by the number of pixels the ROI made up within the cell in each iteration. The resulting fixation count, corrected for ROI size, was then divided by the number of frames which contained that ROI type per video. As the average fixation durations per participant were not normally distributed, we subsequently submitted these values to two Wilcoxon signed rank tests, one contrasting human face fixations in the human video category with artificial face fixations and another contrasting the fixations of human and artificial faces in the videos in which both faces are presented simultaneously.

We furthermore determined fixation latencies as the point in time when each ROI was first fixated in each video by each participant. These values were aggregated individually for each participant, across all videos of the same type (i.e., videos containing only real faces vs. only artificial faces vs. both real and artificial faces). Some participants never looked at a ROI in some of the videos. On average, this was the case for 0.51 (*SD* = 0.95, range = 0 – 4) videos containing only real human faces and 0.14 (SD = 0.55, range = 0 – 3) videos containing only artificial faces. In the videos containing both real and artificial faces, no real face was looked at in 0.17 (SD = 0.71, range = 0 – 4) of the videos, and no artificial face was looked at in 1.17 (SD = 1.32, range = 0 – 7) of the videos. Analyses therefore focused on the subset of videos within each participant in which a specific ROI was regarded at least once. Since latencies were not normally distributed, we again performed Wilcoxon signed rank test to first compare latencies for real and artificial face fixations in the videos containing only one face type and subsequently in the videos in which both faces were presented simultaneously. Effect sizes for all Wilcoxon signed rank test were calculated according to the suggestion of Rosenthal (Rosenthal, 1994) with the formula $r = Z/\sqrt{N}$.

To more elaborately investigate the individual contributions of centrality, physical saliency, human and artificial faces on gaze allocation, we calculated nine separate generalized linear mixed models (GLMM) in R using the package lme4 (Bates et al.,

2014) and the *bobyqa* optimizer. Mixed-effect models have been explicitly suggested as an excellent tool to predict fixation patterns in naturalistic scenes based on image features (Nuthmann & Einhäuser, 2015). The criterion variable in these models was defined by the current fixation in each video frame. In order to reduce biases between looked at and not-looked-at locations in the statistical analyses, two cells of the 32 x 18 grid were selected for each video frame and used in the GLMM. This included the currently fixated cell, as revealed by the eye-tracking data, and one randomly chosen non-fixated cell. The response variable thus described whether a grid cell was fixated or not and we chose to model this binary event using a binomial error distribution and the probit link function. Centrality, physical saliency, human and artificial faces served as quantitative predictors in the models (see Figure 9 for an illustration of the procedure).

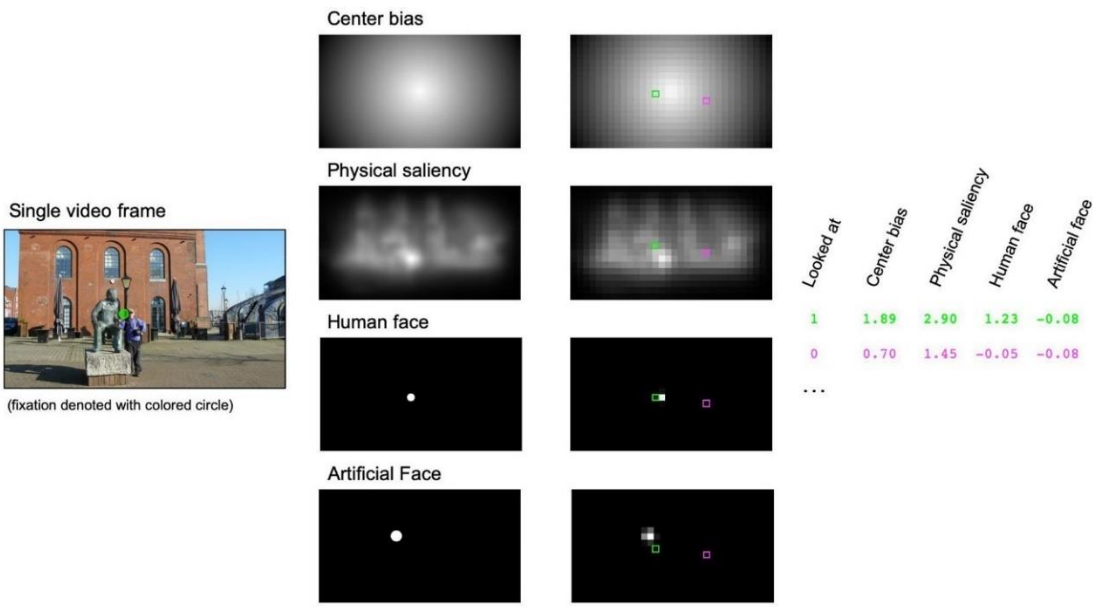

Figure 9. Illustration of the generalized linear mixed model (GLMM) approach for predicting individual fixations. A sample video frame with a possible fixation is shown on the left side. For each video frame, we defined center bias and physical saliency (as calculated by the Graph Based Visual Saliency algorithm) and circular regions of interest for human and/or artificial faces. These maps (middle column) were tiled into a regular 32 x 18 grid with individual cells reflecting the average of the raw values within each cell (right column). Finally, values were z-standardized within each map. Within the GLMM approach, we tried to predict whether a given cell was looked at (here denoted with a green square) or not (randomly selected cell marked with a magenta square) by using center bias, physical saliency and, if appropriate, the presence of human and/or artificial faces. Please note that the image depicted here was not part of the videos and is only shown for illustration.

We used an incremental approach and initially calculated a simplified model which only included distance to center and saliency values as fixed predictors for each video category. Secondly, we added the respective ROI predictors (i.e., human and artificial faces) in a separate model for each video category, yielding a total of two models for both the real human faces and the artificial face videos. For the videos containing both artificial and human faces our incremental approach yielded four different models, the simplest one including only saliency and distance to center as predictors, one model adding only one of the respective ROIs and a final model including both ROIs in addition to the saliency and centrality predictors. To account for within-subject and within-video effects, subject and video numbers were entered as random intercepts. We considered the size of beta weights (β) to estimate which predictor predominantly influenced gaze allocation and evaluated R^2 of the models to assess which model performed best. As the non-fixated grid cell was randomly chosen for each frame of each video, we decided to apply a bootstrapping procedure to validate our model outcomes and to ensure that results do not depend on an individual selection of cells. Herein, the process of randomly choosing a non-fixated grid cell was repeated over 100 iterations and 100 respective GLMMs were calculated for each of the 9 different models. Based on the results of this bootstrapping procedure, we subsequently calculated mean beta weights, mean R^2 and 95% Confidence Intervals (CIs) for each predictor and considered beta weights and R^2 significantly different from one another when the CIs did not overlap.

Results

Fixation durations

To investigate whether fixation durations differed significantly between ROIs, we first calculated a Wilcoxon signed rank test comparing fixations on human faces in the videos in which exclusively human faces were shown with fixations on artificial faces in the videos in which only the artificial faces were shown. The results revealed no significant differences between the two face types ($W = 717$, $p = 0.089$, $r = 0.23$). As the comparison rests on two entirely different sets of videos, we subsequently calculated a Wilcoxon signed rank test to compare fixation on human and artificial faces in the video category which contained both ROI types. Here, human faces were significantly prioritized ($W = 1100$, $p < .001$, $r = 1.24$, see Figure 10a).

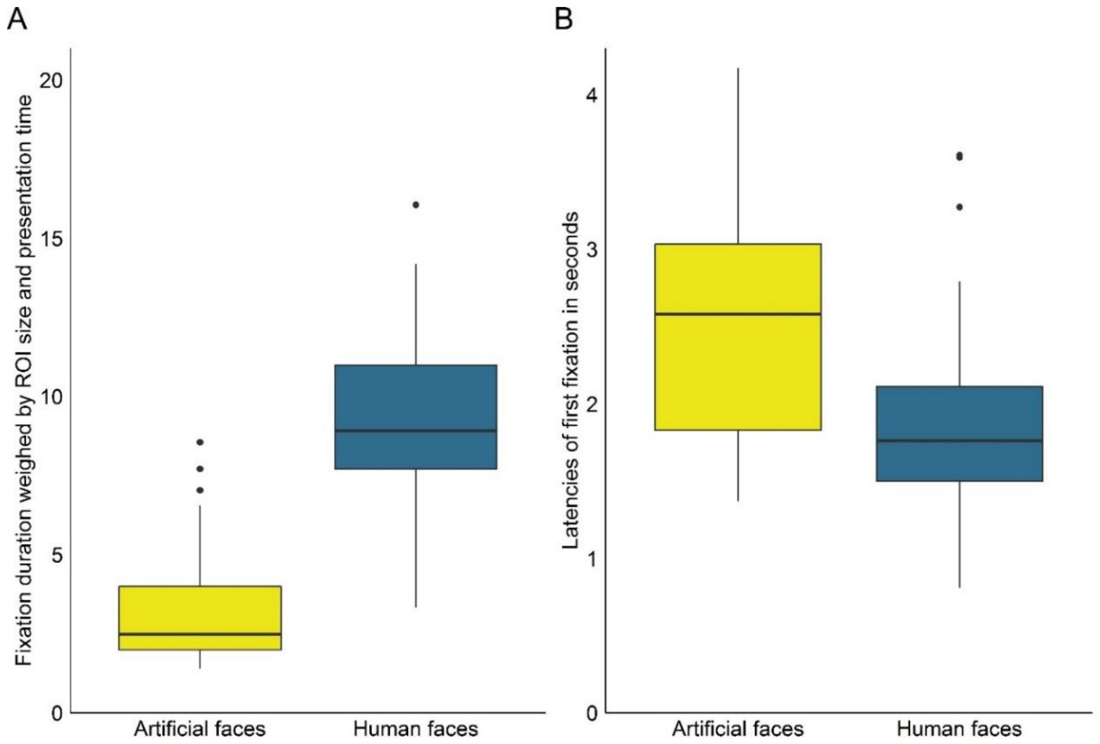

*Figure 10. A) Average duration of fixations on artificial and human faces weighed by ROI size per frame and presentation time of ROIs per video. B) Average latencies of fixations on artificial and human faces. Outliers are defined as points further than 1.5 * interquartile range of the lower or upper hinge.*

Fixation latency

In the videos in which a specific ROI was regarded at least once, the real faces were, on average, first gazed at 2.52s after video start in the videos containing only real faces (*SD* = 0.79s, range = 1.35s - 5.03s), whereas the artificial faces were first gazed at 1.27s after video start in the videos containing only artificial faces (*SD* = 0.52s, range = 0.65s - 3.24s). This difference in latencies was significant (*W*= 1199, *p* < .001, *r* = 1.15) but it should be noted that this comparison involved two different sets of video clips. In the videos containing both face types, real faces were, on average, first gazed at 1.91s after video start (*SD* = 0.65s, range = 0.81s - 3.61s) and the artificial faces, by contrast, at 2.57s (*SD* = 0.79s, range = 1.37s - 4.17s). Fixation latencies were thus significantly reduced for real versus artificial faces (*W*= 898, *p* < .001, *r* = 0.54, see Figure 10b) when both faces were presented simultaneously.

GLMM Results

We used an incremental approach consisting of nine generalized linear mixed models by which we could estimate the individual contributions of each predictor to each model for each video subset. All respective results are summarized in Table 2. Overall, the bootstrapping procedure over 100 iterations showed that both central bias and saliency greatly influenced gaze allocation throughout all video types. However, when the respective ROIs were added as predictors to the models, the explained variance increased significantly as revealed by non-overlapping confidence intervals of the R^2s. A direct comparison between real human and artificial faces in the video subset including both face types additionally showed a higher influence of real human faces ($\beta = 0.289$, 95 % CI [0.285,0.292]) than artificial faces ($\beta = 0.156$, 95 % CI [0.153,0.159]) on fixation selection while both predictors contributed significantly to gaze allocation.

Table 2. Results of incremental generalized linear mixed models (GLMMs) investigating the contribution of individual predictors to gaze patterns.

	Beta Weights of Predictors				
Video subset	Central bias	Saliency	Human Faces	Artificial Faces	R^2
Non-social videos (*n* = 15)	0.410 [0.407,0.413]	0.518 [0.514, 0.523]			0.296 [0.295, 0.297]
Only human face videos (*n* = 15)	0.209 [0.206, 0.212]	0.548 [0.544, 0.551]			0.210 [0.209, 0.212]
	0.240 [0.237, 0.243]	0.526 [0.522,0.529]	0.322 [0.317,0.327]		0.254 [0.253,0.256]
Only artificial face videos (*n* = 15)	0.198 [0.195, 0.201]	0.440 [0.437, 0.443]			0.180 [0.179, 0.181]
	0.164 [0.161, 0.167]	0.431 [0.428,0.433]		0.205 [0.202, 0.208]	0.204 [0.203, 0.205]
Human and artificial face videos (*n* = 15)	0.142 [0.139, 0.145]	0.483 [0.479, 0.486]			0.160 [0.159, 0.161]
	0.135 [0.132, 0.138]	0.438 [0.434, 0.441]	0.277 [0.273,0.280]		0.213 [0.212, 0.215]
	0.145 [0.141, 0.148]	0.456 [0.453, 0.460]		0.131 [0.128, 0.134]	0.172 [0.171, 0.173]
	0.139 [0.135,0.141]	0.398 [0.394,0.401]	0.289 [0.285, 0.292]	0.156 [0.153,0.159]	0.230 [0.229, 0.232]

Mean beta weights and explained variance (R^2) for models comprising an increasing number of predictors. Models are nested and include predictors in models shown above for the specific set of videos. All values were calculated by bootstrapping 100 sets of not-looked-at grid cells and performing GLMMs for each set. Estimates represent means of weights from each bootstrapping iteration. Values in brackets represent the 2.5th and 97.5th percentile rank as an unbiased estimate of the 95% confidence interval.

Discussion

It is generally established that faces elicit an attentional bias towards them. In the current study, we examined whether this attentional bias persists for various face types or whether the presence of real human and artificial faces differentially impacts gaze allocation when viewing videos of complex, naturalistic scenes. While both face types significantly predicted gaze, the relative influence of artificial faces was reduced when real human faces were presented simultaneously. This result was also evident in longer fixation durations on and faster gaze orienting towards real human faces suggesting that real faces are more relevant to observers than artificial ones.

These findings add to pre-existing knowledge on social attention by disentangling the contributions of different face types in naturalistic scenes. Previously, a general strong prioritization of social features (e.g., human heads or bodies) had been described in the literature (Bindemann et al. 2005; Flechsenhar and Gamer 2017; Flechsenhar et al. 2018; Rösler et al. 2017; End and Gamer 2017; Birmingham et al. 2008; Coutrot and Guyader 2014), yet the use of stimulus material varied widely. While many researchers relied on isolated or schematic, artificial faces (e.g. Bindemann, Mike Burton, and Langton 2008; Theeuwes and Van der Stigchel 2006; Bindemann et al. 2005), others employed static or dynamic stimuli representing real humans in naturalistic settings (e.g. Birmingham et al. 2008; End and Gamer 2017). In order to be able to systematically differentiate between artificial and real human faces, we utilized videos containing either only one of the two face types or both human and artificial faces. We were thereby able to see that artificial faces predict gaze when presented exclusively and remain to influence fixations patterns when presented in competition with real human faces. In direct contrast to real human faces, artificial faces yet attracted gaze considerably less as reflected by an enhanced fixation latency, a substantially lower average fixation duration and beta estimate.

These findings are seemingly at odds with a study by Laidlaw and colleagues who used mobile eye-tracking to differentiate gaze patterns when participants viewed either a real or a video-taped person in a waiting room scenario (Laidlaw et al., 2011). As observers fixated the video-taped person displayed on a PC screen more frequently than the live person in the room, it was hypothesized that humans might reduce eye contact when it could lead to a social interaction. A potential interaction is indeed one key difference between real human and artificial faces, yet in our study none of the two face types truly give room for an interaction. While it is therefore not surprising that we cannot replicate the effects observed by Laidlaw and colleagues, it would be interesting to investigate the impact of real and artificial faces in live conditions. The general predictive power of artificial faces observed in our study is yet in line with a previous observation that eyes attract gaze even when they are presented on non-human monsters and independent of where they are located on the body (Levy et al. 2013). Similarly, studies investigating human-robot-interactions have shown that people can make use of referential gaze cues elicited by robots (Mutlu et al., 2009) and that this gaze-following already becomes evident during infancy and occurs even for non-humanoid robots (Movellan & Watson, 2002). These findings are further corroborated by a recent study which reported a preference of face-like stimuli in the human fetus, suggesting that our tendency to fixate face-like structures evolves *a priori* (Reid et al., 2017). Face processing is indeed known to occur holistically such that different components of a face are integrated and interpreted together (Goffaux & Rossion, 2006; Maurer et al., 2002; Van Belle et al., 2010). Face inversion disrupts this process leading face recognition accuracies to drastically decline when inverted faces are presented (Yin, 1969). While the vast majority of studies examined face processing in two-dimensional faces, it was recently reported that recognition is improved for 3D versus 2D faces but not when they are inverted (Eng et al., 2017). This refined recognition is likely due to improved holistic processing when faces are more realistic and depth information is enriched. Similar enhanced holistic processing effects might underlie the increased fixations on real versus artificial faces in the current study which would potentially result in less pronounced gaze differences when both face types are inversed.

The use of generalized linear mixed models further enabled us to investigate the relative contributions of additional predictors on gaze patterns, while allowing for correlations between the individual predictors. Nuthmann and Einhäuser suggested this framework as particularly advantageous for the analysis of gaze during the observation of complex stimuli as their low-level features often tend to be correlated (Nuthmann & Einhäuser, 2015). In all of our models, however, low-level saliency contributed critically to gaze allocation and was even seen to explain eye movements significantly better than faces. It is generally known that both low-level physical saliency and higher-level semantic saliency contribute to attentional selection (Santangelo et al. 2015; Flechsenhar and Gamer 2017; Henderson et al. 2007; Einhäuser, Spain, and Perona 2008) and a recent review showed that both contribute to the likelihood of an item being remembered (Santangelo, 2015). Considering the higher-level semantic relevance of faces in social scenes, the substantial role of both lower-level physical saliency and the presence of faces in the prediction of gaze provide further support for models which claim that perceptual and semantic saliency drives attentional allocation. Contrary to our observation, Coutrot and Guyader found that faces most prominently influenced eye movements of participants who viewed dynamic conversations, whereas saliency did not crucially account for the recorded gaze (Coutrot & Guyader, 2014). However, while our videos did not contain any relevant auditory information, Coutrot and Guyader solely presented conversations rendering the faces displayed in the scene even more relevant to the understanding of its gist. The observed discrepancy in results once again stresses how many factors need to be taken into account when attempting to investigate the mechanisms underlying gaze allocation in naturalistic scenes. Multisensory approaches as, for instance, employed by Nardo and colleagues to study spatial attention, might therefore be helpful in disentangling the various factors influencing the perception of faces in complex naturalistic scenes (Nardo et al., 2014).

The examination of fixation preferences for real human and artificial faces can also further our understanding of mental disorders in which alterations of gaze behavior are implicated. Although children are overall more susceptible to distractions by physically salient image regions than adults (Cavallina et al., 2018), children with autism spectrum disorder display particularly decreased attention to fellow humans, especially faces, (Dawson et al., 1998, 2004) and are less likely to follow gaze than their peers (Leekam et al. 2000). These difficulties do not decline with age (Baron-Cohen et al., 2001; Spezio et al., 2007) and it is generally assumed that the higher-level saliency of social features is reduced for patients with autism-spectrum disorder (Dawson et al., 1998; Klin et al., 2003; Wang et al., 2015). There is some evidence that reduced social attention in autism does not transfer to artificial faces since children with autism spectrum disorder were seen to use regular processing strategies for cartoon faces while processing real faces atypically (Rosset et al., 2008). Additionally, healthy peers performed better in a discrimination task when presented with real versus cartoon faces, whereas patients with autism spectrum disorder did not exhibit a difference in performance (Rosset et al., 2010). The current study provides additional information on gaze allocation towards real and artificial faces than previously established and thereby offers a more elaborate framework for the examination of gaze alterations in autism.

One potential pitfall of our experimental design is that the majority of artificial faces did not exhibit movement. Although their position within the video could change because of smooth pan shots or slight camera movements, real human faces were more likely to move. However, the Graph Based Visual Saliency algorithm, which we used to calculate the physical saliency of different image regions, considers movement across frames. In our statistical model, we were therefore able to take a disparate percentage of motion between face types into account and thus assume that the differences in gaze behavior towards artificial versus real faces cannot be solely explained by motion. Additionally, we need to bear in mind that faces are typically connected to bodies - the extent of which might differ between artificial and real faces in our study. While we attempted to find comparable stimulus material, artificial faces were more frequently presented without being connected to a meaningful bodily extension. Various studies investigating gaze patterns in social scenes (e.g. End and Gamer 2017; Flechsenhar and Gamer 2017; End and Gamer 2019) have yet shown that faces attract decisively more fixations than other body parts, and we hence believe that differences in the presence of extremities do not influence our findings gravely.

To conclude, the current study used multiple generalized linear mixed models to identify several crucial predictors of gaze allocation when viewing complex dynamic scenes. Saliency and central bias had highest predictive power, while both real human and artificial faces also substantially contributed to the prediction of gaze patterns. Taken together, these findings shed further light on the mechanisms underlying the distribution of social attention and highlight the role both real human and artificial faces play in the visual exploration our surroundings.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

MR and MG developed the study concept and design. MR performed and supervised data collection. MR and LR analysed and interpreted the data under supervision of MG. LR drafted the manuscript and MR and MG provided critical revisions. All authors approved of the final version of the manuscript.

Funding

This work was supported by the European Research Council (ERC- 2013-StG #336305).

Acknowledgements

The authors thank André Wittenborn and Michael Strunz for their help in stimulus preparation and data collection as well as Aki Schumacher for providing brief descriptions of the videos for the supplementary material.

SOCIAL FEATURE LOCATION IN REAL-WORLD SCENES CAN BE DECODED FROM EARLY VISUAL CORTEX

Lara Rösler & Matthias Gamer

Department of Psychology, Julius Maximilians University of Würzburg, Würzburg, Germany

Manuscript in preparation for submission.

Abstract

Humans rapidly allocate attention to social features (i.e. human heads or bodies) in complex naturalistic scenes. Various eye-tracking studies have confirmed that this social attention takes place reflexively and independently from the physical saliency of competing image areas. What remains unknown, however, is which neural mechanisms facilitate this rapid attentional prioritization of social information. A viable candidate region is the amygdala which might modulate local activity in cortical regions and thereby facilitate saccade preparation or execution towards social features. Alternatively or additionally, the categorization of social and non-social areas might be achieved in the visual cortices which are known to parse the visual elements of our environment. In the current study, we presented naturalistic scenes with social features in one quadrant of the visual field for 200 ms to 36 participants while simultaneously recording eye movements and brain activity using functional magnetic resonance imaging. On the behavioral level, participants made significantly more saccades towards social cues than a distribution of saccades at chance level would imply. The location of social features in the image could be decoded from early visual cortex but not amygdala activity patterns. We were further unable to detect changes in amygdala activity as a response to saccade events. Collectively, our findings replicate the reflexive prioritization of social elements in complex scenes and show that this mechanism is reflected in early visual cortex activity.

Introduction

Social features, in the form of human faces or bodies, readily grab our attention. When presented with pictures or videos of complex scenes, we preferentially scan these social features (e.g. Birmingham, Bischof, & Kingstone, 2008; End & Gamer, 2017; Rubo & Gamer, 2018) in a reflexive manner (Rösler et al., 2017) and independent of task demands (Flechsenhar & Gamer, 2017; Flechsenhar, Rösler, & Gamer, 2018). While this automatic processing of social information is a vital prerequisite to successful interactions with fellow humans (Langton & Bruce, 1999), the neural mechanisms enabling social attention are not sufficiently understood. Various studies have highlighted an involvement of the amygdala in higher-order social cognition; accordingly, the amygdala plays a central role in emotional processing (Sergerie et al., 2008), reacts selectively to parts of faces (Rutishauser et al., 2015) and is particularly known for its contribution to automatic threat detection (Öhman, 2005).

A growing number of studies yet points to a role of the amygdala which transcends the detection and processing of emotional signals. Evidence from electrophysiological recordings in primates suggests that the amygdala integrates reward-relevant and spatial information as it responded with enhanced activation dependent on reward-magnitude and location of presented stimuli (Peck et al., 2013). These findings are corroborated by studies in humans in which amygdala activation was associated with the execution of saccades towards facial features (Gamer et al., 2010; Gamer & Büchel, 2009; Scheller et al., 2012). The amygdala might thus facilitate attentional shifts towards motivationally relevant aspects of our surroundings (Adolphs, 2006, 2010). Indeed, a patient with amygdala lesion, initially unable to detect fearful faces, could overcome this deficit in fear recognition when instructed to fixate the eyes of viewed faces (Adolphs et al., 2005). It is hence conceivable that amygdala activation lies at the core of the neuronal processes enabling social attention.

Another brain region that is known to be involved in the categorization of visual input and has repeatedly been implicated in attentional processes is the visual cortex. As evidenced by numerous electrophysiological studies, attended stimuli trigger larger responses in visual cortices than unattended information (e.g. Reynolds et al., 2000; for a review see Ungerleider, 2000). In other words, attention-capturing stimuli are more likely to receive cortical representation (Desimone & Duncan, 1995). Within the visual cortex, a ventral information processing stream originating in V1 is assumed to register previously unattended stimuli and monitor attentional shifts towards these inputs (Corbetta & Shulman, 2002), thereby contributing to visual object recognition (Mishkin et al., 1983). The ventral processing stream has further been associated with attention priority maps (Sprague & Serences, 2013). Several priority maps are assumed to co-exist the brain, encoding information about areas in the visual field that carry behavioral importance (Fecteau & Munoz, 2006; Itti & Koch, 2001). The traditional view holds that early visual cortex areas contain priority maps which primarily contain information about physical saliency of the observed scene, while later cortical areas reflect the behavioral or contextual relevance of stimuli (Serences & Yantis, 2006). Mo and colleagues yet demonstrated that differential first saccade targets, which are markers of behavioral relevance, could be decoded in the early visual cortex (Mo et al., 2018). They presented upright, inverted or scrambled faces to participants and used the target of the first saccade as a marker of priority. Using population receptive field mapping, these first saccade targets could be predicted in the primary visual cortex (V1), as well as extrastriate (V2/3) cortices. Saccade target and reconstructed neural representation increased from V1 to V2/3 for upright but not for inverted faces, suggesting that the appropriate image configuration contributes to its behavioral relevance. The attentional shifts towards and recognition of social features in complex scenes might thus be, at least in part, mediated by the early and extrastriate visual areas.

To test these assumptions, we simultaneously recorded eye movements and brain activity of participants using fMRI while presenting complex social scenes for a very brief duration of 200 ms. In a previous eye-tracking study, we had already established that participants reflexively attend to social information within these images (Rösler et al., 2017). In the present study, we aimed at replicating these behavioral results and performed a multivariate pattern analysis (MVPA) in early and extrastriate visual cortices and amygdala to examine whether multivariate activation within these regions is associated with reflexive social attention.

Methods

Participants

We recruited 40 participants via an online recruitment system hosted by the university. We aimed at obtaining a final sample size of 36 to detect a within-subject effect of at least medium size ($d > 0.5$) with 80% power using a two-tailed one-sample or paired t-test. One participant ended the experiment prematurely and therefore had to be excluded from the analysis. Another participant had to be excluded because of a neurological condition that only became apparent on the day of testing. Lastly, one participant had to be excluded from the fMRI analysis because visual inspection revealed issues with MR quality. Another participant had to be excluded from all analyses because of too many missing baselines in the eye-tracking data, rendering it unclear to what extent the stimuli were attended to. This resulted in a final sample size of 37 participants (males = 17; mean age: *M* = 25.83 years, *SD* = 5.80) for the behavioral analyses and 36 participants for the fMRI analyses without eye-tracking regressors. The fMRI analysis using eye-tracking regressors consisted of 34 participants because another 2 participants were missing eye-tracking data for one and two runs, respectively. Ethical approval was obtained by the ethics committee of German Psychological Society (DGPs) and performed in compliance with Declaration of Helsinki guidelines. All participants provided written informed consent and received monetary compensation.

Experimental Paradigm and Procedure

We presented photographs of naturalistic scenes with social and non-social content to participants for 200 ms. In total, participants underwent 5 experimental runs which each consisted of 106 images presented in pseudo-randomized order. These included 80 social images depicting parts of one or multiple human beings, 20 non-social images which did not contain any human features and 6 colored fractal images. The images were obtained from various image databases (NAPS; Marchewka et al., 2014; Spanky fractal database) and the Internet (e.g. Google picture search, flickr) and, where necessary, reduced to a common size of 800 x 600 pixels. Social images were chosen in such a way that the human features were always restricted to one quadrant of the image. By using mirroring and different cut-outs, we created four different versions of each image with the social feature being depicted once in each quadrant. This procedure enabled the subsequent quadrant-based analysis of social information processing. Additionally, we used a saliency-algorithm developed by Itti & Koch to ensure that mean saliency was balanced across quadrants and did thus not differ systematically between social and non-social quadrants (for more details on stimulus selection and preparation see Rösler et al., 2017). To ensure that participants remained attentive throughout the entire experiment, we instructed them to react with a button press whenever a fractal was shown. Each trial began with the presentation of a fixation cross for random period between 1 and 8 seconds, followed by the actual stimulus displayed for 200 ms and a subsequent blank screen shown for 1800 ms during which a response could be made. The eye-tracker was re-calibrated before each of the 5 functional runs.

Apparatus

Eye-Tracking

Eye movements were recorded with a camera-based MRI compatible eye-tracker (SR Research, EyeLink 1000 Plus) at a sampling rate of 250 Hz. Stimulus presentation and eye movement recording was monitored via the Presentation software (17.0, Neurobehavioral Systems). Stimuli were displayed centrally on a grey background of a 31.55” BOLD LCD screen (Cambridge Research Systems, 698.4mm x 392.9mm; 1920 x 1080 Pixel, 120 Hz). The distance from eye to screen amounted to 120 cm yielding a 12.45°x10.45° degree of visual angle of the stimuli. Manual reactions to target stimuli were recorded via a 5-button response box.

Functional Imaging

Functional images were acquired with a 3-Tesla MR-scanner (Siemens, Magnetom Skyra) equipped with a 32-channel head coil. 40 transverse slices (thickness: 2 mm; 1 mm gap) were recorded in each volume using a T2*-sensitive gradient echoplanar imaging (EPI) sequence (repetition time: 2470 ms, echo time: 26ms, flip angle: 80°, field of view: 220 x 220 m^2, in-plane resolution: 2x2 mm^2) . Additionally, high-resolution (1 x 1 x1 mm^3) structural images using a T1-weighted gradient echo sequence with 240 slices were acquired for each participant at the end of the scanning session.

Data Processing

Eye-Tracking

Saccades and fixations were read out for each trial. Saccades were defined as eye movements with a velocity of at least 30° per second or an acceleration of at least 8000°/s and fixations were defined as the breaks in between saccades. These eye movements were then further processed using R (version 3.3.2, Core Team, 2016). Following the same procedure as in our previous study (Rösler et al., 2017), we extracted the first saccade after stimulus onset of all trials in which a stable baseline was present. To determine baseline stability, we evaluated the first 300 ms prior to stimulus onset which is when the fixation cross was presented. For each participant, we conducted an iterative outlier removal procedure treating x and y- baseline coordinates separately. The smallest and the largest values from all baseline coordinates were temporarily removed from the distribution and permanently excluded if their values were more than three standard deviations away from the mean baseline coordinates of the remaining data. This procedure was repeated with the remaining baseline coordinates until no more exclusions were necessary.

Subsequently, saccade x and y coordinates were drift-corrected by subtracting the baseline from the actual values. Only those trials in which a first saccade occurred between 150 and 1000 ms with an amplitude of at least 0.5° of visual angle were considered for the final analyses.

Functional Imaging

Raw DICOM data images were converted to NIFTI format and organized according to the 'Brain Imaging Data Structure' specifications (BIDS; Gorgolewski et al., 2016). Preprocessing was conducted using FMRIPREP (Esteban et al., 2019). Within the FMRIPREP framework, each T1 weighted volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.1.0 (Tustison et al., 2010) and skullstripped using antsBrainExtraction.sh v2.1.0 (using OASIS template). Spatial normalization of the skullstripped T1w volume to the CBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009) was performed using nonlinear transformation implemented in ANTs v2.1.0 (Avants et al., 2008). Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL v5.0.9, RRID:SCR_002823; Zhang et al., 2001). Functional data was slice time corrected using 3dTshift from AFNI v16.2.07 (Cox, 1996) and motion corrected using mcflirt (FSL v5.0.9 ; Jenkinson et al., 2002). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve & Fischl, 2009) with six degrees of freedom, using flirt (FSL). Motion correcting transformations, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.1.0) using Lanczos interpolation.

Physiological noise regressors were extracted applying CompCor (Behzadi et al., 2007). Principal components were estimated for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). A mask to exclude signal with cortical origin was obtained by eroding the brain mask, ensuring it only contained subcortical structures. Six tCompCor components were then calculated including only the top 5% variable voxels within that subcortical mask. For aCompCor, six components were calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (Power et al., 2014) was calculated for each functional run using the implementation of Nipype.

Statistical Analyses

Eye-Tracking

To test whether we could replicate the behavioral findings of Rösler et al (2017), we first analyzed saccadic latencies as a function of saccade target by counting the number of saccades in 50 ms bins ranging from 150 ms to 1000 ms for saccades targeting social and non-social quadrants separately. To investigate whether frequencies of saccadic latency differed between these social and nonsocial target saccades, we performed a 2 (saccade target) x 17 (50 ms saccade latency bins) ANOVA. Additionally, we calculated the percentage score of saccades targeting the social feature for each social feature position (left upper, left lower, right upper and right lower quadrant). If the presence of social features in a quadrant did not influence saccade allocation, we would expect a chance level distribution of saccades targeting the social quadrant (i.e. 25%). In order to investigate whether saccades landed significantly more often in quadrants containing social elements than chance level would suggest, we subtracted 25% from the four percentage scores of each participant. We then submitted these values to a 2 x 2 repeated measures ANOVA with the factors horizontal (left versus right) plane and vertical (upper versus lower) plane of the saccade to investigate whether the distribution differed between quadrants (main and interaction effects) as well as from chance level (intercept of the ANOVA).

Functional Imaging

We performed multivariate analyses to test whether changes induced by the experimental manipulation are reflected in distinct patterns of voxel activity in our regions of interest. To this end, we first used general linear models (GLM) to conduct first-level analyses for each individual subject. To test which brain regions were activated by the position of a social feature in the image, we created subject-specific regressors indicating which social feature location was associated with which onset time. For each of the four positions, we created a different design matrix describing onsets of one specific positions against all others. Onsets of non-social images and fractals were included as nuisance regressors. Taking into account the available eye-tracking data, we analyzed another design matrix per participant with separate regressors for trials in which a saccade was made and in which no saccade was made. A last design matrix was analyzed with separate regressors for saccades targeting the social feature position and saccades that target non-social image areas. For all first-level analyses, functional images from all five sessions were concatenated and five session-specific constants were further included in the model.

Subsequently, we used the Decoding Toolbox (Hebart et al., 2015) to decode distinct activation patterns between contrasts of interest (upper left social feature position versus other, upper right social feature position versus other, lower left social feature position versus other, lower right social feature position versus other, trials in which a saccade was made versus trials in which no saccade was made, trials in which the first saccade targeted the social quadrant versus trials in which the first saccade targeted other quadrants) using pairwise classification. Here, a support vector machine (LIBSVM) trained and classified data from four of the five runs, with the remaining run was used to test the classifier. For each task pair, a five-fold cross-validation was used in which this procedure was repeated for each one of the five left-out runs. The area under the receiving-operating characteristic curve (AUC) for a given ROI was averaged across these folds, which resulted in one AUC value for each task pair in each ROI, in each individual as a measure of classification performance. Based on previous literature, our regions of interest comprised early visual cortices along the ventral processing stream (V1, V2, V4) and the amygdala (see Figure 11) which we obtained from the SPM anatomy toolbox (version 2.1; Eickhoff et al., 2007) . To test whether the position of the social features can be decoded from these chosen ROIs, we averaged the AUC values for the four respective social feature locations to code the overall response to social feature position. Since Allefeld and colleagues recently argued that t-tests against the null are inadequate for the statistical evaluation of classification accuracies given their non-normal distribution (Allefeld et al., 2016), we opted for permutation testing to generate tests statistics. Accordingly, we shuffled regressor labels and onsets a 100 times per participant to generate a 100 AUC values based on permuted data. These were then averaged within participants and, subsequently, the permuted distribution containing averaged 100 AUC values of all participants was tested against the 100 AUC values resulting from the true regressor labels.

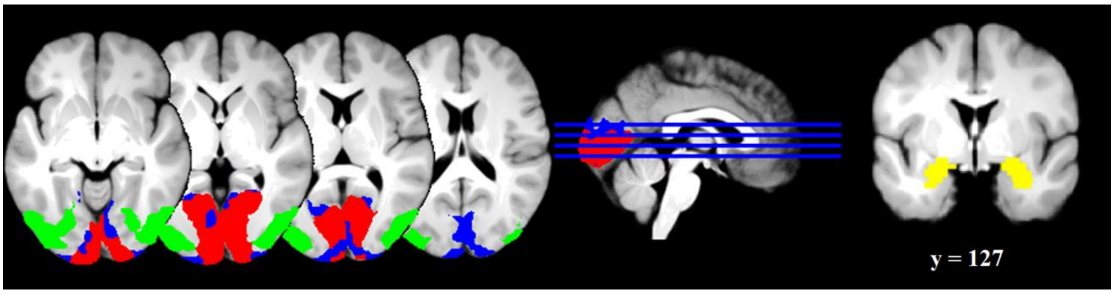

Figure 11. Regions of interest. Red = V1; blue = V2; green = V4, yellow = amygdala.

Results

Behavioral Analyses

Subjects responded with a mean saccade latency of 601.14 ms (*SD* = 210.10 ms) to the appearing stimuli. A 2 (social or non-social saccade target) x 17 (50 ms saccade latency bins) ANOVA on saccade frequency revealed a significant interaction between saccade target and saccade latency bin ($F(16,576) = 15.88$, , $\varepsilon = 0.37$, $p < .001$, $\eta_p^2 = .08$). In line with our previous finding reported in Rösler et al (2017), saccades seem to follow a bimodal distribution with saccades towards social quadrants occurring earlier than saccades targeting non-social quadrants (see Figure 12A). The analysis also revealed a main effect of saccade latency bin on saccade frequency ($F(16,576) = 11.63$, $\varepsilon = 0.29$, $p < .001$, $\eta_p^2 = .16$) and a main effect of saccade target ($F(1,36) = 50.91$, $p < .001$, $\eta_p^2 = .07$). A 2 x 2 ANOVA, investigating the percentage scores of saccades to quadrants with social information, revealed that participants looked significantly more often at social features than a chance distribution of saccades would suggest, as indicated by an intercept significantly different from 0 ($F(1,36) = 48.18$, $p < .001$). We also observed a main effect of horizontality (F(1,36) = 6.22, p = .02, ηp2 = .15) and a main effect of verticality (F(1,36) = 12.06, p = .001, ηp2 = .25) as participants preferred to look up rather than down and had a tendency to look left rather than right when the social feature was presented in the respective hemifield (see Figure 12B).

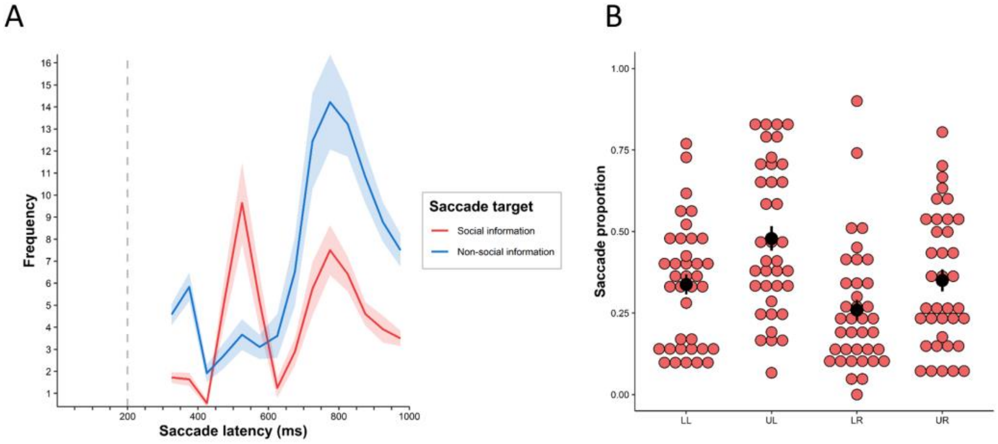

Figure 12. Behavioral evidence of social feature prioritization. A) Latencies of saccades towards social and non-social image regions. Distribution of the mean frequency of saccades towards social image regions (in red) and saccades towards non-social image regions (in blue) per 50 ms latency bin. Shaded areas are defined by the standard errors of the mean. The dashed grey line indicates stimulus offset. B) Proportion of saccades targeting social image regions. Here, proportions were calculated for all trials in which social features appeared in one of the four quadrants (LL = lower left, UL = upper left, LR = lower right, UR = upper right). Each circled red dot represents one participant. Black dots denote the mean proportion of all participants and error bars depict the standard error of the mean.

Representation of social feature position in early visual cortices but not the amygdala

After having established that quadrants containing a social feature are preferentially looked at, we set out to test whether the social feature position can be predicted from activation patterns in the early visual cortices and the amygdala using MVPA. This analysis (Figure 13) demonstrated significant decoding of social feature position in early visual cortex area V2 (t_{35}=2.24, p = 0.016) and associate visual cortex area V4 (t_{35}=5.202, p < 0.001) but only marginally significant classification in early visual area V1 (t_{35}=1.62, p = 0.058). We were further unable to classify social feature position from activity in the amygdala (p= 0.422).

As an additional measure of social attention, we attempted to classify trials in which a saccade occurred from trials in which no saccade occurred. However, the resulting decoding accuracies (Figure 14) did not differ from chance in any visual area (all *ps* > 0.107), nor in the amygdala (p = 0.103). We additionally attempted to decode trials in which a saccade targeted a quadrant containing a social feature from trials in which a saccade targeted on of the remaining quadrants from amygdala activity. Again, we were unable to detect decoding accuracies that differed from chance (p = 0.065).

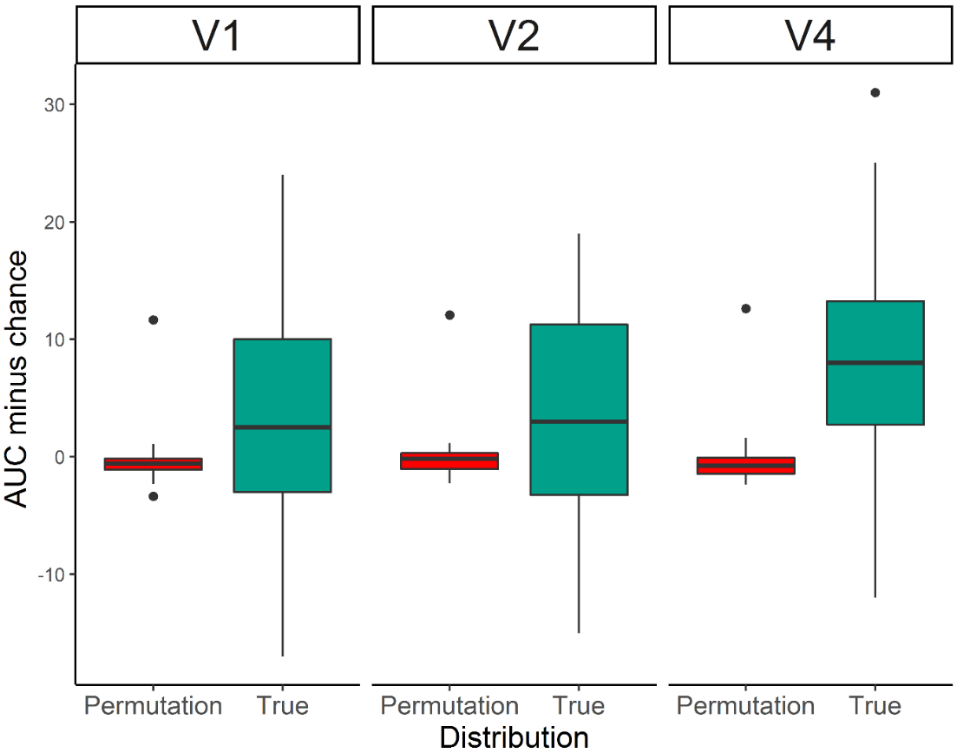

Figure 13. Neural representation of social feature position. Decoding accuracies, as reflected by area under the receiving-operating characteristic curve (AUC) minus chance, measured in the early ventral processing areas V1, V2 and V4 for the prediction of social feature position (n = 36 participants). Decoding accuracies for pairwise comparisons between every one of the four image quadrants containing social features versus the non-social quadrants were averaged within individuals to represent overall encoding of the social feature position. These accuracies (here depicted in green) were tested against decoding accuracies resulting from the average of 100 permutations (here depicted in red) per individual.

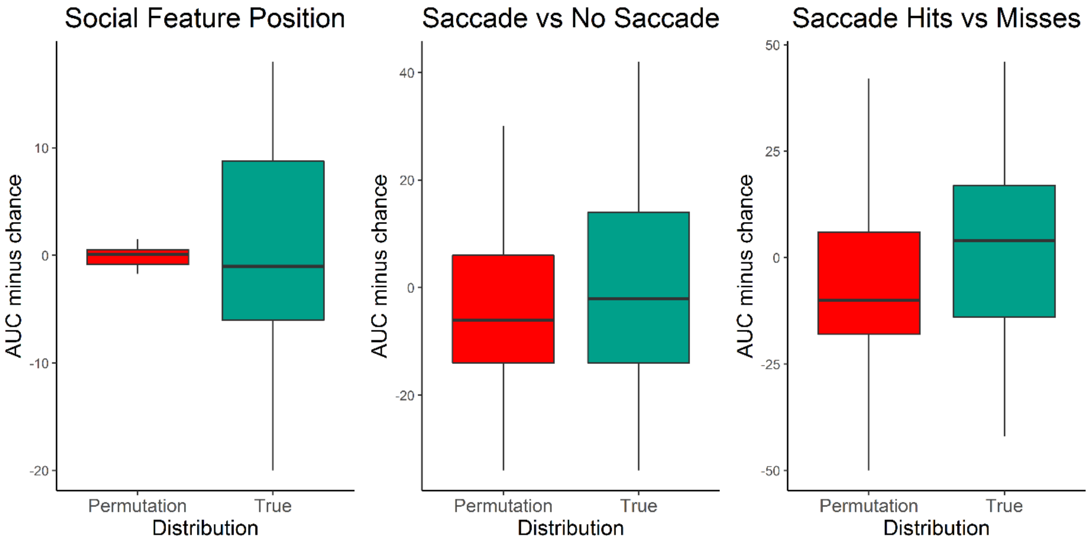

Figure 14. Measures of social attention cannot be classified in the amygdala. Decoding accuracies, as measured by area under the receiving-operating characteristic curve (AUC), for different measures of social attention. True decoding accuracies (green) did not differ from decoding accuracies resulting from a permuted distribution for the classification of social feature position (n = 36 participants), trials with a saccade versus no saccade (n = 34 participants) and trials in which the saccade targeted the quadrant containing a social feature versus a non-social quadrant (n = 34 participants).

Discussion

Social features in complex naturalistic image are processed preferentially and swiftly. In the present study, we demonstrate that this automatic prioritization of social features might be in part achieved through early visual cortical areas. Using a multi-voxel pattern analysis, we show that the position of a social feature in a complex naturalistic image can be decoded from striate and extrastriate visual cortex activity. In contrast, we did not find evidence for a representation of social feature position in the amygdala, nor were we able to decode saccade-related activity in amygdala activation patterns. These results are in line with the traditional idea of an involvement of occipital areas in attention priority maps of visual scenes (Mo et al., 2018; Preston et al., 2013) but provide novel insights into the spatial encoding of social features in real world scenes.

Previous literature has demonstrated that the ventral visual pathway crucially contributes to the high-level interpretation of visual input. In the current study, we show that both early (V1 + V2) as well as extrastriate (V4) visual areas encode the position of social feature in a complex scene which is presented for only 200 ms. These findings extend previous findings of attention priority maps in the extrastriate visual areas (Mo et al., 2018; Sprague & Serences, 2013) by demonstrating that early visual areas seemingly parse complex naturalistic scenes according to the relevance of their image regions. Although we were unable to show a coupling of behavior and neural activity patterns on a trial-by-trial level, the decoding of social feature location corresponds to our behavioral findings of increased first saccades to quadrants containing a social feature in the presented complex scenes. Our findings thereby demonstrate two crucial properties of attention priority maps (Serences & Yantis, 2006) – (1) an increase of the cortical representation of the relevance of social stimuli along the ascension of the visual cortex and (2) a matching of behavioral and neural representation of relevance (i.e. social features are looked at more often and earlier in time and can be decoded from visual cortex activity). Image configuration was already reported to influence extrastriate priority maps, as first saccade targets on upright faces received increase neural representation than those on inverted faces (Mo et al., 2018). Similarly, the visual system (fusiform face area and inferior occipital gyrus) demonstrates enhanced representation of faces when they are shown on typical location in the visual field, in line with natural scene viewing (de Haas et al., 2016). Our results provide a relevant extension of this finding by showing that differential locations of social feature in complex scenes can be decoded from early visual cortex.

Various studies have highlighted a role of the amygdala in social cognition more broadly and, recently, a growing number of scientists pointed to an involvement of the amygdala in the allocation of attention to relevant stimuli (Gamer & Büchel, 2009; Méndez-Bértolo et al., 2016; Peck et al., 2013; Peck & Salzman, 2014). This interpretation does away with the long-standing idea that the amygdala predominantly helps to process threatening or emotionally salient information by suggesting its more general role in spatial decoding and attentional allocation to relevant features in our environment. Our current study provides no evidence for this account as we were unable to decode any measure of social attention from amygdala activity. This finding is in line with a recent lesion study in which aberrant visual exploration in autism spectrum disorder (ASD) could not be explained by amygdala functioning as a patient with an amygdala lesion showed visual exploration patterns more similar to controls than to patients with ASD (Wang, 2019). Numerous recent electrophysiological studies providing evidence for a correspondence between amygdala activity and the facilitation of saccades to emotionally or contextually relevant stimuli (Maeda et al., 2019; Peck et al., 2013; Putnam & Gothard, 2019) in primates yet beg the question whether our results represent a true null finding or a false negative finding influenced by methodological issues. Amygdala BOLD responses to emotional tasks were indeed reported to be least reproducible of a series of investigations into the reproducibility of task-related activations of various regions of interest (Johnstone et al., 2005; Lipp et al., 2014; Plichta et al., 2012). It is yet also conceivable that the social features in our scenes were not relevant enough to be mirrored in amygdala activations patterns. Possibly the use of a design with higher ecological validity, such as a virtual environment scenario, would yield decodable saccade-related activity because the social features in these scenes would be of imminent relevance with possible behavioral consequences to the observer.

Although this study substantially adds to previous insights into the neural processing of social features in complex scenes, there are many questions that remain unanswered. While our inability to decode social features positions and the occurrence of the amygdala might represent a true finding, it is also possible that social attention is only associated with specific sub-nuclei within the amygdala. In total, the human amygdala comprises at least certain functionally and connectivity distinct nuclei (Freese & Amaral, 2009) and previous electrophysiological studies in primates demonstrated that specific subpopulations of amygdala neurons respond to relevant stimuli (Hoffman et al., 2007; Peck et al., 2013). More conclusive insights into the involvement of the human amygdala in social attention might therefore be gained by using multiple regions of interest for the different amygdala sub-nuclei. Additionally, for a complete picture of how social attention is accomplished by the brain, it is crucial to understand how different brain regions cooperate to achieve the fast prioritization of social stimuli. We do currently not know whether the neural representation of social feature location we observe in striate and extrastriate visual cortex stem from top-down modulations or are the mere result of bottom-up processes. Future research should therefore consider examining the functional and structural connectivity across hypothesized neural hubs of social attention.

In summary, the present study demonstrates decodable social attention in striate and extrastriate visual cortices. We show that the location of social features in complex social scenes can be inferred from early and late visual cortex pattern activity, whereas a possible role of the amygdala in social attention allocation cannot be confirmed by the current observations. Our findings extend previous literature on attention priority maps and help to better understand the mechanisms that contribute to social information perception in complex environments.

SOCIAL ANXIETY IS ASSOCIATED WITH ALTERATIONS IN HEART RATE BUT NOT GAZE IN A REAL SOCIAL INTERACTION

Lara Rösler, Stefan Göhring, Michael Strunz, Matthias Gamer

Department of Psychology, Julius Maximilians University of Würzburg, Würzburg, Germany

Manuscript under review.

Abstract

Much of our current understanding of social anxiety rests on the use of simplistic stimulation material in laboratory settings. Latest technological developments now allow the investigation of eye movements and physiological measures during real interactions with adequate recording quality. Considering the wealth of conflicting findings on gaze behavior in social anxiety, the current study aimed at unraveling the mechanisms contributing to differential gaze patterns in a naturalistic setting in the general population and in social anxiety. We introduced participants with differing social anxiety symptoms to a waiting room situation while recording heart rate and electrodermal activity using mobile sensors and eye movements using mobile eye-tracking glasses. We observed fewer fixations on the head of the confederate in the initial waiting phase of the experiment. These head fixations increased when the confederate was involved in a phone call and head fixations were most pronounced during the actual conversation. In opposition to gaze-avoidance models of social anxiety, we did not observe any correlations between social anxiety and visual attention. Social anxiety was, however, associated with elevated heart rate throughout the entire experiment suggesting that physiological hyperactivity constitutes a cardinal feature of the disorder.

Introduction

Understanding social contexts is essential for human interactions. At the core of interpreting the social dynamics of a scene lies an attentional bias towards the fellow humans taking part in it (Birmingham & Kingstone, 2009). Various laboratory studies have shown increased gaze towards human bodies and, in particular, faces, independent of the task at hand (Flechsenhar & Gamer, 2017) or the surrounding visual information (Birmingham et al., 2009b; Boggia & Ristic, 2015; End & Gamer, 2017; Rösler et al., 2017). However, recent advances in technology have enabled researchers to move the studies on social attention from laboratory settings to the real world. After all, the viewing of stimuli on a PC screen lacks many aspects a genuine social encounter entails. A major concern of the proponents of ecological validity (Kingstone, 2009; Kingstone et al., 2008; Risko et al., 2012; Smilek et al., 2006) is that most of the psychological stimulation material does not adequately approximate the complexity of daily life situations. While participants are known to preferentially fixate eyes of schematic or photographed faces presented in isolation (e.g. Friesen & Kingstone, 1998; Frischen et al., 2007; Langton et al., 2000), more intricate visual elements of real world scenes might draw attention away from the eyes. Indeed, numerous studies have shown substantially reduced overt attention to humans when participants are moved from the lab into environments with higher ecological validity (Foulsham et al., 2011; Hayward et al., 2017; Laidlaw et al., 2011). Participants look significantly less frequently at an actual person sitting in a waiting room with them than if that same person is not physically present but displayed on a computer screen. These gaze reductions possibly stem from an active avoidance of a social interaction with the confederate (Laidlaw et al., 2011). These results argue for further investigations of gaze behavior in everyday situations to conclusively elucidate the mechanisms that underlie these observed reductions.

While an attentional shift towards humans is vital for various higher socio-cognitive tasks, several psychiatric disorders display altered processing of human features. Social anxiety, which is characterized by a penetrating fear of social encounters and social-evaluative situations (Clark & Wells, 1995) and typically has a life-long impact on the relationships of the affected individuals (Wittchen et al., 1999), is assumed to feature differential processing of social information as a core symptom (Wells et al., 1995). There is, however, vast disagreement regarding the characteristics of these gaze behavior alterations (Chen & Clarke, 2017). Studies using emotional faces as probes to investigate attentional biases have shown that socially anxious patients predominantly avoid these threatening stimuli (Chen et al., 2002; Mansell et al., 1999). As heightened processing of emotional faces has also been reported in the social anxiety literature (Asmundson & Stein, 1994; Gilboa-Schechtman et al., 1999), Mogg and colleagues have proposed to focus on the temporal dynamics of gaze in social anxiety and suggested an initial hypervigilance to social or emotional stimuli followed by a subsequent avoidance (Mogg et al., 1997).

Up to date, only a handful of studies investigated gaze behavior in social anxiety during a real interaction and, again, provided conflicting results (Farabee et al., 1993; Gilbert, 2001; Howell et al., 2016; Langer et al., 2017; McManus et al., 2008). Considering that viewing preferences of social features change from lab to real-world environments (Foulsham, Walker, et al., 2011; Hayward et al., 2017; Laidlaw et al., 2011; Rubo et al., 2019) and that elevated skin conductance levels in social anxiety are primarily associated with real eye contact rather than the observation of photographed faces (Myllyneva et al., 2015), an investigation of gaze and physiological responses in social anxiety towards different degrees of social interaction in a real-world setting will help to clarify the mechanisms underlying anxious behavior. The aim of the present studies was hence to investigate whether the presence or absence of a potential social interaction influences gaze allocation and whether social anxiety modulates viewing behavior and physiological responses across these different social situations. We measured eye gaze, heart rate and electrodermal activity (EDA) of participants with different degrees of (sub-)clinical social anxiety using a mobile eye-tracking device and mobile physiological sensors while they were sitting in a waiting room with a male confederate who they believed to be another participant. Throughout the entire recording, participants were not aware that the experiment had already started. To modulate the degree to which a social interaction was possible, the confederate received a phone call after the initial two minutes of the experiment rendering him unavailable for a conversation. After two minutes on the phone, he addressed the participant directly and initiated a semi-stereotyped conversation. Based on the results of the waiting room study of Laidlaw and colleagues (2011), we expected that head fixations would be most reduced in the initial waiting phase when an interaction had to be actively avoided. Additionally, we expected that socially anxious participants would display reduced social gaze accompanied by increased physiological arousal throughout all conditions but most prominently during the interaction phase.

Methods

Participants

Prior to recruitment, we asked potential participants to fill in several questionnaires via an online recruiting platform hosted by the university. The pre-screening enabled us to include a wide spectrum of social anxiety manifestations in our sample. Of 409 people who participated in the online screening, we invited 98 people to join the study. A few participants had to be excluded because of various issues with data acquisition or quality, including too large gaze drifts evoked by moving glasses (*n* = 4), hardware recording issues on the day of testing (*n* = 5), too many missing valid fixations (> 30% of recorded frames, *n* = 7) and insufficient quality of the electrocardiogram (ECG) data (*n* = 2). Additionally, unforeseeable events during the recording led to further exclusions, to be specific the non-permitted use of mobile phones (*n* = 4), the initiating of a conversation with the confederate during the waiting phase of the experiment (*n* = 2), disbelief in the cover story of the experiment (*n* = 2) and acquaintance with the confederate (*n* = 1). Since we preselected participants according to their social anxiety and higher levels are more prevalent among females than males (McLean et al., 2011), the final analysis included 71 participants of which 61 were female (mean age = 24.50 years, *SD* = 6.37 years). The study was approved by the local ethics committee. Each participant provided written informed consent prior to the start of the experiment for the experiment they believed to take part in. Participants were informed about the purpose of the actual experiment and their rights to withdraw consent after recordings were complete. Participants were awarded monetary compensation for their participation.

As part of the experiment, participants completed the following questionnaires to allow for a comprehensive characterization of the sample regarding anxiety, autism and general personality traits (see Table 3): The German versions of the Social Interaction Anxiety Scale (SIAS; English original version: Mattick & Clarke, 1998; German version: Stangier et al., 1999), the Brief Fear of Negative Evaulation Scale – Revised (BFNE; English original version: Carleton et al., 2006; German version: Reichenberger et al., 2016), the short version of the Autism Spectrum Quotient (AQ; English original version: Baron-Cohen et al., 2001; German version: Freitag et al., 2007), State-Trait Anxiety Inventory (STAI; English original version: Spielberger, 1983; German version: Laux et al., 1981) and the short version of the Big Five Inventory (BFI; English original version: John et al., 1991; German version: Rammstedt & John, 2005). The range of SIAS scores that were observed in the current sample indicates that we were successful in recruiting participants ranging from very low social anxiety to levels that are also observed in clinical samples which is usually determined by a cut-off score of 30 (Stangier et al., 1999).

Table 3. Sample characteristics.

	Mean	SD	Minimum	Maximum
SIAS	25.96	15.66	6	71
BFNE	37.49	11.89	13	60
STAI	42.04	12.04	24	73
AQ	8.63	5.04	1	20
BFI Extraversion	13.59	4.13	5	20
BFI Agreeableness	11.24	2.93	4	18
BFI Conscientiousness	15.06	2.63	9	20
BFI Neuroticism	13.85	3.54	4	20
BFI Openness	20.42	3.23	9	25

Measurement devices

We used mobile SMI ETG 2W eye-tracking glasses to record eye movements of both eyes with a frame rate of 60 Hz (SensoMotoric Instruments GmbH, 2014). The front camera of the glasses had a resolution of 1280 x 980 pixels and recorded the field of view with a frame rate of 24 Hz. A Samsung Galaxy Note 4, which the participant was carrying in a waist bag, stored all recorded data.

To measure heart rate and heart rate variability, we attached a mobile EcgMove 3 Sensor (Movisens GmbH, 2017) to the sternum of participants by means of a chest belt. The ECG raw signal was sampled with a frequency of 1024 Hz. We additionally used a mobile EdaMove 3 Sensor (Movisens GmbH, 2017) attached to the thenar and hypothenar eminences of the participant's non-dominant hand to record electrodermal activity with a sampling rate of 32 Hz by a constant voltage system (0.5 V).

Apart from the paper-pencil sociodemographic form, questionnaires were filled in on a Dell Latitude 11" Pro 5170 Tablet using the software Presentation (Neurobehavioral Systems, 2016).

Figure 15. Experimental setup. A = Participant position, B = Confederate position, 1 = Posters used for three-point eye-tracker calibration and validation, 2 = Measurement devices, 3 = Confederate questionnaires, 4 = Experimental PC with putative experiment message.

Procedure

Participants expected to take part in an experiment which investigated pupil size and physiological responses in response to differing visual stimuli. Upon their arrival at the institute, the experimenter informed participants that the prior participant (confederate) experienced a slight delay and was therefore still filling in some questionnaires. To allegedly save time, ECG- and EDA-electrodes as well as mobile eye-tracking glasses were attached to the participant in the meantime. The eye-tracker was subsequently calibrated by using three points of reference on a poster (see Figure 15). Upon successful calibration, eye-tracking recording started. Simultaneously, successful calibration cued the confederate to remark that one of the questionnaires had not been copied properly. To minimize movements as all devices were already attached and calibrated, the participant was asked to take a seat facing the confederate (see Figure 16). Under the pretense of fetching a new questionnaire, the experimenter left the room, which marked the onset of the actual experiment. During the first two minutes, which will be referred to as the waiting phase throughout the article, confederate and participant did not interact, but the confederate was occupied filling in the remaining questionnaires. After two minutes, he received a phone call, allegedly by a friend but in truth by the experimenter, which lasted approximately another two minutes (phone phase). Upon hanging up the phone, the confederate initiated a two-minute interaction phase discussing a catalogue of statements and questions regarding previous experiences with psychological experiments which was standardized as much as a natural conversation allows. Although we attempted to generate experimental phases with a length of exactly 2 minutes, these durations varied slightly due to several reasons (e.g., time to establish and finish the phone call, responses of the participant in the interaction phase). On average, the waiting phase lasted 114.72 s (*SD* = 8.62 s), the phone phase 119.58 s (*SD* = 4.64 s) and the interaction phase 115.27 s (*SD* = 7.88 s) respectively.

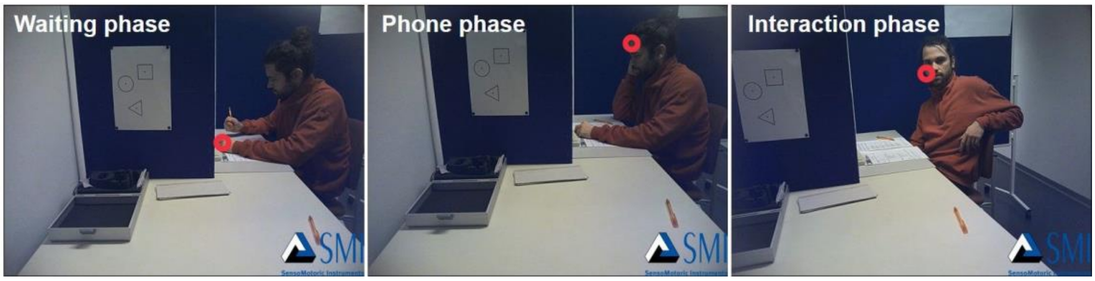

Figure 16. Experimental procedure. *Exemplary images of each of the three experiment phases as viewed by the participant.*

Throughout all three phases, we measured ECG, EDA and gaze by the previously attached mobile devices. After completion of the interaction phase, the three-point poster was used to re-assess the initial calibration and enable a later drift correction. The participants were then informed about the actual purpose of the study and asked to fill in a sociodemographic and the other trait questionnaires.

Data processing

Eye-tracking and video data were handled within the BeGaze software (Version 3.6; SensoMotoric Instruments GmbH, 2016) and, as an initial step, field-of-view videos were exported as full-length AVI-files excluding any fixation data. Instead, fixation coordinates for each frame were exported separately as CSV-files. The AVI-files were then converted into single pictures per video frame using MATLAB (The MathWorks Inc, 2018). As many participants displayed a slight gaze drift tested by the final three-point validation at the end of the experiment, this drift was corrected per participant. For this purpose, the distance between the three actual fixations and points to be fixated was measured manually using the GIMP compass tool (Version 2.8; The GIMP Documentation Team, 2015). The mean of the deviations was then used for a linear gaze drift correction of the extracted fixation coordinates. If the mean deviation was larger than 200 pixels, equivalent to half of the confederate's head size, the participant was excluded entirely from the analyses. Using MATLAB, the drift-corrected coordinates were projected onto the video frames creating a ring shape with an inner radius of 9 pixels and an outer circle diameter of 25 pixels around the fixation (see Figure 16 for an example image). We evaluated fixations manually in an application created specifically for this coding purpose at 6 Hz (i.e., every fifth frame to reduce the time spent on the still lengthy manual procedure). One independent coder who was not involved with the design or implementation of the study rated which region of interest (ROI; confederate head, body or surroundings) was fixated or if no valid fixation was present in the respective frame. A subset of participants (n = 6) was rated by a second rater to establish that sufficient inter-rater reliability was present (Cohen's κ = .93). Based on these ratings, fixation proportion per ROI for all valid fixations within a phase were calculated for each participant.

EDA and ECG raw data were initially read and exported to CSV-files via the unisens4matlab toolbox (Movisens GmbH, 2018). During this step, ECG data were downsampled to 512 Hz. Heartbeat detection was then performed using specifically developed scripts within the open-source statistical programming language R (R Core Team, 2018). First, ECG data were filtered using a 2 Hz high-pass filter in order to remove slow signal drifts. Subsequently, R-waves were detected from the ECG recordings using a semi-automatic method and R-R-intervals were converted to HR (in beats per minute, bpm ; for a similar approach see Rubo & Gamer, 2018). Heart rates below 50 or above 130 bpm were visually inspected for plausibility and corrected if necessary (i.e., in cases of undetected or erroneously detected R-waves). Subsequently, the data were transferred to a real time scale (Velden & Wolk, 1987) and the average heart rate (HR) per experimental phase was calculated within each participant.

Additionally, heart rate variability (HRV) in the high frequency (HF) band was calculated per phase using the R package RHRV (Rodríguez-Liñares et al., 2011) analogous to the procedure of Pittig and colleagues (Pittig et al., 2013). In detail, Fourier analysis (window length = 60 s, window displacement = 5 s) was used to determine the spectral power density of HRV in the high frequency range of 0.15–0.40 Hz. These values were expressed in normative units by dividing them by the total absolute power across all frequencies between 0.003 and 0.40 Hz and multiplying the resulting fraction by 100. Since HRV is affected by the length of the recording period (Berntson et al., 1997), we calculated HF-HRV only for the first 90 s of each experimental phase.

From the EDA recordings, we calculated the skin conductance level (SCL) for each experimental phase as the average of the low-pass filtered (1 Hz cutoff-frequency) skin conductance signal for each participant. These values are expressed in µS.

Statistical Analyses

Data were analyzed using R (R Core Team, 2018). An a priori significance level of $\alpha = .05$ was specified for all statistical tests. As our key question was to discern whether social anxiety impacts either gaze or physiological parameters, we computed four separate analyses of covariance (ANCOVAs) using the R package afex (Singmann et al., 2015). For heart rate, heart rate variability and skin conductance levels, we included the mean-centered covariate SIAS score and the three-level factor experimental phase as well as their interaction term in the model. To model fixation proportions, we again included the mean-centered SIAS score and experimental phase as predictors but added the two-level factor ROI (head versus body), all possible two-way interactions and the triple interaction term to the model (for analyses of eye versus mouth region fixations see Supplemental Material and Supplemental Figures S1 and S2). We additionally included gender as a covariate in all models to correct for potential effects of gender on our dependent variables. Post-hoc tests were performed using the R package emmeans. Generalized η^2 values (Bakeman, 2005) are reported as estimates of the effect size for linear model fixed effects. Correlation coefficients between questionnaire data and the dependent variables are reported in Supplemental Tables S2 to S4, while correlations between our dependent variables per phase are reported in Supplemental Figure S4.

As an additional analysis, we computed split-half consistencies of head fixations per phase across participants since recent studies showed stable fixation patterns within individuals across trials and we were interested to see if we can replicate this finding in our real-life dataset (De Haas et al., 2019; Guy et al., 2019). To this end, we split the three experimental phases per participant into two equally long periods and computed head fixations within each half. We then calculated Pearson's correlation coefficients between halves across all participants for each phase.

Results

Gaze data

To investigate how fixation proportions on the confederate are impacted by social anxiety, experiment phase and ROI we performed an ANCOVA with fixation proportions as the dependent variable and ROI (head versus body) and experimental phase (waiting, phone and interaction phase) as factorial predictors and SIAS score and gender as covariates. We found a significant main effect of ROI ($F_{(1,68)} = 19.01$, $p < .001$, $\eta^2 = .07$) as there were overall considerable differences between body and head fixations (see Figure 3 and Supplemental Table S5). A significant main effect of experiment phase ($F_{(2,136)}$ = 101.54, ε = 0.95, p < .001, η^2 = .21) is driven by higher fixation densities on the confederate during the interaction phase (see Supplemental Table S5). Importantly, we observed a significant interaction of experiment phase and ROI ($F_{(2,136)}$ = 101.94, ε = 0.88, p < .001, η^2 = .30) which mainly describes an increase of head fixations throughout the experiment (see Figure 17 and Supplemental Table S5). Regarding potential influences of social anxiety on fixation proportion, we did not observe a significant main effect ($F_{(1,68)}$ = 1.03, p =.313, η^2 = .004) and none of the individual interactions with experimental phase and ROI, nor the triple interaction between all three predictors reached statistical significance (all p > .49). To test whether no effect of social anxiety on gaze behavior was more likely than social anxiety having an effect on gaze data, we compared the Bayesian ANCOVA models with and without SIAS as a covariate using the BayesFactor package (Morrey & Rouder, 2018). Using the default prior, the resulting Bayes Factor in favor of the model without the SIAS predictor amounted to 4.26 ± 0.79% suggesting that a model which does not take social anxiety into account is approximately 4 times more likely to be the true data-generating model.

Examinations of split-half consistencies in our sample revealed strong and highly significant correlations in all phases (waiting phase: r = 0.56, p < .001; phone phase: r = 0.80, p < .001, interaction phase: r = 0.71, p < .001). Thus, viewing patterns were individually stable across participants.

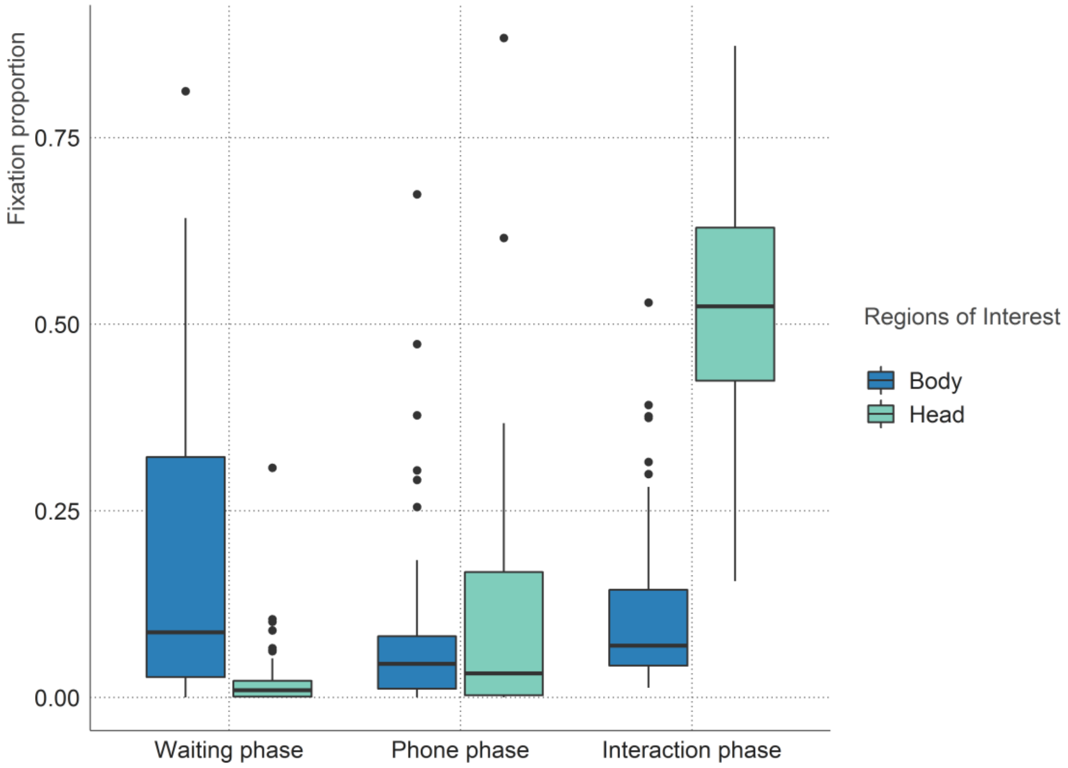

*Figure 17. Body and head fixation proportions during waiting, phone and interaction phase. Outliers are denoted by black dots and defined as points further than 1.5 * interquartile range of the lower or upper hinge.*

Physiological data

To investigate whether mean HR differed across phases and was impacted by social anxiety, we calculated an ANCOVA with the three-level factor experimental phase (waiting, phone and interaction phase) and the continuous SIAS score and the factorial variable gender as covariates. A main effect of phase confirmed that heart rate differed between experimental phases ($F_{(2,136)} = 22.12$, $\varepsilon = 0.78$, $p < .001$, $\eta^2 = .02$, see Figure 18 and Supplemental Table S6) and a main effect of SIAS revealed that social anxiety was also associated with heart rate levels ($F_{(1,68)} = 6.54$, $p = .01$, $\eta^2 = .08$). The interaction term did not reach statistical significance ($F_{(2,136)} = 0.08$, $\varepsilon = 0.78$, $p = .87$, $\eta^2 < .001$), indicating that there was a stable influence of social anxiety on heart rate independent of the phase. The main effect of gender ($F_{(1,68)} = 0.45$, $p = .51$, $\eta^2 = .006$) and the interaction term with experimental phase ($F_{(2,136)} = 1.22$, $\varepsilon = 0.78$, $p = .29$, $\eta^2 = .001$) did not reach statistical significance. To further assess the relationship between heart rate and social anxiety, we calculated Pearson's correlation coefficients comparing the association between SIAS scores and mean heart rate for each phase individually. Indeed, SIAS scores were significantly correlated with mean heart rate across all phases (waiting phase: $r = 0.30$, $p = .012$; phone phase: $r = 0.27$, $p = .024$; interaction phase: $r = 0.30$, $p = .012$, see Figure 19).

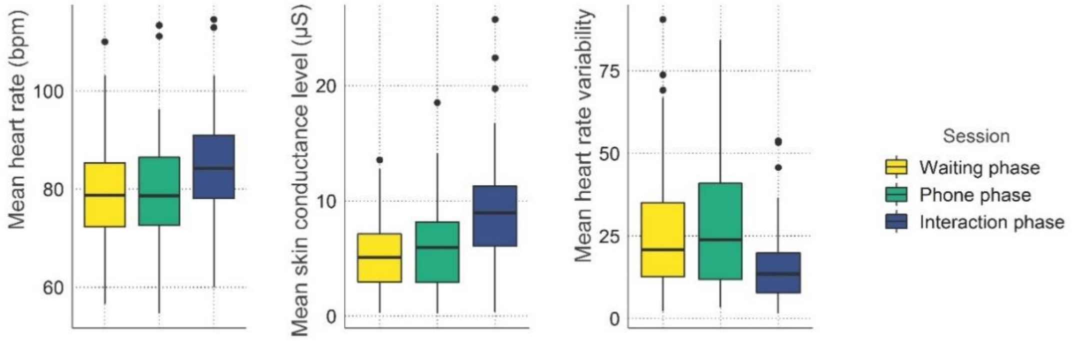

*Figure 18. Mean heart rate, skin conductance levels and heart rate variability across the waiting, phone and interaction phase. Heart rate variability is represented in normalized units. Outliers are denoted by black dots and defined as points further than 1.5 * interquartile range of the lower or upper hinge.*

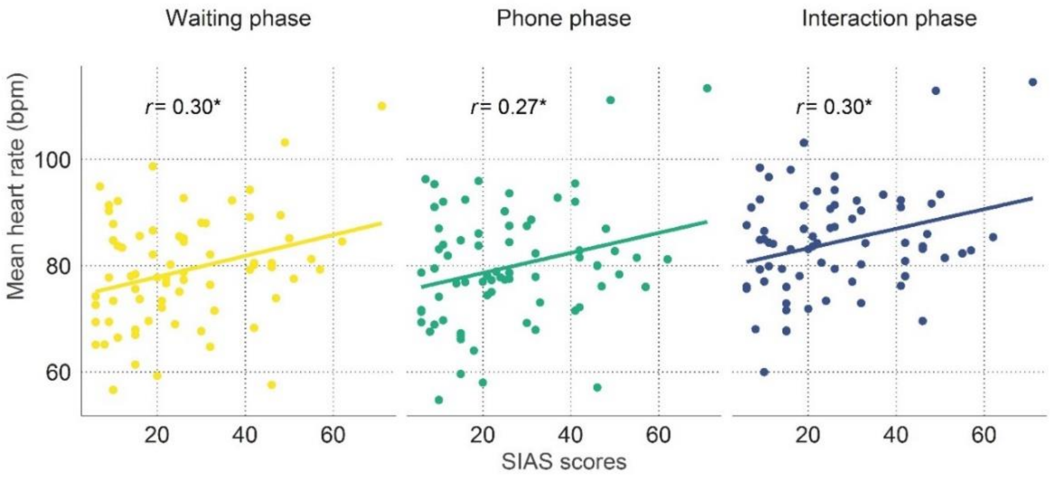

Figure 19. Mean heart rate in beats per minute as a function of social interaction anxiety scale (SIAS) scores per phase. Asterisks denote p < .05.

An analogous ANCOVA model investigating influences of social anxiety and experimental phase on HF-HRV revealed that only the main effect of experimental phase was statistically significant ($F_{(2,134)} = 7.38$, $\varepsilon = 0.96$, $p = .001$, $\eta^2 = .04$), while the main effect of SIAS ($F_{(1,67)} = 0.03$, $p = .86$, $\eta^2 < .001$) and the interaction term ($F_{(2,134)} = 0.19$, $\varepsilon = 0.96$, $p = .82$, $\eta^2 = .001$) did not reach statistical significance (for post-hoc comparisons see Supplemental Table S7). Moreover, there was a significant main effect of gender on HF-HRV ($F_{(1,67)} = 4.80$, $p = .03$, $\eta^2 = .04$) but not of its interaction with experimental phase ($F_{(2,134)} = 0.38$, $\varepsilon = 0.96$, $p = .67$, $\eta^2 = .002$).

Similar results were obtained regarding the influences of social anxiety and experimental phase on mean SCL. Skin conductance levels varied across phase as revealed by a main effect of experimental phase ($F_{(2,136)} = 37.76$, $\varepsilon = 0.72$, $p < .001$, $\eta^2 = .060$) but we did neither observe a significant effect of SIAS ($F_{(1,68)} = 0.10$, $p = .75$, $\eta^2 = .001$), nor for the interaction term ($F_{(2,136)} = 0.96$, $\varepsilon = 0.72$, $p = .36$, $\eta^2 = .001$). There was no statistically significant effect of gender ($F_{(1,68)} = 0.45$, $p = .51$, $\eta^2 = .006$), nor of its interaction with experimental phase ($F_{(2,136)} = 3.41$, $\varepsilon = 0.72$, $p = .05$, $\eta^2 = .005$) on SCL (for post-hoc comparisons see Supplemental Table S8). For this reason, we did not conduct any follow-up correlational analyses between social anxiety and HF-HRV or SCL, respectively.

Discussion

Most of our current knowledge on the behavioral and physiological manifestation of social anxiety is based on the processing of simple social stimuli in a laboratory setting. The aim of the present study was to investigate autonomic responses and gaze behavior in (sub-)clinically socially anxious individuals in a naturalistic scenario while manipulating the degree to which a social interaction is likely to develop. Participants were not informed about the start of the recordings but waited to start a different experiment while another participant, the confederate, filled in questionnaires. We generally found that, independent of social anxiety, head fixations are most prominently reduced when a social interaction is possible but likely undesired as in this initial waiting phase. By subsequently rendering the confederate unavailable to talk through responding to a phone call, fixations on his head were seen to significantly increase. The likelihood of the confederate looking back or initiating a conversation hence seems to strongly affect gaze behavior. Overall, the final interaction, in which the confederate openly sought a conversation with the participant, was most demanding as indicated by elevated heart rate and skin conductance levels and a reduced heart rate variability. Surprisingly, social anxiety levels did not impact gaze behavior in any of the phases of the experiment. With regard to physiological measurements, however, heart rate consistently correlated with social anxiety scores independent of the phase. These findings suggest that individuals with (sub-)clinical social anxiety might be able to behaviorally compensate their fear of direct eye contact at the expense of increased physiological arousal in the presence of or the interaction with other individuals.

These findings represent an important addition to observations made in simplistic laboratory scenarios.

By immersing the participants into a naturalistic setting without their awareness of experiment and recording commencement, we were able to gain essential insights into the development of gaze throughout a real social interaction. The presence of another person had previously been shown to induce norm-following and facilitate differential task processing (Guerin, 1986; Zajonc, 1965). As hypothesized by Laidlaw and colleagues (Laidlaw et al., 2011) who also investigated eye movements during a waiting room situation, our results confirm that the chance of interacting with another unknown person clearly influences gaze behavior. We observed increased fixations on the head of the confederate when the possibility of an interaction decreased as the confederate was occupied with a phone call. We thereby provide further support for the notion of *civil inattention* initially formulated by Goffman which suggests that gaze avoidance is a characteristic feature of public anonymous settings (Goffman, 2008). While cardiovascular responses did not differ between the waiting and the phone phase, heart rate was significantly more elevated and heart rate variability significantly decreased during the conversation with the confederate. Skin conductance levels already increased slightly when the confederate started talking on the phone, possibly because of heightened attention to the sudden auditory stimulation. Again, electrodermal activity was highest during the interaction emphasizing that conversations with strangers seem to come at cost of certain physiological arousal. It is arguably easiest to avoid conversations in a waiting room scenario with strangers by reducing eye contact (Lalljee, 1978).

Considering the large body of conflicting literature on gaze behavior in social anxiety, the current results aid the interpretation of findings from simulated social exchanges and their potential translation to live social interactions. In line with an abundance of studies drawing on gaze recordings during photographed or video-animated faces (Beidel et al., 1985; Boll et al., 2016; Glasgow & Arkowitz, 1975; Hofmann et al., 1997; Walters & Hope, 1998; Weeks et al., 2011), we did not observe any evidence for gaze avoidance in social anxiety. To be more specific, we did not find reduced fixations on the head or body of the confederate in any of the experimental phases. How can this finding be reconciled with the reported observations (Farabee et al., 1993; Howell et al., 2016; Langer et al., 2017; Moukheiber et al., 2010) of gaze avoidance in social anxiety? Social settings are diverse and different social encounters might facilitate distinct behavioral reactions. Possibly, the presence of a real person induces norm-activating behavior (Guerin, 1986; Zajonc, 1965) which, specifically during conversations, requires a certain amount of attention towards the interaction partner. The use of images or videos as a simulation of an interaction might therefore fail to enforce this heightened attention on the interaction partner in socially anxious individuals. Furthermore, the degree to which the social situation is perceived as threatening varies widely across reported experiments. We deliberately chose to investigate a neutral everyday situation to be able to draw inferences about non-threatening daily encounters while observed reductions in gaze in live interactions were yet often in response to conflicts or a disagreeing confederate (Farabee et al., 1993; Langer et al., 2017).

The elevated heart rate observed in socially anxious participants in our study also suggests that overcoming fear of eye contact might ensue at the cost of higher physiological arousal. As this association has also been observed in another study with regard to electrodermal activity (Myllyneva et al., 2015), it is plausible that social anxiety is primarily associated with either compensatory or concomitant physiological mechanisms in real interactions. One might wonder why we exclusively observed correlations between heart rate and social anxiety and not any of the other physiological measures. However, the literature on physiological responses in social anxiety has been inconsistent and it is plausible that this inconsistency and our failure to observe altered electrodermal activity and heart rate variability in social anxiety stems from too low statistical power and the investigation of subclinical samples.

While the use of an everyday interaction allows us to investigate which alleged features of social anxiety survive outside of a standard laboratory setting, several limitations need to be considered. We aimed at maximizing the naturalness of our experimental design, which is why we decided against very standardized conversations between confederate and participant, a specific sequence of direct gaze or gaze avoidance of the confederate or a forced conflict situation. Consequently, we are unable to draw inferences about specific reactions to evaluative gaze or interactions with different emotional valence. Additionally, a potential argument against our interpretation of the observed increase in gaze towards the confederate's head is that it was also the sole source of auditory information in this moment. The auditory stimulation rather than the change in interaction possibility could have impacted the participant's increased attention. Future studies should therefore consider including a non-social auditory source to allow comparisons in attentional capture. While physiological reactions were most prominently increased in the conversation phase suggesting that this part of the experiment was also the most arousing to participants, an increase of movement during speech could also contribute to these observations. As social anxiety was yet correlated with increases in heart rate throughout the entire experiment, the association between physiological responses and anxiety symptoms cannot be fully explained by movement artefacts.

This finding nevertheless needs to be taken with a grain of salt since we are unable to conclude whether the increased heart rate is caused by the impending or current social interaction or by a higher baseline heart rate in socially anxious participants. The participation in an experiment inevitably induces brief social contact with the experimenter, even if no fellow participant is in the room, and it is therefore very difficult to isolate the causes of the observed heart rate increase. Future research should nevertheless aim to elucidate the circumstances under which elevated heart rates can be observed in socially anxious participants. Such examination might include methods of ambulatory assessment to examine heart rate changes across a variety of everyday situations in socially anxious participants (Sperry et al., 2018). Additionally, socially anxious participants in our sample were not clinically diagnosed and altered gaze patterns might only become apparent in a clinical sample. However, 26 out of the 71 participants demonstrated high SIAS scores (30 or higher) that also occur in clinical samples (Stangier et al., 1999). Since we observed virtually no correlations between social anxiety symptoms and measures of gaze (see Supplementary Table S2 and Supplementary Figure S3), the current study does not provide evidence for the assumption that results would look different in a clinical sample. Whether our failure to observe gaze differences in social anxiety is due to homogenous viewing behavior in anxious and non-anxious populations or related to differential responses to social threat remains to be elucidated by future studies. Interestingly, non-clinical socially anxious participants responded with increased gaze to social cues when social evaluation was enhanced (Müller-Pinzler et al., 2015), suggesting that differences in gaze between socially anxious and non-anxious groups might only become apparent when the perceived social evaluation is sufficiently high. Controlling for state anxiety and the level of perceived social threat might help to interpret gaze patterns in the future.

The current results present evidence for increased heart rate but not for differences in gaze behavior in social anxiety during a real social interaction. Considering the inconsistent literature on social attention in socially anxious individuals, the current null findings regarding this measure suggest that gaze reduction might be either primarily a laboratory phenomenon or restricted to specific situations (e.g., including evaluative threat) or very high (clinically relevant) levels of social anxiety. The reported fear of eye contact might therefore be a more cognitive rather than a behavioral feature of the disorder. By contrast, increases in cardiovascular responses in social situations seem to represent a cardinal feature of social anxiety and current and future therapeutic interventions could thus consider targeting the awareness and regulation of cardiovascular activity.

Acknowledgements

Declarations of interest: none

This work was supported by the European Research Council (ERC-2013-StG-336305) and the German Research Foundation (Collaborative Research Center "Fear, Anxiety, Anxiety Disorders", SFB-TRR 58 project C10).

DISCUSSION

Humans in our surroundings are of special importance to us. Eye-tracking studies have demonstrated the prioritization of humans in complex scenes reflected in increased fixations on human features. In the present thesis, we aimed at elucidating the specific mechanisms by which this social attention operates. On a behavioral level, we were able to show that social features in complex naturalistic scenes are prioritized in an automatic fashion. After 200 ms of stimulus presentation, which is too early for top-down processes to occur, participants targeted those image areas that contained humans in them significantly more often than would be expected from a chance distribution of saccades. Additionally, saccades towards social areas occurred earlier in time than saccades towards non-social image regions. The application of a gaze-contingent paradigm allowed us to demonstrate that social features are preferentially targeted even when bottom-up information is restricted; that is, participants scanned the masked images, in which only the fixated region was clearly visible, in a way that was comparable to free-viewing of these images (i.e. social regions were fixated drastically longer than non-social image regions). A comparison of real and artificial faces during the observation of dynamic naturalistic videos further revealed that real faces, belonging to actual humans in the scene, were preferentially viewed. Artificial faces, belonging to human-like statues or machines, also significantly predicted gaze allocation but to a lesser extent than real faces. An fMRI study investigating the neural correlates of reflexive social attention, pointed to an involvement of early and extrastriate visual cortices in the encoding of social feature space.

Overall, these studies helped to clarify under which circumstances social features are prioritized in a laboratory setting. However, to be able to draw inferences about real life situations, we also conducted a study in which participants interacted with a confederate which they believed to be a fellow participant while waiting to begin an experiment. This study provided crucial insights into how real-life behavior differs from viewing behavior in the lab. Overall, fixations on the confederate were most reduced when an interaction was being avoided, increased when the interaction was rendered temporarily impossible due to a telephone conversation of the confederate and were, unsurprisingly, most frequent when engaging in a conversation with the confederate. These results highlight that gaze behavior heavily depends on the social setting. As we recruited the sample by inviting participants with a range of social anxiety symptoms, we were also able to assess whether social anxiety impacts gaze behavior during this real-life interaction. We did not find any evidence for differences in measures of visual exploration but detected a correlation between heart rate and social anxiety across all phases of the experiment. Alleged altered gaze behavior in social anxiety might thus represent a laboratory phenomenon that occurs only in very specific real-life conditions (Rubo et al., 2019). Altogether the experiments described in the current thesis thus helped to refine our understanding of social attention and raise awareness of the challenges the real world imposes on the inferences we draw from laboratory research.

Translating Findings on Social Attention into the Real Life

Considering the discrepancy between our null finding on gaze alterations in social anxiety (see thesis chapter 6) and the laboratory literature on the topic (e.g. Lazarov et al., 2016; Horley et al. 2004; Mansell et al. 1999), it is necessary to reflect on the extent to which findings from the lab translate to the actual world. A great advantage of laboratory experiments is the high experimental control implicated in the design. Accordingly, the presented stimuli can be chosen so that they fulfill certain *a priori* defined criteria, commonly a matching of lower-level visual features between test and control stimuli in the field of attention research. Additionally, the experimenter has high control over the duration of the stimulus presentation and the order in which these stimuli are shown. In the real world, however, it is extremely challenging to warrant even minimal control over the experimental design. While the team of researchers can ascertain that all participants are tested in the same geographic environment, the exact time during which certain aspects of the environment are perceived is contingent on the participants viewing behavior. The investigation of interactive behavior yields even more challenges as even a pseudo-scripted dialogue will have to be adapted to the input of the participant to permit plausibility. Nevertheless, recent investigations into the translatability of laboratory findings emphasize the need to conduct increased research outside of the confines of the highly controlled lab scenarios (Foulsham et al., 2011; Gallup et al., 2012; Hayward et al., 2017; Laidlaw et al., 2011; Rubo et al., 2019). Foulsham and colleagues pointed to subtle but possibly crucial differences in viewing behavior in the real moving world. When participants walked across the campus, they tended to fixate less on another pedestrian located closely to them than if they viewed a close pedestrian on a video tape. In line with this finding, another study demonstrated that people were not inclined to look at another person situated in the same waiting room as them but, instead, spent more time fixating the same chair when it was empty. However, when a video of the person seated on the chair was presented on a computer screen, participants exhibited increased fixations on the person as compared to the real life condition (Laidlaw et al., 2011).

These results are not surprising. Imagine yourself seated in the waiting room of your local dentist. Will you shamelessly scan the eyes of the other people in the room as if their faces were presented to you on a computer screen? Probably not. As formulated many decades ago by proponents of the dual function of gaze (Benthall et al., 1976), social attention is interactive. The eyes not only serve to perceive visual input in our surroundings but also communicate our inner processes to the outer world. This denotes the striking difference between the observation of stimuli inside versus outside the lab – the person seated next to me at my dentist might actually look back and wonder what all my staring is about.

For this reason, many researchers have argued that the static or dynamic stimuli commonly employed in laboratory science do not trigger the same behavior in participants as a real-world encounter with humans (Foulsham & Kingstone, 2017; Foulsham & Underwood, 2008; Risko et al., 2016; Tatler et al., 2011). To systematically investigate the effect of "believing to be seen" on viewing behavior, Gobel and colleagues asked participants to watch videos of faces with different social ranks, while being filmed themselves (Gobel et al., 2015). Participants were either instructed that their recordings would subsequently be seen by the people in the videos or that they would not be shown to anyone. The expectancy of being observed later on was associated with decreased fixations on the eyes of those people with alleged higher social ranks. Similarly, participants spent significantly more time visually exploring an alluring swimsuit calendar when they believed that their eye movements were not monitored (Risko & Kingstone, 2011). This effect, which can be interpreted as the eye-tracker introducing a form of social presence, was replicated in another sample which also showed that fixations on the provocative stimulus increased when participants forgot about wearing eye-trackers (Nasiopoulos et al., 2015). As already suggested in the previously reported real-life studies (Foulsham et al., 2011; Laidlaw et al., 2011), the communicative function of one's own gaze thus seems to significantly influence gaze behavior.

The Relationship Between Prosociality and the Belief of Being Watched

Altered gaze as a consequence of the presence of others, the so-called "audience effect" as first introduced by Zajonc in the 1960s (Zajonc, 1965), is assumed to be largely driven by reputation management (Cañigueral & Hamilton, 2019; Tennie et al., 2010). Creating a good standing in one's environment requires the ability to infer the information one makes available to surrounding people and the ability to assess how current behavior will be received (Izuma, 2012). One means to achieving a good reputation is behaving prosocially, as indicated by increased benevolent and generous behavior (Bereczkei et al., 2007; Smith & Bird, 2000). The presence of an audience was indeed associated with larger contribution to public instead of private goods in a study conducted by Filiz-Ozbay and Ozbay (Filiz-Ozbay & Ozbay, 2014). Interestingly, prosociality also seems to be intricately linked with eye gaze; eye contact accordingly triggers increased prosocial behavior (Bull & Gibson-Robinson, 1981; Goldman & Fordyce, 1983) and the presence of attentive eyes was associated with a greater tendency to offer help (Manesi et al., 2016) and the number of donations in a field experiment (Powell et al., 2012). To gain insight into the mechanisms by which the belief of being watched influences gaze patterns, Canigueral and Hamilton tested how prosocial decisions differ between live and pre-recorded interactions (Cañigueral & Hamilton, 2019). Participants believed to take part in a study assessing social attention during charitable behavior. During the experiment, they interacted with a confederate, an alleged charity worker, in a web-based conference call. While all calls were actually pre-recorded, only half of the participants received this information and the other half believed that the confederate was online. All participants had to perform two tasks during the web call. In one task, the confederate asked participants how they would behave in everyday dilemmas (e.g. would you give money to a homeless person) and, in the second task, resembling the dictator game used in previous research (Izuma et al., 2011), the confederate asked participants whether they would accept a monetary offer of which a certain amount of money would be allocated to charity. A decline of the offer would always result in a bonus of 4 pounds for the participant.

The results revealed that the belief of being watched marginally increases prosocial behavior, as reflected in both increased expression of prosocial tendencies and monetary donations when participants believed the interaction partner was actually online. Additionally, participants showed reduced gaze behavior towards the interaction partner in the live condition, which has been reported in other studies and is in line with social norm-following (Foulsham et al., 2011; Laidlaw et al., 2011). Illuminating the relationship between gaze and prosocial behavior, Canigueral and Hamilton also report increased gaze towards the confederate after non-prosocial decisions versus prosocial decision. This altered gaze following socially non-desired behavior is conformant with the idea of a link between the audience effect and reputation management, as participants possibly increase their gaze on the interaction partner to seek information as to how they are currently being evaluated. In line with the idea of fear of social evaluation, contagious yawning, a generally undesired signal of boredom or disinterest, was reduced when the experimenter remained in the experimentation room, both when watching a yawn contagion video on a screen (Gallup et al., 2016) and when viewing yawning agents in a virtual reality environment (Gallup et al., 2019). These results thus argue for caution when interpreting findings on social attention from laboratory settings as many behavioral measures of social attention are influenced by social presence.

Differences in Visual Input Between Real-Life and Laboratory Conditions

Next to the effects introduced by the presence of other people in real life scenarios, challenging differences in visual features of the presented stimuli also need to be considered. During a real world encounter, humans typically align their head and body with the areas or stimuli they are most interested in (Foulsham & Kingstone, 2017). When viewing a series of pictures on a PC screen, this alignment is reduced from our three-dimensional environment to a two-dimensional space. Tatler and colleagues argue that the frequent use of smaller stimulation windows led to an overemphasis of low-level image features in the attention literature (Tatler et al., 2011). Next to grave changes in the studied visual space, additional biases introduced by laboratory stimulations include the sudden appearance and disappearance of stimulus material and the inclination to fixate the center of the screen (Foulsham & Underwood, 2008). The continuity and context of the visual stimulation also frequently undergo a drastic change when transitioning from laboratory to real world studies. However, the temporal predictability of the real world is likely to influence how we scan our environment. Foulsham and colleagues presented static images derived from a mobile real-world eye-tracking study to participants in the lab to assess whether the presentation sequence of these static images affected how well gaze behavior within the lab could predict gaze behavior outside of the lab (Foulsham & Kingstone, 2017). The selected images were derived from 10 second intervals of real-world campus walks and were either shown in a random order or presented sequentially, thereby reestablishing the continuity of the real-world movie. Overall, viewing behavior was more consistent across participants for the images presented in continuous fashion than for randomly presented images. However, none of these fixation patterns proved to be a useful predictor of gaze behavior in the real world. The best prediction was instead provided by a model considering where the eyes are positioned in the presented videos. In other words, it was more helpful to simply read out the center of the recorded videos and correct it by the position of the eyes during the recordings rather than collect actual fixations data on static images. It thus becomes evident how important temporal continuity and spatial context are for our understanding of attentional allocation in scene perception. Taken together, the dual function of gaze which comes into play when another person is around, the influence of social presence on prosocial behavior and the differences in the characteristics of visual stimulation material between laboratory and real-world environments emphasize the volatility of the meaning of social stimuli. A person that I view on a video in the lab is of different relevance to me than the experimenter standing by my side and this difference in meaning demonstrably affects my attentional allocation. Putative mechanisms of social attention are thereby highly susceptible to the environment under which they are studied.

The Complexity of Psychiatric Disorders

When we investigate attentional biases in psychiatric disorders, we do not only need to consider the ecological validity of the research design but also the manner by which we categorize and recruit the participants that make up the sample of interest. Since the introduction of disorder classification systems such as the DSM and ICD in the mid-20th century, mental disorders have been commonly understood and classified based on the description and observation of specific symptoms (Clark et al., 2017). These diagnoses typically define or guide the prescribed treatment, following the general assumption that the observed symptoms of a specific disorder can be explained by an underlying neurobiological cause. While the field of neuroscience has greatly advanced in the past decades, with refined technological tools enabling a supposedly better understanding of the mechanisms at work, little progress has been made towards improving the understanding and outcome of mental disorders (Kendler, 2012; Kendler et al., 2011). Effective treatment, in the form of pharmaceutical interventions or behavioral treatments, does exist but these treatments are often not very precise. For instance, antipsychotic medication is commonly administered to both patients with schizophrenia and bipolar disorders (Correll et al., 2015), whereas antidepressants are used across a wide spectrum of affective disorders (Bandelow et al., 2012). Crucially, despite the wealth of research into the origins of mental illnesses, the burden of suffering does not seem to have significantly improved for any psychiatric disorder as reflected in unchanged mortality rates (Kessler et al., 2005).

Considering these dismaying outcome reports and the overall vague boundaries between different disorders, the National Institute of Mental Health suggested a different research-based framework for the classification of mental disorders, the Research Domain Criteria (RDoC), (Cuthbert & Insel, 2013; Insel et al., 2010). RDoC adopts a translational perspective on mental disorders, where disruptions of established healthy working mechanisms of the brain present the starting point of classification rather than going from symptoms to psychopathology. Within the RDoC, five major domains of mental functioning (positive valence systems, negative valence systems, cognitive systems, social processes and arousal systems) are outlined and alterations in disorders are studied on levels ranging from genes and cells to complex behavior and self-report (Cuthbert & Insel, 2013). A decade after its initial formulation, it remains to be seen to what extent focusing on dimensions of impaired mental health can help to bridge insights gained from genetics and neuroscience and can further our understanding of the causes of disorders (Patrick & Hajcak, 2016).

A different strand in the literature on psychiatry nosology proclaims a necessary shift from understanding mental disorders as clusters of symptoms with an underlying latent cause to regarding disorders as complex dynamic systems of directly interacting variables (Borsboom, 2017; Borsboom & Cramer, 2013). Rather than having one clear-cut neurological origin, mental disorders are thought to arise because of the causal connections between its symptoms. Taking depression as an example, the disorder itself can be understood as feelings of sadness leading to a loss of motivation. Decreased motivation, in turn, might result in increased feelings of guilt as a response to decreased productivity. It is easy to imagine how these feelings of guilt might again act as a catalyst for experiencing unhappiness. These symptoms might naturally be associated with modulations in neurotransmitter or hormone levels but variations in neurobiology alone are not constitutive of the disorder (Borsboom, 2017). Network theorists therefore suggest the use of network analyses, in which symptoms functions as nodes and correlations between symptoms are represented as edges between nodes, to elucidate the mechanisms of mental disorders (Borsboom & Cramer, 2013; Bringmann et al., 2013; Cramer et al., 2010). As these networks are able to incorporate longitudinal datasets, it is possible to estimate whether one node can predict another node at a subsequent time point. This type of analysis allows important insights into the transitions from a non-pathological to a pathological state. It has additionally been argued that the traditional approach of using sum scores to investigate the underlying biology of mental disorders overlooks crucial differences in the symptomatology between individual participants (Fried & Nesse, 2015). The consideration of different symptoms or items of a disorder in a network overcomes the oversimplification of a disease state.

This network approach presents an interesting avenue for social attention researchers, possibly allowing a more adequate understanding of the behavioral changes observed in psychiatric disorders with affected social attention. With regard to the discrepant literature on altered attention in social anxiety (as discussed in chapter 6 of this thesis), a network analysis identified a general attentional orienting bias to non-emotional stimuli, as well as fear and avoidance of social situations as the most central components of a social anxiety disorder network (Heeren & McNally, 2016). The often-discussed attentional engagement with threat or difficulty disengaging from it did not have high predictive value within the network. Within the field of social attention, there has been an increasing interest in the correlation between specific items of psychopathology or personality questionnaires and gaze behavior (Rauthmann et al., 2012; Wu et al., 2014). As a next step, these analyses on questionnaire items and behavioral measures of social attention could be subjected to a network analysis. While this approach requires large sample sizes for sufficient statistical power, it might help to elucidate which attentional deficits are, if at all, causally linked to which specific symptomatology.

The (Ir-)Reproducibility of fMRI Findings

During the investigation of the neural correlates of social attention, we were able to decode social feature position from early and extrastriate visual cortices (see chapter 5). Based on the recent literature on the amygdala’s role in fast processing of threat and attention allocation to relevant stimuli (Gamer & Büchel, 2009; Méndez-Bértolo et al., 2016; Peck et al., 2013; Peck & Salzman, 2014), we had hypothesized that the amygdala might be involved in a network specialized for social attention. However, we were not able to detect saccade-related amygdala activity that would have supported a role of the amygdala in attention allocation. We were further unable to decode the location of the social feature from patterns of amygdala activations. These results should yet be interpreted with caution since there is overall increasing consensus that the freedom of choices in fMRI design, preprocessing and statistical analyses can yield contradicting results for the same datasets.

After Ioannidis’ proclamation that “most published research findings are false” (Ioannidis, 2005), the field of psychology has undergone a thorough revision of contemporary research practices. This did not leave the field of fMRI untouched; in contrast, various renowned fMRI experts dedicated time and effort to raising awareness of the pitfalls of fMRI research. Possibly the most famous effort was undertaken by Craig Bennett and co-workers who reported an activation cluster during social perspective taking in a dead Atlantic salmon (Bennett et al., 2009). When adjusting the statistical threshold for multiple comparisons, the observed activation within the identified voxel cluster disappeared. While the report triggered a lot of undesired bad press for fMRI in general, it did drive home the message that multiple comparison corrections are essential in fMRI analyses (Lyon, 2017). Other influential articles called for set standards in reporting fMRI results (Poldrack et al., 2008), warned against the dangers of circular analyses when using regions of interest for both selection and selective analysis (Kriegeskorte et al., 2009) and pointed to issues in low statistical power and too high flexibility in analyses leading to low generalizability of fMRI results (Poldrack et al., 2017).

Cumulatively, the rise in awareness of potential challenges in fMRI prompted various initiatives aimed at making fMRI research more reproducible and generalizable. Open data repositories like NeuroVault (Gorgolewski et al., 2015) encourage researchers to upload and share their unthresholded statistical fMRI maps to allow comparisons and replications across datasets. Similarly, the online platform Neurosynth permits the automated synthesis of fMRI literature, frequently used to conduct large-scale meta-analyses (Yarkoni et al., 2011). Additionally, the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016) is increasingly adopted across universities for the organization of neuroimaging data in a standardized, easy to understand manner. This allows researchers from different labs to easily pick up or expand the work on existing datasets. Various other tools, drawing on BIDS-formatted input, have been developed by the neuroscience community to facilitate standardized data handling and analyses. MRIQC (Esteban et al., 2017) allows the automatic extraction of quality measures from MRI data, whereas FMRIPREP (Esteban et al., 2019) has been developed for standardized fool-proof data preprocessing.

In our fMRI study reported in chapter 6, we have implemented the standardized FMRIPREP preprocessing pipelines but there are nonetheless many other caveats that could have potentially influenced our amygdala null finding. Overall, there is vast heterogeneity regarding amygdala activation by emotion tasks (Lipp et al., 2014). Various reproducibility studies investigating multiple regions of interest found previously reported results on amygdala activation to be least reproducible (Johnstone et al., 2005; Lipp et al., 2014; Plichta et al., 2012). After physiological noise correction, task-related amygdala activations were even less reproducible, which suggests that replicable amygdala findings might have been driven by physiological effects (Boubela et al., 2015; Lipp et al., 2014). Using a fMRI sequence with a high temporal resolution, Boubela and colleagues showed that signal changes captured by emotional tasks are commonly not located in the amygdala itself but seem to occur in the neighboring Basal Vein of Rosenthal (Boubela et al., 2015). Capitalizing on the idea of spurious amygdala activation, a recent study showed that the choice of smoothing kernel and motion correction during preprocessing yield altered amygdala activation patterns (Murphy et al., 2020). We did not spatially smooth our fMRI data at all, as FMRIPREP does not include any spatial smoothing by default and as it is not recommended to smooth data for multivariate pattern analyses. It therefore remains unclear whether observed null findings stem from physiological noise or because our contrasts of choice did not capture differential amygdala activation.

As already pointed out in our discussion of our fMRI results in chapter 6, there is increasing evidence for a role of the amygdala in the oculomotor control of (social) attention allocation. Accordingly, a recent study showed that primate amygdala activation was associated with saccade facilitation in emotional contexts (Maeda et al., 2019). Muscimol-induced inactivation of the central nucleus of the amygdala reduced contralateral saccades during a cueing task and reduced the frequency of saccades targeting a social interaction during the free-viewing of a movie. Another electrophysiology study revealed multidimensional selectivity, comprising of differential responses to stimulus timing, categories (social versus non-social) and reward magnitude, in the primate amygdala (Putnam & Gothard, 2019). Taken together, there is compelling evidence for a potential involvement of the amygdala in social attention, rendering it likely that our null findings do not capture the ground truth but that methodological issues or problems in experimental design prevents us from detecting social attention-related amygdala activation. A recent study demonstrated that a sample size of 80 participants is necessary to detect medium effects ($0.5 < d < 0.8$) in fMRI group analyses. Our final sample only included 34 participants for the saccade-related fMRI analyses, and it is therefore recommendable to repeat this study with an increased sample size, possibly via multi-site collaborations.

Outlook

While the research presented in the current thesis highlighted the reflexive nature of social attention, which is likely at least in part mediated by early visual cortex function, is also became evident that some questions are still left unanswered. Future replications of our or similar fMRI studies are necessary to identify the brain regions that are recruited during the allocation of reflexive social attention. Once crucial areas are recognized, it will be particularly helpful to assess their functional and structural connectivity to better understand how a fast, automatic detection of social features is achieved in the brain. A subcortical pathway to the amygdala is thought to mediate fast detections of threat (Méndez-Bértolo et al., 2016) and it is conceivable that a similar route enables the swift prioritization of human features in complex environments. While fMRI studies are limited with regards to their temporal resolution, intracranial recordings from humans or electrophysiological recordings in primates could help to clarify the direction and timing of signal propagation through the brain. Research into such network dynamics will help to elucidate whether the neural system for social attention is fundamentally different from the perception of non-social scenes (also see (Sylvester et al., 2020).

On both the behavioral and neural level, investigations into psychopathological mechanisms of social attention deserve further consideration. Our examination of the relationship of social attention and social anxiety during a real-life social interaction did not demonstrate any differences in gaze behavior between socially non-anxious and anxious individuals. It is worth addressing potential factors influencing this finding in future studies, such as perceived social evaluation or state anxiety. As differences in gaze towards social information have long been implicated in autism spectrum disorder (ASD) (Frazier et al., 2017), it could be fruitful to assess our paradigm on reflexive social attention in ASD to test whether deviations already become apparent in the reflexive responses to complex social scenes. If behavioral differences are established, an examination of the neural responses to these complex social scenes might yield valuable insights. Bearing in mind the advantages of network analyses (Borsboom & Cramer, 2013), it would be particularly interesting to focus on specific symptoms of ASD and social anxiety and evaluate their position in a network of symptoms rather than to focus on a single categorization into participants with and without the disorder in question. As these types of analyses require large sample sizes, this might only be achievable through multi-site collaborations.

Concluding Remarks

When we perceive and interact with our environment, we are biased to spend a disproportionate amount of time on the visual exploration of fellow human beings. The experimental work in the present thesis demonstrated that this bias is reflexive and occurs even after very brief presentation durations of complex social scenes. However, when only parts of these scenes are invisible, thereby limiting bottom-up information, scan paths of observers still focus on social elements, indicating that voluntary mechanisms contribute to fixations on social features. Social attention is thus not only reflexive but also encompasses top-down attentional control. We have further shown that gaze is not only impacted by real human faces but also human-like, artificial faces, although to a lesser extent. Social attention consequently varies depending on the type of stimulus material employed. Supporting this notion, we showed that gaze towards a fellow human being is impacted by a potential social interaction during a real-life study. Diverging from laboratory studies, we were further unable to detect differential gaze patterns in social anxiety during the non-interactive and interactive phases of this real-life experiment. Our results thus substantiate recent initiatives calling for increased ecological validity in social neuroscience research (Foulsham & Kingstone, 2017; Risko et al., 2012). Lastly, our neuroimaging results demonstrate that the fast prioritization of social features seems to be, at least partly, achieved by striate and extrastriate visual areas. Social information thus impacts early visual attention both on a behavioral and a neural level.

Attentional Selection of Social Features Persists Despite Restricted Bottom-Up Information and Affects Temporal Viewing Dynamics

S1 Eye-Tracking Preprocessing Details

Saccades and were detected from the recorded eye-tracking data by using a velocity and an acceleration threshold of 30°/s or 8000°/s^2, respectively. Time periods between saccades were defined as fixations and their coordinates (x, y) and durations saved for subsequent analyses. Fixations were drift-corrected with reference to a baseline period of 300 ms during the presentation of the fixation cross directly preceding stimulus presentation. Similar to previous studies, fixations that deviated from this baseline were identified by a recursive outlier removal procedure that was applied separately to x- and y-baseline-coordinates (see End & Gamer, 2017; Flechsenhar & Gamer, 2017; Rösler et al., 2017). In detail, this procedure temporarily removed the highest and lowest coordinates for each participant from the baseline distribution and compared it to the mean and standard deviation of the remaining data. If these values were more than three standard deviations below or above this mean, they were marked as outliers, otherwise, they were returned to the distribution. This procedure was repeated until no more values were defined as outliers. Baseline outliers or missing baseline coordinates (social scene trials: *M* = 5.89%, *SD* = 6.15%; non-social scene trials: *M* = 4.78%, *SD* = 4.73%) were replaced with the mean baseline position of all scenes with valid baseline position data of the respective participant.

Following baseline correction of all fixations within each trial, a fixation density map was created by storing fixation coordinates in an empty matrix with the same dimensions as the currently used stimuli (1200 x 900 pixels). Fixations were weighted by their duration in ms. The resulting map was smoothed with an isotropic Gaussian kernel with a standard deviation of 32 pixels corresponding to 1° visual angle in positive and negative direction using the R package *spatstat* (version 1.47.0; Baddeley et al., 2015). The resulting 2° of visual angle correspond to the functional field of the human fovea centralis. In a final step, the fixation density maps were normalized to values between 0 and 1.

S2 Region of Interest Details

Saliency maps were used to identify regions of high saliency (above the eighth percentile of the saliency map) and areas of low saliency (below the eighth percentile) for all stimuli. Additionally, we manually defined regions for head and body of depicted human beings for social scenes using the software GNU Image Manipulation Program (GIMP; Version 2.8.10). A ROI could only be defined once, so that areas of high and low saliency for social scenes were restricted to those that had not yet been defined by head and body ROIs. In a previous study, we already demonstrated that social ROIs (head and body) had a lower mean saliency than highly salient non-social image regions for this stimulus set (End & Gamer, 2017). To determine the extent to which each ROI was fixated by the participant, we calculated the sum of fixation density values for each ROI and divided it by the sum of fixation density values for the whole stimulus. To take into account the different sizes of ROIs, this proportion was then normalized by dividing it by the area of the ROI. These area-normed fixation density scores were analyzed using a 2 x 4 repeated-measures ANOVA with factors viewing condition (free-viewing, gaze-contingent) and ROI (head, body, low saliency, high saliency).

S3 References of Stimuli from Databases

Stimuli taken from different databases (*n* = 67) with according reference and content differentiation for this study. The remaining stimuli (*n* = 93) were taken from internet sources (e.g., Google, Flickr etc.)

Database	Reference	Content
Emotional Picture Set	9.jpg	social
Emotional Picture Set	119.jpg	social
Emotional Picture Set	131.jpg	social
Emotional Picture Set	133.jpg	social
Emotional Picture Set	138.jpg	social
Emotional Picture Set	191.jpg	social
Emotional Picture Set	196.jpg	social
Emotional Picture Set	197.jpg	social
Emotional Picture Set	205.jpg	social
Emotional Picture Set	267.jpg	non-social
Emotional Picture Set	280.jpg	non-social
International Affective Picture System	5199.jpg	social
International Affective Picture System	9150.jpg	social
International Affective Picture System	9186.jpg	non-social
International Affective Picture System	9422.jpg	non-social

McGill Calibrated Colour Image Database	Merry_0005_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_0014_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_0060_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_0064_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_florida0011.jpg	social
McGill Calibrated Colour Image Database	Merry_florida0017.jpg	non-social
McGill Calibrated Colour Image Database	Merry_mexico0072.jpg	social
McGill Calibrated Colour Image Database	Merry_mexico0143.jpg	social
McGill Calibrated Colour Image Database	Merry_0081.jpg	non-social
McGill Calibrated Colour Image Database	Pippin_city6.jpg	social
McGill Calibrated Colour Image Database	Pippin_city66.jpg	social
Nencki Affective Picture System	Animals_025.jpg	non-social
Nencki Affective Picture System	Animals_048_h.jpg	non-social
Nencki Affective Picture System	Animals_074_h.jpg	non-social
Nencki Affective Picture System	Animals_102_h.jpg	non-social
Nencki Affective Picture System	Animals_128_h.jpg	non-social
Nencki Affective Picture System	Animals_194_h.jpg	non-social
Nencki Affective Picture System	Animals_195_h.jpg	non-social
Nencki Affective Picture System	Animals_201_h.jpg	non-social
Nencki Affective Picture System	Animals_218_h.jpg	non-social
Nencki Affective Picture System	Faces_023_h.jpg	social
Nencki Affective Picture System	Faces_265_h.jpg	social
Nencki Affective Picture System	Faces_290_h.jpg	social
Nencki Affective Picture System	Faces_302_h.jpg	social
Nencki Affective Picture System	Landscapes_016_h.jpg	non-social
Nencki Affective Picture System	Landscapes_025_h.jpg	non-social
Nencki Affective Picture System	Landscapes_040_h.jpg	non-social
Nencki Affective Picture System	Landscapes_043_h.jpg	non-social
Nencki Affective Picture System	Landscapes_064_h.jpg	non-social

Nencki Affective Picture System	Landscapes_071_h.jpg	non-social
Nencki Affective Picture System	Landscapes_085_h.jpg	non-social
Nencki Affective Picture System	Landscapes_178_h.jpg	non-social
Nencki Affective Picture System	Objects_002_h.jpg	non-social
Nencki Affective Picture System	Objects_013_h.jpg	non-social
Nencki Affective Picture System	Objects_058_h.jpg	non-social
Nencki Affective Picture System	Objects_183_h.jpg	non-social
Nencki Affective Picture System	Objects_202_h.jpg	non-social
Nencki Affective Picture System	Objects_214_h.jpg	non-social
Nencki Affective Picture System	People_009_h.jpg	social
Nencki Affective Picture System	People_015_h.jpg	social
Nencki Affective Picture System	People_022_h.jpg	social
Nencki Affective Picture System	People_054_h.jpg	social
Nencki Affective Picture System	People_058_h.jpg	social
Nencki Affective Picture System	People_109_h.jpg	social
Nencki Affective Picture System	People_116_h.jpg	social
Nencki Affective Picture System	People_131_h.jpg	social
Nencki Affective Picture System	People_157_h.jpg	social
Nencki Affective Picture System	People_158_h.jpg	social
Nencki Affective Picture System	People_167_h.jpg	social
Nencki Affective Picture System	People_182_h.jpg	social
Nencki Affective Picture System	People_195_h.jpg	social
Object and Semantic Images and Eyetracking dataset	118.jpg	non-social

Table S1. Detailed video content descriptions.

Neutral Videos	
1	A green canyon with a wild river in the middle and several waterfalls streaming into it. Filmed out of an airplane.
2	A rollercoaster ride out of the perspective of the cart. You can see the rail track and bald trees along the rollercoaster.
3	Scene in an automobile factory. You can see a skeleton of a car and many robot arms that are working on it.
4	Walking along a trail out of a forest to a clearing. There is a river to the left.
5	Colorful balloons with tags on them are released into the sky.
6	Two hot-air balloons are floating over a sunny, hilly landscape.
7	At a train track. You can see a wagon train with machines on it on the left and a normal train on the right, both moving.
8	A ski-run down a snowy hill out of the perspective of the skier. Entering a forest after a while.
9	Scene filmed from a boat, looking at the water and the coast with tall buildings and trees.
10	Lava streaming down a volcanic mountain. There is water at the foot of the mountain and steam rising.
11	A small boat is struggling in a very rough sea.
12	A lot of old, rusty vintage cars on a snowy field.
13	The opening of a watergate filmed from inside the watergate.
14	The slow passing of a wagon train, filmed from a bridge. There are trees next to the tracks.
15	A sunflower field, the flowers are moving in the wind.
Real Faces Only Videos	
16	Inside an airport hall. There is a woman (real face) and a man (real face) with luggage on the left, waiting. On the right, an employee (real face) is moving luggage carts. In the end, a man (real face) walks into the frame from the left.
17	A field with sheep with a man (real face) in reflective clothing standing at the fence, looking at the sheep. The sheep are running towards him. After a while, the man (real face) turns around and leaves the frame.
18	A crane moves a big tree trunk. Two men (real faces) in reflective clothing enter the scene and monitor the action and give directions to the person (real face) operating the crane.

19	A big ship enters a harbor with other ships on left and right. A couple of people (real faces) are watching. There is something burning in the sand and smoke emerges.
20	A man (real face) in a workshop puts together a wooden object. There are tools laying around. He takes a screwdriver and screws and is explaining what he is doing.
21	There are three children (real faces) playing on a playground. One is on a swing, another tries to get on another swing and the third one is standing in front of them. There are trees in the background and a cyclist (real face) passes by.
22	Indigenous farmers (real faces) in the mountains, with a herd of alpacas surrounding them. The camera moves and shows the valley with a river.
23	Scene in an indigenous village. A woman (real face) grills an animal on a fire. A child (real face) walks up to her and carries the food in a big palm leaf to a group making music. In the background, you can see people (real faces) working with other big plant leaves.
24	A group of children (real faces) is singing in a choir. On the left, there is the choir director (real face) singing along and on the right, a man (real face) is accompanying the group with a guitar.
25	Three children (real faces) are playing on a trampoline, playfully wrestling for a ball. You can see the neighborhood in the background.
26	A scene on a beach, with people (real faces) going for a walk or exercising by the water.
27	A man (real face) is cooking food outside of a house in a pan. A woman (real face) is standing next to him and a man (real face) is sitting in the background, reading a newspaper. There are two motorcycles parked next to the cooking station.
28	Three young adults (real faces) are playing frisbees. One is giving directions.
29	A scene of a construction site next to a highway. A huge drill creates a hole in the ground and a man (real face) in reflective clothes stands next to it and supervises the action.
30	A girl (real face) is slowly riding a horse in a sand square.
Real and Artificial Faces Videos	
31	Scene at an amusement park. There is a big locomotive with a face on it (artificial face) in the middle with wagons of animal sculptures (artificial faces) in the back. There is water coming out of the locomotive and the animal's mouths. Children and adults (real faces) are gathering around it.
32	The entrance to a Hinduistic temple. There are two sculptures of goddesses (artificial faces) next to the entrance. You can see two men inside (real faces), who then come out of the temple.
33	A big human-like stone sculpture (artificial face) is moved by people (real faces) with two ropes on each side. People (real faces) are standing around, either pulling, taking pictures or watching.

34	A scene from a theatre play. There are two actresses (real faces) dressed up as clowns in the middle, acting. On each side of the stage is a clown-statue made out of balloons (artificial faces).
35	Sculptors (real faces) are shown making human figurines (artificial face) out of clay.
36	A man and an artist (real faces) are posing for pictures in front of a big painting of the man (artificial face).
37	An old man (real face) is sitting in a small room stuffed with human-like figures out of clay and straw (artificial faces). He is sitting on the floor, making new ones out of straw (artificial faces).
38	A scene inside of a Lego store. There are a lot of customers (real faces) looking at products. The room is decorated with big sculptures (artificial faces) and pictures of characters from different movies and TV shows (artificial faces).
39	Four men (real faces) are standing in front of a wall, on which there are paintings of rugby and baseball players (artificial faces).
40	A scene from a cosplay show. First, a host (real face) talks to the audience. Then, people in costumes enter the stage, dressed up as animals or ninjas (artificial faces).
41	A clergyman (real face) is sitting in a temple playing an instrument. He is surrounded by several human-like figurines (artificial faces).
42	A scene inside of a subway with a ventriloquist (real face), who is performing with his doll (artificial face) towards the camera. There are people (real faces) in the back watching.
43	A group of street artists is standing on a scaffold and is busy painting a wall with graffiti letterings and faces (artificial faces).
44	Two men (real faces) play a game in which they throw scraps of paper in bowls located on a round, rotating table. In the background of the game are figurines (artificial faces) and paintings of people (artificial faces).
45	A man (real face) is standing in front of many portrait paintings (artificial faces) and giving a presentation to an audience.
Artificial Faces Only Videos	
46	A scene from a puppet show in a church. There is one puppet (artificial face) shown above a table. In the background are church windows with biblical scenes on them.
47	A big sculpture (artificial face) of the manger scene is shown inside of a church.
48	Scene walking towards a visitor platform in front of a big human-like stone sculpture (artificial face) in a plateau landscape.
49	Scene filming a wall with graffiti art with letterings and comic characters (artificial faces).
50	Two brazen, realistic sculptures of men (artificial faces).
51	An exhibition on Native Americans. Moving figurines (artificial faces) are shown doing different kinds of work in a village.

52	The entrance of a gothic church with several figurines of clergymen (artificial faces).
53	A big boundary wall with Spanish writing and a painting of a cartoon-like human (artificial face) lying on the floor.
54	A Disney rollercoaster. There is a big poster in the middle with cartoon characters (artificial faces) on it.
55	A scene filmed out of a car. The camera is moving through an entrance gate and passing a buddha statue (artificial face) next to the road. There are other cars on the street.
56	A scene filmed out of a cart of an amusement park ride. On the left and right is a fictional forest landscape with moving animal figurines (artificial faces).
57	A big building with a statue of a meditating man (artificial face) in the front.
58	A big painting of two people (artificial faces) and a Mickey Mouse (artificial face) standing in a desert catching fire.

Social anxiety is associated with alterations in heart rate but not gaze in a real social interaction

Fixation proportions on eye and mouth region

A large part of the literature on gaze behavior in social anxiety does not focus on head but on eye fixations (e.g., Horley et al. 2004; Boll et al. 2016; Moukheiber et al. 2012; Horley et al. 2003; Moukheiber et al. 2010). To investigate whether we can observe a relationship between social anxiety and gaze behavior with a more fine-grained spatial resolution, we re-coded fixations such that head fixations were labeled as either eye or mouth fixations. The nose tip was used to determine higher or lower face regions which were respectively labeled as eye or mouth region. The reported results yet need to be interpreted with caution because we cannot confidently state that the eye-tracking procedure employed in our study allows for such high resolution. While the accuracy of the glasses in itself should be high enough to distinguish between upper and lower face regions at a distance of approximately 2.5 meters, a three-point validation at the end of the experiment revealed gaze drifts in virtually all participants. On average, even after exclusion of the most extreme drift outliers, the absolute value of the gaze drift amounted to 10 (*SD* = 12) pixels on the x- and 37 pixels (*SD* = 39) on the y-axis. We applied a linear drift correction on our data assuming that the drift increases linearly throughout the experiment. As this is an assumption which we cannot test, fixations of the eye and the mouth region might be hard to tell apart (see Figure S1).

Figure S1. View of the confederate from the position of the participant. The yellow line indicates the average drift gaze across participants (x = 10 pixels, y = 37 pixels, here depicted as a negative value because the large majority of drifts on the y-axis were negative).

We nevertheless performed an ANCOVA with fixation proportions as the dependent variable, ROI (eyes versus mouth) and experimental phase (waiting, phone and interaction phase) as factorial predictors and SIAS score and gender as covariates. We found a significant main effect of ROI ($F_{(1,68)}$ = 6.56, p = .013, η^2 = .10), as fixation proportions differed overall between mouth and eye regions (see Figure S2). Higher fixations on eye and mouth regions in the interaction phase become apparent in a significant main effect of experiment phase ($F_{(2,136)}$ = 165.15, ε = 0.79, p < .001, η^2 = .37). We also observed a significant interaction of experiment phase and ROI ($F_{(2,136)}$ = 3.92, ε = 0.58, p = .045, η^2 = .42) which mainly describes a larger increase of mouth as compared to eye fixations throughout the three experiment phases (see Figure 3). Social anxiety did not seem to have a significant effect on fixation proportions ($F_{(1,68)}$ = 0.62, p =.433, η^2 = .004) and none of the individual interactions with experimental phase and ROI, nor the triple interaction between all three predictors or the effects of gender reached statistical significance (all p > .32).

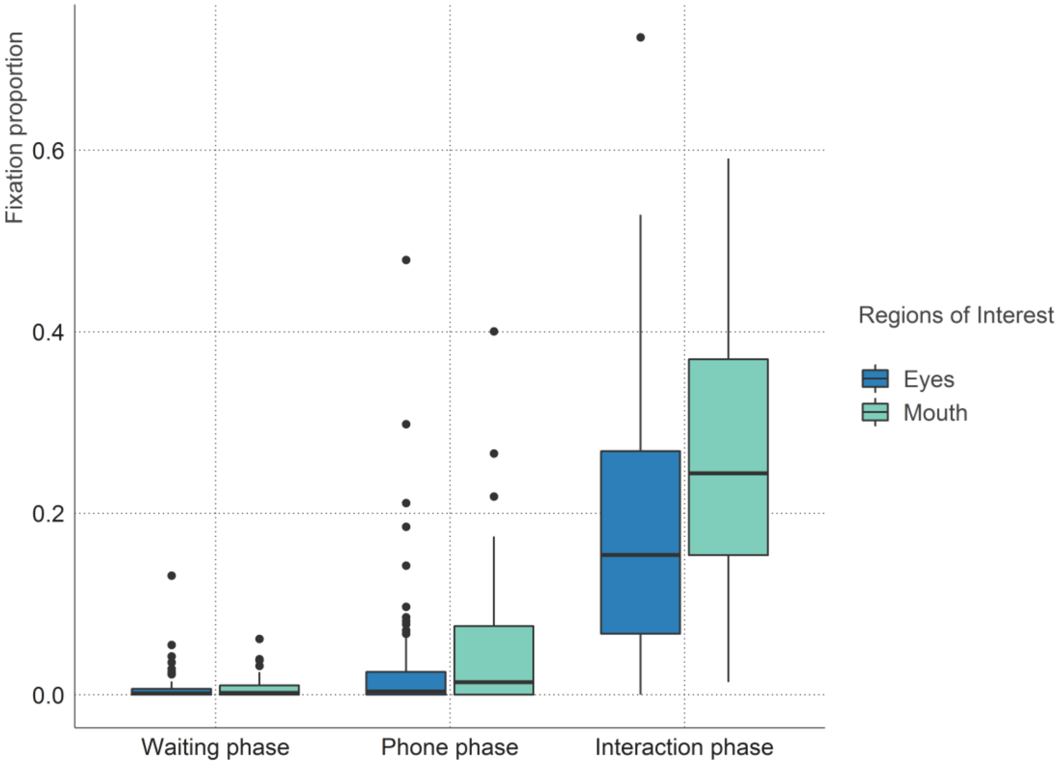

*Figure S2. Eye and mouth region fixation proportions during waiting, phone and interaction phase. Outliers are denoted by black dots and defined as points further than 1.5 * interquartile range of the lower or upper hinge.*

Table S2. Pearson correlations between questionnaire scores and the number of fixations on the body and head of the confederate during the three experimental phases

	Body			Head		
	Waiting phase	Phone phase	Interaction phase	Waiting phase	Phone phase	Interaction phase
SIAS	-.11	-.12	.03	-.09	-.08	-.03
BFNE	-.05	-.11	.05	.05	-.03	-.04
AQ	-.15	-.12	-.01	-.08	-.02	-.06
AQ-S	-.09	-.07	-.16	-.12	-.02	.01
AQ-F	-.21	-.16	.06	-.07	-.06	-.06
AQ-C	-.06	-.05	.13	.02	.04	-.11
STAI	.03	-.02	.06	.05	.07	-.01
BFI-E	.15	.06	.01	.08	.07	.13
BFI-A	-.18	-.09	-.12	.07	.09	.10
BFI-C	.05	-.09	-.10	.04	.13	.18
BFI-N	-.15	-.22	-.01	-.01	.04	-.10
BFI-O	-.02	-.03	.13	.00	-.02	-.10

Note. n = 71. SIAS = Social Interaction Anxiety Scale; BFNE = Brief Fear of Negative Evaluation Scale; AQ-K =Autism Quotient short version (Global Score); AQ-S = Autism Quotient - Social Competence; AQ-F = Autism Quotient - Fantasy; AQ-C = Autism Quotient - Communication; STAI = State Trait Anxiety Inventory; BFI-E = Big Five Inventory (short version) – Extraversion; BFI-A = Big Five Inventory – Agreeableness; BFI-C = Big Five Inventory – Conscientiousness; BFI-N = Big Five Inventory – Neuroticism; BFI-O = Big Five Inventory – Openness. None of the correlations was statistically significant (even without controlling for multiple comparisons).

Table S3. Pearson correlations between questionnaire scores, heart rate (HR) and high frequency heart rate variability (HF-HRV) during the three experimental phases

	HR			HF-HRV		
	Waiting phase	Phone phase	Interaction phase	Waiting phase	Phone phase	Interaction phase [a]
gender	.033	.070	.120	.21	.19	.22
SIAS	.30*	.27*	.30*	-.03	.01	.03
BFNE	.02	.03	.12	.03	.11	.02
AQ	.21	.22	.18	.15	-.20	.01
AQ-S	.19	.19	.16	-.14	-.13	.02
AQ-F	.13	.17	.14	-.17	-.30*	.01
AQ-C	.15	.16	.12	-.02	-.03	.00
STAI	.16	.14	.11	-.03	.01	.08
BFI-E	-.16	-.16	-.13	.05	.08	-.02
BFI-A	-.34**	-.27*	-.22	.22	.14	.10
BFI-C	-.41**	-.41**	-.26*	.25	.14	.12
BFI-N	.12	.12	.15	.05	.04	.19
BFI-O	-.10	-.13	-.16	.08	.08	-.06

Note. n = 71. SIAS = Social Interaction Anxiety Scale; BFNE = Brief Fear of Negative Evaluation Scale; AQ-K =Autism Quotient short version (Global Score); AQ-S = Autism Quotient - Social Competence; AQ-F = Autism Quotient - Fantasy; AQ-C = Autism Quotient - Communication; STAI = State Trait Anxiety Inventory; BFI-E = Big Five Inventory (short version) – Extraversion; BFI-A = Big Five Inventory – Agreeableness; BFI-C = Big Five Inventory – Conscientiousness; BFI-N = Big Five Inventory – Neuroticism; BFI-O = Big Five Inventory – Openness

[a] n = 70

* p < .05, not corrected for multiple comparisons

** p < .01, not corrected for multiple comparisons

Table S4. Pearson correlations between questionnaire scores and skin conductance levels during the three experimental phases.

	SCL		
	Waiting phase	Phone phase	Interaction phase
gender	.07	.00	.14
SIAS	.01	-.03	-.07
BFNE	-.06	-.06	-.14
AQ	-.11	-.08	-.16
AQ-S	-.06	-.06	-.08
AQ-F	-.14	-.05	-.17
AQ-C	-.08	-.08	-.16
STAI	-.05	-.10	-.14
BFI-E	.19	.19	.23
BFI-A	.04	.07	-.12
BFI-C	-.11	-.12	-.02
BFI-N	-.04	-.07	-.07
BFI-O	-.16	-.17	-.24*

Note. n = 71. SIAS = Social Interaction Anxiety Scale; BFNE = Brief Fear of Negative Evaluation Scale; AQ-K =Autism Quotient short version (Global Score); AQ-S = Autism Quotient - Social Competence; AQ-F = Autism Quotient - Fantasy; AQ-C = Autism Quotient - Communication; STAI = State Trait Anxiety Inventory; BFI-E = Big Five Inventory (short version) – Extraversion; BFI-A = Big Five Inventory – Agreeableness; BFI-C = Big Five Inventory – Conscientiousness; BFI-N = Big Five Inventory – Neuroticism; BFI-O = Big Five Inventory – Openness

* p < .05, not corrected for multiple comparisons

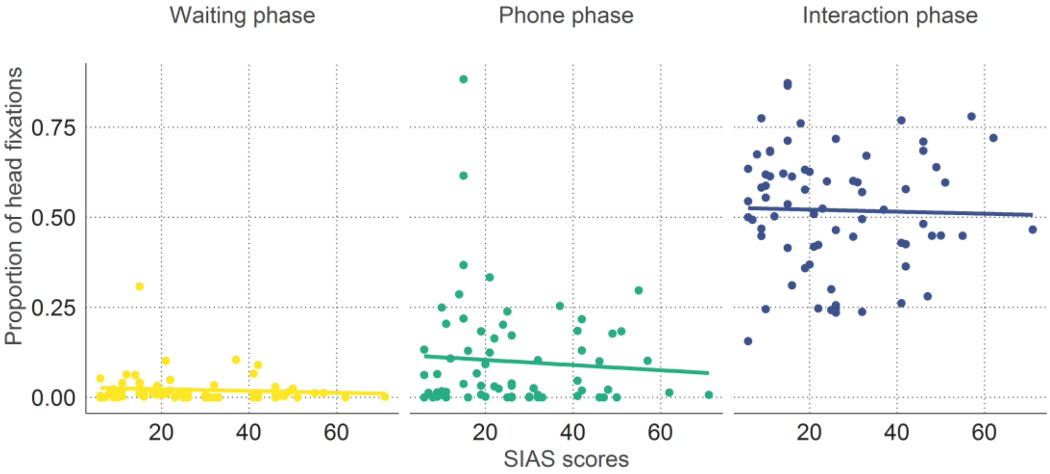

Figure S3. Mean fixation proportion on heads as a function of social interaction anxiety scale (SIAS) scores per phase.

Table S5. Post-hoc contrasts of the interaction effect between ROI and experimental phase on fixation proportion, all p-values are Tukey-corrected for multiple comparisons.

Contrast	Estimate	SE	df	t	p
Waiting.phase,head - Phone.phase,head	-0.107	0.028	257.660	-3.888	0.002
Waiting.phase,head - Interaction.phase,head	-0.513	0.028	257.660	-18.591	0.000
Waiting.phase,head - Waiting.phase,body	-0.178	0.035	187.133	-5.133	0.000
Waiting.phase,head - Phone.phase,body	-0.054	0.032	186.115	-1.699	0.534
Waiting.phase,head - Interaction.phase,body	-0.077	0.032	186.115	-2.423	0.154
Phone.phase,head - Interaction.phase,head	-0.406	0.028	257.660	-14.702	0.000
Phone.phase,head - Waiting.phase,body	-0.070	0.032	186.115	-2.201	0.242
Phone.phase,head - Phone.phase,body	0.053	0.035	187.133	1.538	0.640
Phone.phase,head - Interaction.phase,body	0.030	0.032	186.115	0.945	0.934
Interaction.phase,head - Waiting.phase,body	0.336	0.032	186.115	10.532	0.000
Interaction.phase,head - Phone.phase,body	0.459	0.032	186.115	14.402	0.000
Interaction.phase,head - Interaction.phase,body	0.436	0.035	187.133	12.608	0.000
Waiting.phase,body - Phone.phase,body	0.123	0.028	257.660	4.468	0.000
Waiting.phase,body - Interaction.phase,body	0.100	0.028	257.660	3.632	0.005
Phone.phase,body - Interaction.phase,body	-0.023	0.028	257.660	-0.836	0.961

Table S6. Post-hoc contrasts for experimental phase on heart rate, all p-values are Tukey-corrected for multiple comparisons.

Contrast	Estimate	SE	df	t	p
Waiting.phase - Phone.phase	-0.256	0.755	136.000	-0.339	0.939
Waiting.phase - Interaction.phase	-4.472	0.755	136.000	-5.922	0.000
Phone.phase - Interaction.phase	-4.217	0.755	136.000	-5.584	0.000

Table S7. Post-hoc contrasts for experimental phase on heart rate variability, all p-values are Tukey-corrected for multiple comparisons.

Contrast	Estimate	SE	df	t	p
Waiting.phase - Phone.phase	-1.292	2.939	134.000	-0.440	0.899
Waiting.phase - Interaction.phase	9.069	2.939	134.000	3.085	0.007
Phone.phase - Interaction.phase	10.361	2.939	134.000	3.525	0.002

Table S8. Post-hoc contrasts for experimental phase on skin conductance level, all p-values are Tukey-corrected for multiple comparisons.

Contrast	Estimate	SE	df	t	p
Waiting.phase - Phone.phase	-0.908	0.384	136.000	-2.362	0.051
Waiting.phase - Interaction.phase	-3.238	0.384	136.000	-8.423	0.000
Phone.phase - Interaction.phase	-2.330	0.384	136.000	-6.061	0.000

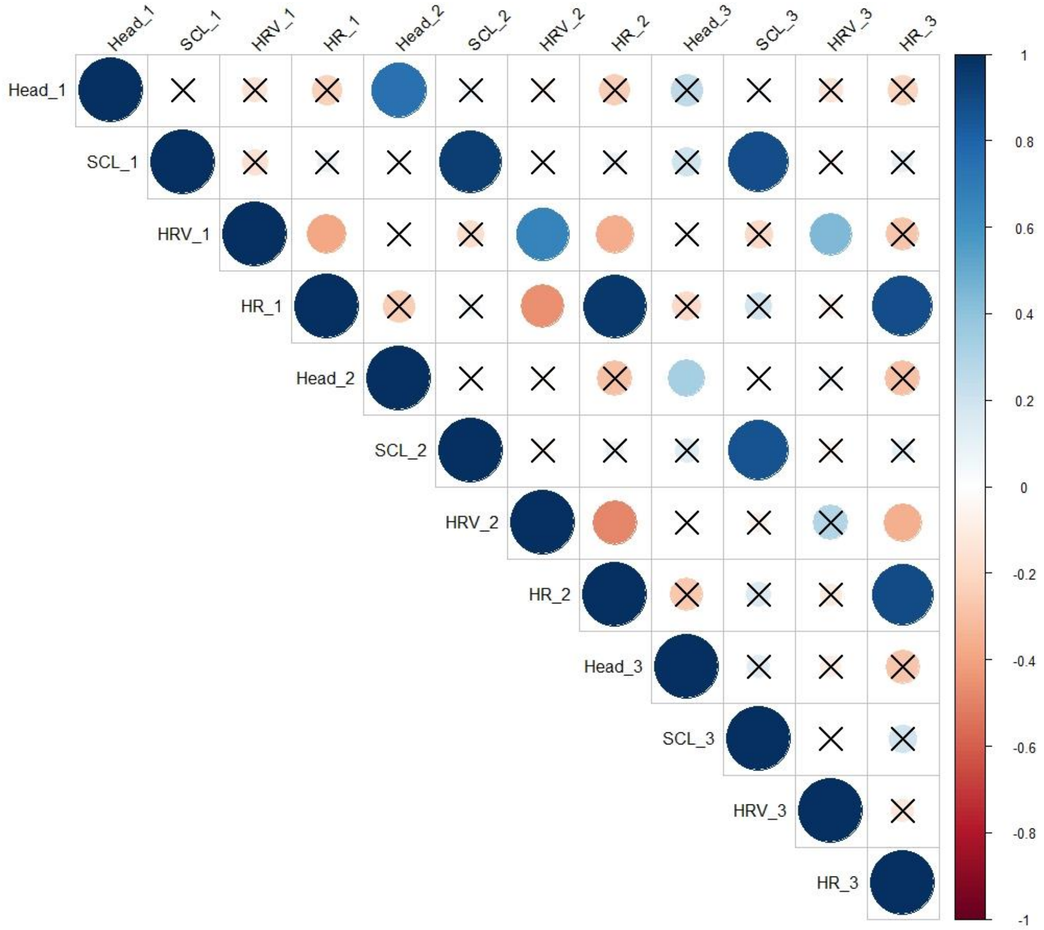

Figure S4. Correlation matrix of all dependent variables. Head = proportion of head fixations, SCL = mean skin conductance level, HRV = mean heart rate variability, HR = mean heart rate; 1: waiting phase, 2: phone phase, 3: interaction phase. All correlations p > .01, uncorrected for multiple comparisons, are crossed out.

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature, 372(6507), 669–672. https://doi.org/10.1038/372669a0

Adolphs, R. (2006). Is the Human Amygdala Specialized for Processing Social Information? *Annals of the New York Academy of Sciences*, *985*(1), 326–340. https://doi.org/10.1111/j.1749-6632.2003.tb07091.x

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, *1191*(1), 42–61. https://doi.org/10.1111/j.1749-6632.2010.05445.x

Adolphs, R, Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72. https://doi.org/10.1038/nature03086

Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Nakachi, R., & Kashima, H. (2006). A deficit in discriminating gaze direction in a case with right superior temporal gyrus lesion. *Neuropsychologia*, *44*(2), 161–170. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2005.05.018

Allefeld, C., Görgen, K., & Haynes, J. D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, *141*, 378–392. https://doi.org/10.1016/j.neuroimage.2016.07.040

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

Altman, M. N., Khislavsky, A. L., Coverdale, M. E., & Gilger, J. W. (2016). Adaptive attention: How preference for animacy impacts change detection. *Evolution and Human Behavior*, *37*(4), 303–314. https://doi.org/10.1016/j.evolhumbehav.2016.01.006

Anderson, E. R., Veed, G. J., Inderbitzen-Nolan, H. M., & Hansen, D. J. (2010). An Evaluation of the Applicability of the Tripartite Constructs to Social Anxiety in Adolescents. *Journal of Clinical Child & Adolescent Psychology*, *39*(2), 195–207. https://doi.org/10.1080/15374410903532643

Anderson, N. C., Bischof, W. F., Laidlaw, K. E. W., Risko, E. F., & Kingstone, A. (2013). Recurrence quantification analysis of eye movements. *Behavior Research Methods*, *45*(3), 842–856. https://doi.org/10.3758/s13428-012-0299-5

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, *15*(5), 1–22. https://doi.org/10.1167/15.5.9

Antony, M. M., Federici, A., & Stein, M. B. (2008). Overview and Introduction to Anxiety Disorders. In *Oxford Handbook of Anxiety and Related Disorders* (pp. 3–15). https://doi.org/10.1093/oxfordhb/9780195307030.013.0001

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, *32*(8), 704–723. https://doi.org/10.1016/J.CPR.2012.09.004

Asmundson, G. J. G., & Stein, M. B. (1994). Selective processing of social threat in patients with generalized social phobia: Evaluation using a dot-probe paradigm. *Journal of Anxiety Disorders*, *8*(2), 107–117. https://doi.org/10.1016/0887-6185(94)90009-4

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R.*

Chapman and Hall/CRC Press.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384. https://doi.org/10.3758/BF03192707

Bandelow, B., Sher, L., Bunevicius, R., Hollander, E., Kasper, S., Zohar, J., Möller, H. J., Thibaut, F., Baranska-Rybak, W., Cubala, W. J., Fiellin, D., Kranzler, H. R., Moore, A., Rankans, E., Rasmussen, J., Saitz, R., Saravane, D., Schlaepfer, T. E., Tang, S. W., … Vega, J. (2012). Guidelines for the pharmacological treatment of anxiety disorders, obsessive-compulsive disorder and posttraumatic stress disorder in primary care. *International Journal of Psychiatry in Clinical Practice*, *16*(2), 77–84. https://doi.org/10.3109/13651501.2012.667114

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry*, *42*(2), S0021963001006643. https://doi.org/10.1017/S0021963001006643

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Baskerville, T. A., & Douglas, A. J. (2010). Dopamine and Oxytocin Interactions Underlying Behaviors: Potential Contributions to Behavioral Disorders. *CNS Neuroscience & Therapeutics*, *16*(3), e92–e123. https://doi.org/10.1111/j.1755-5949.2010.00154.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4*. https://doi.org/10.18637/jss.v067.i01

Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., Johnson, M. H., & team, the B. (2012). Precursors to Social and Communication Difficulties in Infants At-Risk for Autism: Gaze Following and Attentional Engagement. *Journal of Autism and Developmental Disorders*, *42*(10), 2208–2218. https://doi.org/10.1007/s10803-012-1450-y

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Beidel, D. C., Turner, S. M., & Dancu, C. V. (1985). Physiological, cognitive and behavioral aspects of social anxiety. *Behaviour Research and Therapy*, *23*(2), 109–117. https://doi.org/10.1016/0005-7967(85)90019-1

Belova, M. A., Paton, J. J., Morrison, S. E., & Salzman, C. D. (2007). Expectation Modulates Neural Responses to Pleasant and Aversive Stimuli in Primate Amygdala. *Neuron*, *55*(6), 970–984. https://doi.org/10.1016/J.NEURON.2007.08.004

Belova, M. A., Paton, J. J., & Salzman, C. D. (2008). Moment-to-moment tracking of state value in the amygdala. *Journal of Neuroscience*, *28*(40), 10023-10030.

Bennet, C., Baird, A., Miller, M., & Wolford, G. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: An argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results*, *1*(1), 1-5.

Benthall, J., Argyle, M., & Cook, M. (1976). *Gaze and Mutual Gaze*. https://doi.org/10.2307/3032267

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

Bentin, S., & Golland, Y. (2002). Meaningful processing of meaningless stimuli: The influence of perceptual experience on early visual processing of faces. *Cognition*, *86*(1), 1–14.

Bentin, S., Sagiv, N., Mecklinger, A., Friederici, A., & von Cramon, Y. D. (2002). Priming Visual Face-Processing Mechanisms: Electrophysiological Evidence. *Psychological Science*, *13*(2), 190–193.

https://doi.org/10.1111/1467-9280.00435

Bereczkei, T., Birkas, B., & Kerekes, Z. (2007). Public charity offer as a proximate factor of evolved reputation-building strategy: an experimental analysis of a real-life situation. *Evolution and Human Behavior*, *28*(4), 277–284. https://doi.org/10.1016/j.evolhumbehav.2007.04.002

Bermudez, M. A., Göbel, C., & Schultz, W. (2012). Sensitivity to Temporal Reward Structure in Amygdala Neurons. *Current Biology*, *22*(19), 1839–1844. https://doi.org/10.1016/J.CUB.2012.07.062

Berntson, G. G., Thomas Bigger, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., & Van der Molen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*(6), 623–648. https://doi.org/10.1111/j.1469-8986.1997.tb02140.x

Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & de Haan, E. H. F. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, *12*(6), 1048–1053. https://doi.org/10.3758/BF03206442

Bindemann, M., Burton, A. M., Langton, S. R. H., Schweinberger, S. R., & Doherty, M. J. (2007). The control of attention to faces. *Journal of Vision*, *7*(10), 15. https://doi.org/10.1167/7.10.15

Bindemann, M., Burton, A.M., & Langton, S. R. H. (2008). How do eye gaze and facial expression interact? *Visual Cognition*, *16*(6), 708–733. https://doi.org/10.1080/13506280701269318

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. (2008a). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2–3), 341–355. https://doi.org/10.1080/13506280701434532

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). 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). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, *17*(6–7), 904–924. https://doi.org/10.1080/13506280902758044

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009b). 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., & Kingstone, A. (2009). Human Social Attention. *Annals of the New York Academy of Sciences*, *1156*(1), 118–140. https://doi.org/10.1111/j.1749-6632.2009.04468.x

Boggia, J., & Ristic, J. (2015). Social event segmentation. *Quarterly Journal of Experimental Psychology*, *68*(4), 731–744. https://doi.org/10.1080/17470218.2014.964738

Boll, S., Bartholomaeus, M., Peter, U., Lupke, U., & Gamer, M. (2016). Attentional mechanisms of social perception are biased in social phobia. *Journal of Anxiety Disorders*, *40*, 83–93. https://doi.org/10.1016/J.JANXDIS.2016.04.004

Boone, M. L., McNeil, D. W., Masia, C. L., Turk, C. L., Carter, L. E., Ries, B. J., & Lewin, M. R. (1999). Multimodal Comparisons of Social Phobia Subtypes and Avoidant Personality Disorder. *Journal of Anxiety Disorders*, *13*(3), 271–292. https://doi.org/10.1016/S0887-6185(99)00004-3

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207. https://doi.org/10.1109/TPAMI.2012.89

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, *9*, 91-121.

Boubela, R. N., Kalcher, K., Huf, W., Seidel, E. M., Derntl, B., Pezawas, L., Našel, C., & Moser, E. (2015).

FMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. *Scientific Reports*, *5*. https://doi.org/10.1038/srep10499

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0060188

Broussard, J. I., Karelina, K., Sarter, M., & Givens, B. (2009). Cholinergic optimization of cue-evoked parietal activity during challenged attentional performance. *European Journal of Neuroscience*, *29*(8), 1711–1722. https://doi.org/10.1111/j.1460-9568.2009.06713.x

Buckner, J. D., Maner, J. K., & Schmidt, N. B. (2010). Difficulty Disengaging Attention from Social Threat in Social Anxiety. *Cognitive Therapy and Research*, *34*(1), 99–105. https://doi.org/10.1007/s10608-008-9205-y

Bull, R., & Gibson-Robinson, E. (1981). The Influences of Eye-Gaze, Style of Dress, and Locality on the Amounts of Money Donated to a Charity. *Human Relations*, *34*(10), 895–905. https://doi.org/10.1177/001872678103401005

Byrow, Y., Broeren, S., De Lissa, P., & Peters, L. (2016). Anxiety, Attachment & Attention: The Influence of Adult Attachment Style on Attentional Biases of Anxious Individuals. *Journal of Experimental Psychopathology*, 7(1), jep.046714. https://doi.org/10.5127/jep.046714

Byrow, Y., Chen, N. T. M., & Peters, L. (2016b). Time Course of Attention in Socially Anxious Individuals: Investigating the Effects of Adult Attachment Style. *Behavior Therapy*, *47*(4), 560–571. https://doi.org/10.1016/J.BETH.2016.04.005

Calder, A. J., Jenkins, R., Cassel, A., & Clifford, C. W. G. (2008). Visual representation of eye gaze is coded by a nonopponent multichannel system. *Journal of Experimental Psychology: General*, *137*(2), 244–261. https://doi.org/10.1037/0096-3445.137.2.244

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, *40*(8), 1129–1138. https://doi.org/10.1016/S0028-3932(02)00008-8

Cañigueral, R., & Hamilton, A. F. d. C. (2019). Being watched: Effects of an audience on eye gaze and prosocial behaviour. *Acta Psychologica*, *195*, 50–63. https://doi.org/10.1016/j.actpsy.2019.02.002

Capitanio, J. P., Machado, C. J., & Amaral, D. (2006). Amygdalectomy and responsiveness to novelty in Rhesus monkeys (Macaca mulatta): Generality and individual consistency of effects. https://doi.org/10.1037/1528-3542.6.1.73

Capozzi, F., & Ristic, J. (2018). How attention gates social interactions. *Annals of the New York Academy of Sciences*, *1426*(1), 179–198. https://doi.org/10.1111/nyas.13854

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

Cavallina, C., Puccio, G., Capurso, M., Bremner, A. J., & Santangelo, V. (2018). Cognitive development attenuates audiovisual distraction and promotes the selection of task-relevant perceptual saliency during visual search on complex scenes. *Cognition*, *180*, 91–98. https://doi.org/10.1016/j.cognition.2018.07.003

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, *9*(12), 10–10. https://doi.org/10.1167/9.12.10

Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. In *Advances in neural information processing systems* (pp. 241-248).

Chalmers, J. A., Quintana, D. S., Abbott, M. J.-A., & Kemp, A. H. (2014). Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis. *Frontiers in Psychiatry*, *5*, 80. https://doi.org/10.3389/fpsyt.2014.00080

Chaminade, T., Zecca, M., Blakemore, S.-J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., & Umiltà, M. A. (2010). Brain Response to a Humanoid Robot in Areas Implicated in the Perception of Human Emotional Gestures. *PLoS ONE*, *5*(7), e11577. https://doi.org/10.1371/journal.pone.0011577

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased Spontaneous Attention to Social Scenes in 6-Month-Old Infants Later Diagnosed with Autism Spectrum Disorders. *Biological Psychiatry*, *74*(3), 195–203. https://doi.org/10.1016/J.BIOPSYCH.2012.11.022

Chen, N. T. M., & Clarke, P. J. F. (2017). Gaze-Based Assessments of Vigilance and Avoidance in Social Anxiety: a Review. *Current Psychiatry Reports*, *19*(9), 59. https://doi.org/10.1007/s11920-017-0808-4

Chen, N. T. M., Clarke, P. J. F., MacLeod, C., & Guastella, A. J. (2012). Biased Attentional Processing of Positive Stimuli in Social Anxiety Disorder: An Eye Movement Study. *Cognitive Behaviour Therapy*, *41*(2), 96–107. https://doi.org/10.1080/16506073.2012.666562

Chen, Y. ., Ehlers, A., Clark, D. ., & Mansell, W. (2002). Patients with generalized social phobia direct their attention away from faces. *Behaviour Research and Therapy*, *40*(6), 677–687. https://doi.org/10.1016/S0005-7967(01)00086-9

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Clark, D. M. (2005). A Cognitive Perspective on Social Phobia. In W. R. Crozier & L. E. Alden (Eds.), *The essential handbook of social anxiety for clinicians* (pp. 193–218). John Wiley.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. Social phobia: Diagnosis, assessment, and treatment. In *Social Phobia: Diagnosis, Treatment and Assessment* (68th ed., pp. 22–23).

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, *18*(2), 72–145. https://doi.org/10.1177/1529100617727266

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/J.NEURON.2008.04.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Correll, C. U., Detraux, J., De Lepeleire, J., & De Hert, M. (2015). Effects of antipsychotics, antidepressants and mood stabilizers on risk for physical diseases in people with schizophrenia, depression and bipolar disorder. *World Psychiatry*, *14*(2), 119–136. https://doi.org/10.1002/wps.20204

Coutrot, A., & Guyader, N. (2014). How saliency , faces , and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 1–17. https://doi.org/10.1167/14.8.5.doi

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and brain sciences*, *33*(2-3), 137-150.

Curio, C., Bülthoff, H., & Giese, M. (2011). *Dynamic faces: Insights from experiments and computation*. MIT Press.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

Dal Monte, O., Costa, V. D., Noble, P. L., Murray, E. A., & Averbeck, B. B. (2015). Amygdala lesions in rhesus macaques decrease attention to threat. *Nature Communications*, *6*(1), 10161. https://doi.org/10.1038/ncomms10161

Dare A. Baldwin. (1995). Understanding the Link Between Joint Attention and Language. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development.* (pp. 131-158.).

Davis, M., & Whalen, P. J. (2001). The amygdala: vigilance and emotion. *Molecular psychiatry*, *6*(1), 13-34.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, *20*(03), 775–803. https://doi.org/10.1017/S0954579408000370

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with Autism Fail to Orient to Naturally Occurring Social Stimuli. *Journal of Autism and Developmental Disorders*, *28*(6), 479–485. https://doi.org/10.1023/A:1026043926488

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology*, *40*(2), 271–283. https://doi.org/10.1037/0012-1649.40.2.271

Day-Brown, J. D., Wei, H., Chomsung, R. D., Petry, H. M., & Bickford, M. E. (2010). Pulvinar Projections to the Striatum and Amygdala in the Tree Shrew. *Frontiers in Neuroanatomy*, *4*, 143. https://doi.org/10.3389/fnana.2010.00143

De Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). *Individual differences in visual salience vary along semantic dimensions. 116*(24), 11687–11692. https://doi.org/10.1073/pnas.1820553116

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

De Souza, W. C., Eifuku, S., Tamura, R., Nishijo, H., & Ono, T. (2005). Differential Characteristics of Face Neuron Responses Within the Anterior Superior Temporal Sulcus of Macaques. *Journal of Neurophysiology*, *94*(2), 1252–1266. https://doi.org/10.1152/jn.00949.2004

Deaner, R. O., & Platt, M. L. (2003). Reflexive social attention in monkeys and humans. *Current Biology*, *13*(18), 1609-1613.

Deater-Deckard, K. (2001). Annotation: Recent Research Examining the Role of Peer Relationships in the Development of Psychopathology. *Journal of Child Psychology and Psychiatry*, *42*(5), S0021963001007272. https://doi.org/10.1017/S0021963001007272

Deiters, D. D., Stevens, S., Hermann, C., & Gerlach, A. L. (2013). Internal and external attention in speech anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(2), 143–149. https://doi.org/10.1016/J.JBTEP.2012.09.001

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior Inferotemporal Neurons of Monkeys Engaged in Object Recognition Can be Highly Sensitive to Object Retinal Position. *Journal of Neurophysiology*, *89*(6), 3264–3278. https://doi.org/10.1152/jn.00358.2002

Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., & Bodfish, J. W. (2012). Reward Circuitry Function in Autism During Face Anticipation and Outcomes. *Journal of Autism and Developmental Disorders*, *42*(2), 147–160. https://doi.org/10.1007/s10803-011-1221-1

Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). Screening for Autistic Spectrum Disorder in Children Aged 14–15 Months. II: Population Screening with the Early Screening of Autistic Traits Questionnaire (ESAT). Design and General Findings. *Journal of Autism and Developmental Disorders*, *36*(6), 713–722. https://doi.org/10.1007/s10803-006-0114-

1

Dorris, M. C., Olivier, E., & Munoz, D. P. (2007). Competitive Integration of Visual and Preparatory Signals in the Superior Colliculus during Saccadic Programming. *Journal of Neuroscience*, *27*(19).

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509–540. https://doi.org/10.1080/135062899394920

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M. H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage*, *36*(3), 511–521. https://doi.org/10.1016/j.neuroimage.2007.03.060

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14). https://doi.org/10.1167/8.14.18

Elsabbagh, M., Bedford, R., Senju, A., Charman, T., Pickles, A., & Johnson, M. H. (2014). What you see is what you get: contextual modulation of face scanning in typical and atypical development. *Social Cognitive and Affective Neuroscience*, *9*(4), 538–543. https://doi.org/10.1093/scan/nst012

Elsabbagh, M., Gliga, T., Pickles, A., Hudry, K., Charman, T., & Johnson, M. H. (2013). The development of face orienting mechanisms in infants at-risk for autism. *Behavioural Brain Research*, *251*, 147–154. https://doi.org/10.1016/J.BBR.2012.07.030

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

End, A., & Gamer, M. (2019). Task instructions can accelerate the early preference for social features in naturalistic scenes. *Royal Society Open Science*, *6*(3). https://doi.org/10.1098/rsos.180596

Eng, Z. H. D., Yick, Y. Y., Guo, Y., Xu, H., Reiner, M., Cham, T. J., & Chen, S. H. A. (2017). 3D faces are recognized more accurately and faster than 2D faces, but with similar inversion effects. *Vision Research*, *138*, 78–85. https://doi.org/10.1016/j.visres.2017.06.004

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, *45*(14), 3234–3241. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2007.06.022

Essau, C. A., Conradt, J., & Petermann, F. (1999). Frequency and comorbidity of social phobia and social fears in adolescents. *Behaviour Research and Therapy*, *37*(9), 831–843. https://doi.org/10.1016/S0005-7967(98)00179-X

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PloS One*, *12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Falck-Ytter, T., Rehnberg, E., & Bölte, S. (2013). Lack of Visual Orienting to Biological Motion and Audiovisual Synchrony in 3-Year-Olds with Autism. *PLoS ONE*, *8*(7), e68816. https://doi.org/10.1371/journal.pone.0068816

Farabee, D. J., Holcom, J. . M. L., Ramsey, S. L., & Cole, S. G. (1993). Social Anxiety and Speaker Gaze in a Persuasive Atmosphere. *Journal of Research in Personality*, 27(4), 365–376. https://doi.org/10.1006/JRPE.1993.1025

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. https://doi.org/10.1073/PNAS.152159999

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, *102*(47), 17245-17250.

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

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in cognitive sciences*, *10*(8), 382-390.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/CERCOR/1.1.1

Filiz-Ozbay, E., & Ozbay, E. Y. (2014). Effect of an audience in public goods provision. *Experimental Economics*, *17*(2), 200–214. https://doi.org/10.1007/s10683-013-9363-y

Flechsenhar, A. F., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PLoS ONE*, *12*(8), e0183799. https://doi.org/10.1371/journal.pone.0183799

Flechsenhar, A., Rösler, L., & Gamer, M. (2018). Attentional Selection of Social Features Persists Despite Restricted Bottom-Up Information and Affects Temporal Viewing Dynamics. *Scientific Reports*, *8*(1), 12555. https://doi.org/10.1038/s41598-018-30736-8

Fletcher-Watson, S., Leekam, S. R. R., Benson, V., Frank, M. C. C., & Findlay, J. M. 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, S., 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

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, S102. https://doi.org/10.1016/s1053-8119(09)70884-5

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research*, *50*(8), 779–795. https://doi.org/10.1016/j.visres.2010.01.019

Foulsham, T., & Kingstone, A. (2017). Are fixations in static natural scenes a useful predictor of attention in the real world? *Canadian Journal of Experimental Psychology*, 71(2). https://doi.org/10.1037/cep0000125

Foulsham, T., Teszka, R., & Kingstone, A. (2011). Saccade control in natural images is shaped by the information visible at fixation: Evidence from asymmetric gaze-contingent windows. *Attention, Perception, and Psychophysics*, *73*(1), 266–283. https://doi.org/10.3758/s13414-010-0014-5

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6. https://doi.org/10.1167/8.2.6

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931. https://doi.org/10.1016/J.VISRES.2011.07.002

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

Freese, J., & Amaral, D. (2009). Neuroanatomy of the primate amygdala. In P.J. Whalen & E. A. Phelps (Eds.), *The human amygdala* (pp. 3–42). Guilford Press.

Freeth, M., Chapman, P., Ropar, D., & Mitchell, P. (2010). Do Gaze Cues in Complex Scenes Capture and Direct the Attention of High Functioning Adolescents with ASD? Evidence from Eye-tracking. *Journal of Autism and Developmental Disorders*, *40*(5), 534–547. https://doi.org/10.1007/s10803-009-0893-2

Freeth, M., Foulsham, T., & Chapman, P. (2011). The influence of visual saliency on fixation patterns in individuals with Autism Spectrum Disorders. *Neuropsychologia*, *49*(1), 156–160. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2010.11.012

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLoS ONE*, *8*(1), e53286. https://doi.org/10.1371/journal.pone.0053286

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *36*(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don’t add up: why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(1), 72. https://doi.org/10.1186/s12916-015-0325-4

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, *5*(3), 490–495. https://doi.org/10.3758/BF03208827

Friesen, C. K., & Kingstone, A. (2003). Covert and overt orienting to gaze direction cues and the effects of fixation offset. *NeuoReport*, *14*(3), 489–493.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ‘theory of mind.’ *Trends in Cognitive Sciences*, *7*(2), 77–83. https://doi.org/10.1016/S1364-6613(02)00025-6

Gallup, A., Church, A. M., Miller, H., Risko, E. F., & Kingstone, A. (2016). Social Presence Diminishes Contagious Yawning in the Laboratory. *Scientific Reports*, *6*. https://doi.org/10.1038/srep25045

Gallup, A., Hale, J. J., Sumpter, D. J. T., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7245–7250. https://doi.org/10.1073/pnas.1116141109

Gallup, A., Vasilyev, D., Anderson, N., & Kingstone, A. (2019). Contagious yawning in virtual reality is affected by actual, but not simulated, social presence. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-018-36570-2

Gamble, A. L., & Rapee, R. M. (2009). The time-course of attentional bias in anxious children and adolescents. *Journal of Anxiety Disorders*, *23*(7), 841–847. https://doi.org/10.1016/J.JANXDIS.2009.04.001

Gamer, M., & Büchel, C. (2009). Amygdala activation predicts gaze toward fearful eyes. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(28), 9123–9126. https://doi.org/10.1523/JNEUROSCI.1883-09.2009

Gamer, M., Zurowski, B., & Büchel, C. (2010). Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9400–9405. https://doi.org/10.1073/pnas.1000985107

Garner, M., Mogg, K., & Bradley, B. P. (2006). Orienting and maintenance of gaze to facial expressions in social anxiety. *Journal of Abnormal Psychology*, *115*(4), 760–770. https://doi.org/10.1037/0021-843X.115.4.760

Garrido, M. I., Barnes, G. R., Sahani, M., & Dolan, R. J. (2012). Functional Evidence for a Dual Route to

Amygdala. *Current Biology*, *22*(2), 129–134. https://doi.org/10.1016/J.CUB.2011.11.056

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, *35*(4), 1674–1684. https://doi.org/10.1016/J.NEUROIMAGE.2007.02.003

Geisler, W. S., & Najemnik, J. (2005). Human and optimal eye movement strategies in visual search. *Journal of Vision*, *5*(8), 778–778. https://doi.org/10.1167/5.8.778

George, N., Driver, J., & Dolan, R. J. (2001). Seen Gaze-Direction Modulates Fusiform Activity and Its Coupling with Other Brain Areas during Face Processing. *NeuroImage*, *13*(6), 1102–1112. https://doi.org/10.1006/NIMG.2001.0769

Ghashghaei, H. T., Hilgetag, C. C., & Barbas, H. (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *NeuroImage*, *34*(3), 905–923. https://doi.org/10.1016/J.NEUROIMAGE.2006.09.046

Gilbert, P. (2001). Evolution and social anxiety. The role of attraction, social competition, and social hierarchies. *The Psychiatric Clinics of North America*, *24*(4), 723–751. https://doi.org/10.1016/S0193-953X(05)70260-4

Gilboa-Schechtman, E., Foa, E. B., & Amir, N. (1999). Attentional Biases for Facial Expressions in Social Phobia: The Face-in-the-Crowd Paradigm. *Cognition & Emotion*, *13*(3), 305–318. https://doi.org/10.1080/026999399379294

Ginsburg, G. S., Becker, E. M., Keeton, C. P., Sakolsky, D., Piacentini, J., Albano, A. M., Compton, S. N., Iyengar, S., Sullivan, K., Caporino, N., Peris, T., Birmaher, B., Rynn, M., March, J., & Kendall, P. C. (2014). Naturalistic Follow-up of Youths Treated for Pediatric Anxiety Disorders. *JAMA Psychiatry*, *71*(3), 310. https://doi.org/10.1001/jamapsychiatry.2013.4186

Glasgow, R. E., & Arkowitz, H. (1975). The behavioral assessment of male and female social competence in dyadic heterosexual interactions. *Behavior Therapy*, *6*(4), 488–498. https://doi.org/10.1016/S0005-7894(75)80005-0

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364. https://doi.org/10.1016/j.cognition.2014.11.040

Goffaux, V., & Rossion, B. (2006). Faces Are "Spatial"-Holistic Face Perception Is Supported by Low Spatial Frequencies. https://doi.org/10.1037/0096-1523.32.4.1023

Goffman, E. (2008). *Behavior in Public Places*. The Free Press. https://doi.org/papers3://publication/uuid/4EEC120E-5776-413F-B43F-044C09251F5F

Goldman, M., & Fordyce, J. (1983). Prosocial behavior as affected by eye contact, touch, and voice expression. *Journal of Social Psychology*, *121*(1), 125–129. https://doi.org/10.1080/00224545.1983.9924474

Gomes, N., Soares, S. C., Silva, S., & Silva, C. F. (2018). Mind the snake: Fear detection relies on low spatial frequencies. *Emotion*, *18*(6), 886–895. https://doi.org/10.1037/emo0000391

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*. https://doi.org/10.1038/sdata.2016.44

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J.-B., Yarkoni, T., & Margulies, D. S. (2015). NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00008

Gramer, M., & Saria, K. (2007). Effects of social anxiety and evaluative threat on cardiovascular responses to active performance situations. *Biological Psychology*, *74*(1), 67–74. https://doi.org/10.1016/J.BIOPSYCHO.2006.07.004

Grelotti, D. J., Gauthier, I., & Schultz, R. T. (2002). Social interest and the development of cortical face specialization: What autism teaches us about face processing. *Developmental Psychobiology*, *40*(3), 213–225. https://doi.org/10.1002/dev.10028

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Guastella, A. J., Einfeld, S. L., Gray, K. M., Rinehart, N. J., Tonge, B. J., Lambert, T. J., & Hickie, I. B. (2010). Intranasal Oxytocin Improves Emotion Recognition for Youth with Autism Spectrum Disorders. *Biological Psychiatry*, *67*(7), 692–694. https://doi.org/10.1016/J.BIOPSYCH.2009.09.020

Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Experimental Social Psychology*, 22(1), 38–77. https://doi.org/10.1016/0022-1031(86)90040-5

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, 42, 279–297. https://doi.org/10.1016/J.NEUBIOREV.2014.03.013

Guy, N., Azulay, H., Kardosh, R., Weiss, Y., Hassin, R. R., Israel, S., & Pertzov, Y. (2019). A novel perceptual trait: gaze predilection for faces during visual exploration. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-47110-x

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B. H., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping*, 7(1), 29–37. https://doi.org/10.1002/(SICI)1097-0193(1999)7:1<29::AID-HBM3>3.0.CO;2-R

Hall, C., Hogue, T., & Guo, K. (2011). Differential Gaze Behavior towards Sexually Preferred and Non-Preferred Human Figures. *Journal of Sex Research*, *48*(5), 461–469. https://doi.org/10.1080/00224499.2010.521899

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems 19*, 545–552. https://doi.org/10.7551/mitpress/7503.003.0073

Harel, J., Koch, C., & Perona, P. (2018). Graph-Based Visual Saliency. *Advances in Neural Information Processing Systems 19*. https://doi.org/10.7551/mitpress/7503.003.0073

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Haxby, J. V, Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The Effect of Face Inversion on Activity in Human Neural Systems for Face and Object Perception. *Neuron*, 22(1), 189–199. https://doi.org/10.1016/S0896-6273(00)80690-X

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7), 523–534. https://doi.org/10.1038/nrn1931

Hayward, D. A., Voorhies, W., Morris, J. L., Capozzi, F., & Ristic, J. (2017). Staring reality in the face: A comparison of social attention across laboratory and real world measures suggests little common ground. *Canadian Journal of Experimental Psychology*, *71*(3), 212–225. https://doi.org/10.1037/cep0000117

Hebart, M. N., Görgen, K., & Haynes, J. D. (2015). The decoding toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*(JAN). https://doi.org/10.3389/fninf.2014.00088

Heeren, A., & McNally, R. J. (2016). An integrative network approach to social anxiety disorder: The complex dynamic interplay among attentional bias for threat, attentional control, and symptoms. *Journal of Anxiety Disorders*, 42, 95–104. https://doi.org/10.1016/j.janxdis.2016.06.009

Heimberg, R. G., Brozovich, F. A., & Rapee, R. M. (2010). A Cognitive Behavioral Model of Social Anxiety Disorder: Update and Extension. *Social Anxiety*, 395–422. https://doi.org/10.1016/B978-0-12-375096-9.00015-8

Helminen, T. M., Kaasinen, S. M., & Hietanen, J. K. (2011). Eye contact and arousal: The effects of stimulus duration. *Biological Psychology*, *88*(1), 124–130. https://doi.org/10.1016/J.BIOPSYCHO.2011.07.002

Hendelman, W. (2000). The Amygdala. *Atlas of Functional Neuroanatomy*. https://doi.org/10.1201/9781420048209.supl79a

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye Movements*. https://doi.org/10.1016/B978-008044980-7/50027-6

Heywood, C., & Cowey, A. (2006). The role of the 'face-cell' area in the discrimination and recognition of faces by monkeys. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 31–38. https://doi.org/10.1098/rstb.1992.0004

Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., & Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. *Neuropsychologia*, *46*(9), 2423–2430. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2008.02.029

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, K. L., Gothard, K. M., Schmid, M. C. C., & Logothetis, N. K. (2007). Facial-Expression and Gaze-Selective Responses in the Monkey Amygdala. *Current Biology*, *17*(9), 766–772. https://doi.org/10.1016/j.cub.2007.03.040

Hofmann, S. G. (2007). Cognitive Factors that Maintain Social Anxiety Disorder: a Comprehensive Model and its Treatment Implications. *Cognitive Behaviour Therapy*, *36*(4), 193–209. https://doi.org/10.1080/16506070701421313

Hofmann, S. G., Gerlach, A. L., Wender, A., & Roth, W. T. (1997). Speech Disturbances and Gaze Behavior During Public Speaking in Subtypes of Social Phobia. *Journal of Anxiety Disorders*, *11*(6), 573–585.

Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Sciences*, *3*(2), 65–73. https://doi.org/10.1016/S1364-6613(98)01271-6

Hollander, E., Bartz, J., Chaplin, W., Phillips, A., Sumner, J., Soorya, L., Anagnostou, E., & Wasserman, S. (2007). Oxytocin Increases Retention of Social Cognition in Autism. *Biological Psychiatry*, *61*(4), 498–503. https://doi.org/10.1016/J.BIOPSYCH.2006.05.030

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's Eyes Trigger Shifts of Visual Attention in Human Infants. *Psychological Science*, *9*(2), 131–134. https://doi.org/10.1111/1467-9280.00024

Hooker, C. I., Paller, K. A., Gitelman, D. R., Parrish, T. B., Mesulam, M.-M., & Reber, P. J. (2003). Brain networks for analyzing eye gaze. *Cognitive Brain Research*, *17*(2), 406–418. https://doi.org/10.1016/S0926-6410(03)00143-5

Hopp, J. J., & Fuchs, A. F. (2004). The characteristics and neuronal substrate of saccadic eye movement plasticity. *Progress in Neurobiology*, *72*(1), 27–53. https://doi.org/10.1016/j.pneurobio.2003.12.002

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not see eye to eye:: A visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders*, *17*(1), 33–44. https://doi.org/10.1016/S0887-6185(02)00180-9

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2004). Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia. *Psychiatry Research*, *127*(1–2), 43–53. https://doi.org/10.1016/J.PSYCHRES.2004.02.016

Hosozawa, M., Tanaka, K., Shimizu, T., Nakano, T., & Kitazawa, S. (2012). How children with specific language impairment view social situations: an eye tracking study. *Pediatrics*, *129*(6), e1453-60. https://doi.org/10.1542/peds.2011-2278

Howell, A. N., Zibulsky, D. A., Srivastav, A., & Weeks, J. W. (2016). Relations among Social Anxiety, Eye

Contact Avoidance, State Anxiety, and Perception of Interaction Performance during a Live Conversation. *Cognitive Behaviour Therapy*, *45*(2), 111–122. https://doi.org/10.1080/16506073.2015.1111932

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*(5), 25–25. https://doi.org/10.1167/9.5.25

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. In *American Journal of Psychiatry*, 167(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia*, *45*(5), 1019–1028. https://doi.org/10.1016/j.neuropsychologia.2006.09.004

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Itti, Laurent, & Baldi, P. (2005). Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems*.

Itti, Laurent, & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, Laurent, & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203. https://doi.org/10.1038/35058500

Itti, Laurent, Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell*, *20*(11), 1254–1259.

Izuma, K. (2012). The social neuroscience of reputation. *Neuroscience research*, 72(4), 283-288.

Izuma, K., Matsumoto, K., Camerer, C. F., & Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(42), 17302–17307. https://doi.org/10.1073/pnas.1107038108

Jacques, C., d'Arripe, O., & Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. *Journal of Vision*, 7(8), 3. https://doi.org/10.1167/7.8.3

Jamieson, J. P., Nock, M. K., & Mendes, W. B. (2013). Changing the Conceptualization of Stress in Social Anxiety Disorder. *Clinical Psychological Science*, *1*(4), 363–374. https://doi.org/10.1177/2167702613482119

Jeffreys, D. A. (1996). Evoked Potential Studies of Face and Object Processing. *Visual Cognition*, *3*(1), 1–38. https://doi.org/10.1080/713756729

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H., & Curtis, C. E. (2012). Prioritized maps of space in human frontoparietal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(48), 17382–17390. https://doi.org/10.1523/JNEUROSCI.3810-12.2012

John, O. P., Donahue, E., & Kentle, R. (1991). *The Big Five Inventory*.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, *6*(10), 766–774. https://doi.org/10.1038/nrn1766

Johnson, M. H. (2014). Autism: Demise of the Innate Social Orienting Hypothesis. *Current Biology*, *24*(1), R30–R31. https://doi.org/10.1016/J.CUB.2013.11.021

Johnstone, T., Somerville, L. H., Alexander, A. L., Oakes, T. R., Davidson, R. J., Kalin, N. H., & Whalen, P. J. (2005). *Stability of amygdala BOLD response to fearful faces over multiple scan sessions*. https://doi.org/10.1016/j.neuroimage.2004.12.016

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience*, *29*(19), 6234–6238. https://doi.org/10.1523/JNEUROSCI.5570-08.2009

Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Control of attention and gaze in complex environments. *Journal of Vision*, *6*(12), 9. https://doi.org/10.1167/6.12.9

Judd, T., Durand, F., & Torralba, A. (2012). A Benchmark of Computational Models of Saliency to Predict Human Fixations A Benchmark of Computational Models of Saliency to Predict Human Fixations. *MIT Technical Report*.

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

Kanner, L. (1943). Autistic Disturbances of Affective Contact. *Nervous Child*, *2*(3), 217–250.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kaping, D., Vinck, M., Hutchison, R. M., Everling, S., & Womelsdorf, T. (2011). Specific Contributions of Ventromedial, Anterior Cingulate, and Lateral Prefrontal Cortex for Attentional Selection and Stimulus Valuation. *PLoS Biology*, *9*(12), e1001224. https://doi.org/10.1371/journal.pbio.1001224

Kapp, B. S., Supple, W. F., & Whalen, P. J. (1994). Effects of electrical stimulation of the amygdaloid central nucleus on neocortical arousal in the rabbit. *Behavioral Neuroscience*, *108*(1), 81.

Karnath, H.-O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, *411*(6840), 950–953. https://doi.org/10.1038/35082075

Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Ito, K., Fukuda, H., Kojima, S., & Nakamura, K. (1999). The human amygdala plays an important role in gaze monitoring. *Brain*, *122*(4), 779–783. https://doi.org/10.1093/brain/122.4.779

Kendall, L. N., Raffaelli, Q., Kingstone, A., & Todd, R. M. (2016). Iconic faces are not real faces: enhanced emotion detection and altered neural processing as faces become more iconic. *Cognitive Research: Principles and Implications*, *1*(1), 19. https://doi.org/10.1186/s41235-016-0021-8

Kendler, K. S. (2012). The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Molecular Psychiatry* , 17(4), 377–388. https://doi.org/10.1038/mp.2011.182

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*(6), 1143–1150. https://doi.org/10.1017/S0033291710001844

Kennedy, D. P., & Adolphs, R. (2010). Impaired fixation to eyes following amygdala damage arises from abnormal bottom-up attention. *Neuropsychologia*, *48*(12), 3392–3398. https://doi.org/10.1016/j.neuropsychologia.2010.06.025

Kennerley, S. W., & Wallis, J. D. (2009). Reward-dependent modulation of working memory in lateral prefrontal cortex. *The Journal of Neuroscience, 29*(10), 3259–3270. https://doi.org/10.1523/JNEUROSCI.5353-08.2009

Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E., Wang, P., Wells, K. B., & Zaslavsky, A. M. (2005). Prevalence and Treatment of Mental Disorders, 1990 to 2003. *New England Journal of Medicine*, *352*(24), 2515–2523. https://doi.org/10.1056/NEJMsa043266

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology*, *19*(1), 52–56.

https://doi.org/10.1016/J.CONB.2009.05.004

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. https://doi.org/10.1348/000712607X251243

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, Researchers! It Is Time to Take a Look at the Real World. *Current Directions in Psychological Science*, *12*(5), 176–180. https://doi.org/10.1111/1467-8721.01255

Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: An fMRI investigation. *Brain and Cognition*, *55*(2), 269–271. https://doi.org/10.1016/J.BANDC.2004.02.037

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147. https://doi.org/10.1016/S1364-6613(00)01452-2

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, S14. https://doi.org/10.1068/v070821

Kleinhans, N. M., Richards, T., Sterling, L., Stegbauer, K. C., Mahurin, R., Johnson, L. C., Greenson, J., Dawson, G., & Aylward, E. (2008). Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain*, *131*(4), 1000–1012. https://doi.org/10.1093/brain/awm334

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1430), 345–360.

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

Koch, C., & Ullman, S. (1987). Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. In *Matters of Intelligence* (pp. 115–141). Springer Netherlands. https://doi.org/10.1007/978-94-009-3833-5_5

Kohls, G., Schulte-Rüther, M., Nehrkorn, B., Müller, K., Fink, G. R., Kamp-Becker, I., Herpertz-Dahlmann, B., Schultz, R. T., & Konrad, K. (2013). Reward system dysfunction in autism spectrum disorders. *Social Cognitive and Affective Neuroscience*, *8*(5), 565–572. https://doi.org/10.1093/scan/nss033

Kootstra, G., de Boer, B., & Schomaker, L. R. B. (2011). Predicting Eye Fixations on Complex Visual Stimuli Using Local Symmetry. *Cognitive Computation*, *3*(1), 223–240. https://doi.org/10.1007/s12559-010-9089-5

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/J.TICS.2013.06.007

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Krolak-Salmon, P., Hénaff, M.-A., Vighetto, A., Bertrand, O., & Mauguière, F. (2004). Early Amygdala Reaction to Fear Spreading in Occipital, Temporal, and Frontal Cortex: A Depth Electrode ERP Study in Human. *Neuron*, *42*(4), 665–676. https://doi.org/10.1016/S0896-6273(04)00264-8

Kullback, S. (1960). Information Theory and Statistics. In *SIAM Review, 2*(2). https://doi.org/10.1137/1002033

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5548–5553. https://doi.org/10.1073/pnas.1017022108

Lalljee, M. (1978). The role of gaze in the expression of emotion. *Australian Journal of Psychology*, *30*(1), 59–67. https://doi.org/10.1080/00049537808256040

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*, 39-58.

Lang, P. J., & McTeague, L. M. (2009). The anxiety disorder spectrum: Fear imagery, physiological reactivity, and differential diagnosis. *Anxiety, Stress & Coping*, *22*(1), 5–25. https://doi.org/10.1080/10615800802478247

Lange, W.-G., Heuer, K., Langner, O., Keijsers, G. P. J., Becker, E. S., & Rinck, M. (2011). Face value: Eye movements and the evaluation of facial crowds in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(3), 355–363. https://doi.org/10.1016/J.JBTEP.2011.02.007

Langer, J. K., Lim, M. H., Fernandez, K. C., & Rodebaugh, T. L. (2017). Social Anxiety Disorder is Associated with Reduced Eye Contact During Conversation Primed for Conflict. *Cognitive Therapy and Research*, *41*(2), 220–229. https://doi.org/10.1007/s10608-016-9813-x

Langton, S. R. H., & Bruce, V. (1999). Reflexive Visual Orienting in Response to the Social Attention of Others. *Visual Cognition*, *6*(5), 541–567. https://doi.org/10.1080/135062899394939

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/10.1016/S1364-6613(99)01436-9

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). Das State-Trait-Anxiety-Inventory (STAI).

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments.

Lazarov, A., Abend, R., & Bar-Haim, Y. (2016). Social anxiety is related to increased dwell time on socially threatening faces. *Journal of Affective Disorders*, *193*, 282–288. https://doi.org/10.1016/J.JAD.2016.01.007

LeDoux, J. (1998). The emotional brain: The mysterious underpinnings of emotional life. Simon and Schuster.

Leekam, S. R., López, B., & Moore, C. (2000). Attention and joint attention in preschool children with autism. *Developmental Psychology*, *36*(2), 261–273. https://doi.org/10.1037/0012-1649.36.2.261

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94. https://doi.org/10.1038/82947

Levy, J., Foulsham, T., & Kingstone, A. (2013). Monsters are people too. *Biology Letters*, *9*(1), 20120850. https://doi.org/10.1098/rsbl.2012.0850

Liang, C.-W., Tsai, J.-L., & Hsu, W.-Y. (2017). Sustained visual attention for competing emotional stimuli in social anxiety: An eye tracking study. *Journal of Behavior Therapy and Experimental Psychiatry*, *54*, 178–185. https://doi.org/10.1016/J.JBTEP.2016.08.009

Lipp, I., Murphy, K., Wise, R. G., & Caseras, X. (2014). Understanding the contribution of neural and physiological signal variation to the low repeatability of emotion-induced BOLD responses. *NeuroImage*, *86*, 335–342. https://doi.org/10.1016/j.neuroimage.2013.10.015

Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, *8*(2), 99–117. https://doi.org/10.1037/1076-898X.8.2.99

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism and Developmental Disorders*, *16*(3), 335–349. https://doi.org/10.1007/BF01531663

Lyon, L. (2017). Dead salmon and voodoo correlations: should we be sceptical about functional MRI? *Brain*, *140*(8), e53–e53. https://doi.org/10.1093/brain/awx180

Mack, A., Pappas, Z., Silverman, M., & Gay, R. (2002). What we see: Inattention and the capture of attention by meaning. *Consciousness and Cognition*, *11*(4), 488–506.

https://doi.org/10.1016/S1053-8100(02)00028-4

Mackay, M., Cerf, M., & Koch, C. (2012). Evidence for two distinct mechanisms directing gaze in natural scenes. *Journal of Vision*, *12*(4), 1–12. https://doi.org/10.1167/12.4.9

Maeda, K., Inoue, K., Kunimatsu, J., Takada, M., & Hikosaka, O. (2019). Amygdala controls saccade and gaze physically, motivationally, and socially. *BioRxiv*, 608703. https://doi.org/10.1101/608703

Manesi, Z., Van Lange, P. A. M., & Pollet, T. V. (2016). Eyes Wide Open. *Evolutionary Psychology*, *14*(2), 147470491664078. https://doi.org/10.1177/1474704916640780

Mansell, W., Clark, D. M., Ehlers, A., & Chen, Y.-P. (1999). Social Anxiety and Attention away from Emotional Faces. *Cognition & Emotion*, *13*(6), 673–690. https://doi.org/10.1080/026999399379032

Mansfield, E., Farroni, T., & Johnson, M. (2003). Does gaze perception facilitate overt orienting? *Visual Cognition*, *10*(1), 7–14. https://doi.org/10.1080/713756671

Marat, S., Rahman, A., Pellerin, D., Guyader, N., & Houzet, D. (2013). Improving Visual Saliency by Adding 'Face Feature Map' and 'Center Bias.' Cognitive Computation, 5(1), 63–75. https://doi.org/10.1007/s12559-012-9146-3

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610. https://doi.org/10.3758/s13428-013-0379-1

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., & Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *66*(2), 026702. https://doi.org/10.1103/PhysRevE.66.026702

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6

Maunsell, J. H. R. (2004). Neuronal representations of cognitive state: reward or attention? *Trends in Cognitive Sciences*, *8*(6), 261–265. https://doi.org/10.1016/J.TICS.2004.04.003

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6). https://doi.org/10.1016/S1364-6613(02)01903-4

Mauss, I. B., Wilhelm, F. H., & Gross, J. J. (2003). Autonomic recovery and habituation in social anxiety. *Psychophysiology*, *40*(4), 648–653. https://doi.org/10.1111/1469-8986.00066

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-Specific Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610. https://doi.org/10.1162/jocn.1997.9.5.605

McCulloch, C. E., & Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science*, *26*(3), 388–402. https://doi.org/10.1214/11-STS361

McFadyen, J., Mermillod, M., Mattingley, J. B., Halász, V., & Garrido, M. I. (2017). A Rapid Subcortical Amygdala Route for Faces Irrespective of Spatial Frequency and Emotion. *The Journal of Neuroscience*, *37*(14), 3864–3874. https://doi.org/10.1523/JNEUROSCI.3525-16.2017

McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, *45*(8), 1027–1035. https://doi.org/10.1016/J.JPSYCHIRES.2011.03.006

McManus, F., Sacadura, C., & Clark, D. M. (2008). Why social anxiety persists: An experimental investigation of the role of safety behaviours as a maintaining factor. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(2), 147–161. https://doi.org/10.1016/J.JBTEP.2006.12.002

McPartland, J., Dawson, G., Webb, S. J., Panagiotides, H., & Carver, L. J. (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *45*(7), 1235–1245. https://doi.org/10.1111/j.1469-7610.2004.00318.x

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., Vuilleumier, P., Gil-Nagel, A., & Strange, B. A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience*, *19*(8), 1041–1049. https://doi.org/10.1038/nn.4324

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime Prevalence of Mental Disorders in U.S. Adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980–989. https://doi.org/10.1016/J.JAAC.2010.05.017

Merin, N., Young, G. S., Ozonoff, S., & Rogers, S. J. (2007). Visual Fixation Patterns during Reciprocal Social Interaction Distinguish a Subgroup of 6-Month-Old Infants At-Risk for Autism from Comparison Infants. *Journal of Autism and Developmental Disorders*, *37*(1), 108–121. https://doi.org/10.1007/s10803-006-0342-4

Mesulam, M.-M., Mufson, E. J., Levey, A. I., & Wainer, B. H. (1983). Cholinergic innervation of cortex by the basal forebrain: Cytochemistry and cortical connections of the septal area, diagonal band nuclei, nucleus basalis (Substantia innominata), and hypothalamus in the rhesus monkey. *The Journal of Comparative Neurology*, *214*(2), 170–197. https://doi.org/10.1002/cne.902140206

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417. https://doi.org/10.1016/0166-2236(83)90190-X

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

Mo, C., He, D., & Fang, F. (2018). Attention priority map of face images in human early visual cortex. *Journal of Neuroscience*, *38*(1), 149–157. https://doi.org/10.1523/JNEUROSCI.1206-17.2017

Mogg, K., Bradley, B. P., De Bono, J., & Painter, M. (1997). Time course of attentional bias for threat information in non-clinical anxiety. *Behaviour research and therapy*, *35*(4), 297-303.

Mogg, K., Bradley, B. P., Miles, F., & Dixon, R. (2004). Time course of attentional bias for threat scenes: Testing the vigilance-avoidance hypothesis. *Cognition and Emotion*, *18*(5), 689–700. https://doi.org/10.1080/02699930341000158

Mogg, K., Philippot, P., & Bradley, B. P. (2004). Selective attention to angry faces in clinical social phobia. *Journal of abnormal psychology*, *113*(1), 160.

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, *21*(2), 373–377. https://doi.org/10.1016/S0163-6383(98)90014-5

Morris, J. S., DeGelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain*, *124*(6), 1241–1252. https://doi.org/10.1093/brain/124.6.1241

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*(6684), 467–470. https://doi.org/10.1038/30976

Moukheiber, A., Rautureau, G., Perez-Diaz, F., Jouvent, R., & Pelissolo, A. (2012). Gaze behaviour in social blushers. *Psychiatry Research*, *200*(2–3), 614–619. https://doi.org/10.1016/J.PSYCHRES.2012.07.017

Moukheiber, A., Rautureau, G., Perez-Diaz, F., Soussignan, R., Dubal, S., Jouvent, R., & Pelissolo, A. (2010). Gaze avoidance in social phobia: Objective measure and correlates. *Behaviour Research and Therapy*, *48*(2), 147–151. https://doi.org/10.1016/J.BRAT.2009.09.012

Movellan, J. R., & Watson, J. S. (2002). The development of gaze following as a Bayesian systems identification problem. *Proceedings - 2nd International Conference on Development and Learning, ICDL 2002*, 34–40. https://doi.org/10.1109/DEVLRN.2002.1011728

Müller-Pinzler, L., Gazzola, V., Keysers, C., Sommer, J., Jansen, A., Frässle, S., Einhäuser, W., Paulus, F. M., & Krach, S. (2015). Neural pathways of embarrassment and their modulation by social anxiety. *NeuroImage*, *119*, 252–261. https://doi.org/10.1016/j.neuroimage.2015.06.036

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental psychology: Human perception and performance*, *15*(2), 315.

Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Mundy, P., & Rebecca Neal, A. (2000). Neural plasticity, joint attention, and a transactional social-orienting model of autism. *International Review of Research in Mental Retardation*, *23*, 139–168. https://doi.org/10.1016/S0074-7750(00)80009-9

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry*, *27*(5), 657–669. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x

Murphy, J. E., Yanes, J. A., Kirby, L. A. J., Reid, M. A., & Robinson, J. L. (2020). Left, right, or bilateral amygdala activation? How effects of smoothing and motion correction on ultra-high field, high-resolution functional magnetic resonance imaging (fMRI) data alter inferences. *Neuroscience Research*, *150*, 51–59. https://doi.org/10.1016/j.neures.2019.01.009

Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Nonverbal leakage in robots. *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction - HRI ’09*, 69. https://doi.org/10.1145/1514095.1514110

Myllyneva, A., Ranta, K., & Hietanen, J. K. (2015). Psychophysiological responses to eye contact in adolescents with social anxiety disorder. *Biological Psychology*, *109*, 151–158. https://doi.org/10.1016/J.BIOPSYCHO.2015.05.005

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., Ohta, H., Kato, N., & Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences*, 277(1696), 2935–2943. https://doi.org/10.1098/rspb.2010.0587

Nardo, D., Santangelo, V., & Macaluso, E. (2014). Spatial orienting in complex audiovisual environments. *Human Brain Mapping*, *35*(4), 1597–1614. https://doi.org/10.1002/hbm.22276

Nasiopoulos, E., Risko, E. F., Foulsham, T., & Kingstone, A. (2015). Wearable computing: Will it make people prosocial? *British Journal of Psychology*, *106*(2). https://doi.org/10.1111/bjop.12080

Nasiopoulos, E., Risko, E. F., & Kingstone, A. (2015). Social Attention, Social Presence, and the Dual Function of Gaze. In *The Many Faces of Social Attention* (pp. 129–155). Springer International Publishing. https://doi.org/10.1007/978-3-319-21368-2_5

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

Carleton, R. N., McCreary, D. R., Norton, P. J., & Asmundson, G. J. (2006). Brief fear of negative evaluation scale—revised. *Depression and anxiety*, *23*(5), 297-303.

Noonan, M. P., Sallet, J., Mars, R. B., Neubert, F. X., O’Reilly, J. X., Andersson, J. L., Mitchell, A. S., Bell, A. H., Miller, K. L., & Rushworth, M. F. S. (2014). A Neural Circuit Covarying with Social Hierarchy in Macaques. *PLoS Biology*, *12*(9), e1001940. https://doi.org/10.1371/journal.pbio.1001940

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

Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences*, *1339*(1), 82–96. https://doi.org/10.1111/nyas.12705

Nyström, M., & Holmqvist, K. (2008). Semantic Override of Low-level Features in Image Viewing – Both Initially and Overall. *Journal of Eye Movement Research*, 2(2), 1–11. https://doi.org/10.16910/jemr.2.2.2

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*(6), 261–266. https://doi.org/10.1016/S1364-6613(02)01908-3

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, *30*(10), 953–958. https://doi.org/10.1016/j.psyneuen.2005.03.019

Olmos, A., & Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, *33*(12), 1463–1473. https://doi.org/10.1068/p5321

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders*, *24*(3), 247–257. https://doi.org/10.1007/BF02172225

Oya, H., Kawasaki, H., Howard, M. A., & Adolphs, R. (2002). Electrophysiological responses in the human amygdala discriminate emotion categories of complex visual stimuli. *The Journal of Neuroscience*, *22*(21), 9502–9512. https://doi.org/10.1523/JNEUROSCI.22-21-09502.2002

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226. https://doi.org/10.1038/nature04676

Panayiotou, G., Karekla, M., Georgiou, D., Constantinou, E., & Paraskeva-Siamata, M. (2017). Psychophysiological and self-reported reactivity associated with social anxiety and public speaking fear symptoms: Effects of fear versus distress. *Psychiatry Research*, *255*, 278–286. https://doi.org/10.1016/J.PSYCHRES.2017.05.044

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123. https://doi.org/10.1016/S0042-6989(01)00250-4

Parkington, K. B., & Itier, R. J. (2018). One versus two eyes makes a difference! Early face perception is modulated by featural fixation and feature context. *Cortex*, *109*, 35–49. https://doi.org/10.1016/j.cortex.2018.08.025

Parks, D., Borji, A., & Itti, L. (2015). Augmented saliency model using automatic 3D head pose detection and learned gaze following in natural scenes. *Vision Research*, *116*, 113–126. https://doi.org/10.1016/J.VISRES.2014.10.027

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, *439*(7078), 865–870. https://doi.org/10.1038/nature04490

Patrick, C. J., & Hajcak, G. (2016). RDoC: Translating promise into progress. *Psychophysiology*, *53*(3), 415–424. https://doi.org/10.1111/psyp.12612

Patterson, M. L., & Ritts, V. (1997). Social and Communicative Anxiety: A Review and Meta-Analysis. *Annals of the International Communication Association*, *20*(1), 263–303. https://doi.org/10.1080/23808985.1997.11678944

Peck, C. J., Lau, B., & Salzman, C. D. (2013). The primate amygdala combines information about space and value. *Nature Neuroscience*, *16*(3), 340–348. https://doi.org/10.1038/nn.3328

Peck, C. J., & Salzman, C. D. (2014). The amygdala and basal forebrain as a pathway for motivationally guided attention. *Journal of Neuroscience*, *34*(41), 13757-13767.

Pegna, A. J., Khateb, A., Lazeyras, F., & Seghier, M. L. (2005). Discriminating emotional faces without primary visual cortices involves the right amygdala. *Nature Neuroscience*, *8*(1), 24–25. https://doi.org/10.1038/nn1364

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When Strangers Pass. *Psychological Science*, *15*(9), 598–603. https://doi.org/10.1111/j.0956-7976.2004.00726.x

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *335*(1273). https://doi.org/10.1098/rstb.1992.0003

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London - Biological Sciences*, 223(1232) , 293-317. https://doi.org/10.1098/rspb.1985.0003

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a “low road” to “many roads” of evaluating biological significance. *Nature Reviews. Neuroscience*, *11*(11), 773–783. https://doi.org/10.1038/nrn2920

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397–2416. https://doi.org/10.1016/j.visres.2005.03.019

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *48*(2), 175–187. https://doi.org/10.1016/J.NEURON.2005.09.025

Pittig, A., Arch, J. J., Lam, C. W. R., & Craske, M. G. (2013). Heart rate and heart rate variability in panic, social anxiety, obsessive–compulsive, and generalized anxiety disorders at baseline and in response to relaxation and hyperventilation. *International Journal of Psychophysiology*, *87*(1), 19–27. https://doi.org/10.1016/J.IJPSYCHO.2012.10.012

Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A. B. M., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., & Meyer-Lindenberg, A. (2012). Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *NeuroImage*, *60*(3), 1746–1758. https://doi.org/10.1016/j.neuroimage.2012.01.129

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J. B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Poldrack, R. A., Fletcher, P. C., Henson, R. N., Worsley, K. J., Brett, M., & Nichols, T. E. (2008). Guidelines for reporting an fMRI study. *NeuroImage*, *40*(2), 409–414. https://doi.org/10.1016/j.neuroimage.2007.11.048

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Powell, K. L., Roberts, G., & Nettle, D. (2012). Eye Images Increase Charitable Donations: Evidence From an Opportunistic Field Experiment in a Supermarket. *Ethology*, *118*(11), 1096–1101. https://doi.org/10.1111/eth.12011

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Preston, T. J., Guo, F., Das, K., Giesbrecht, B., & Eckstein, M. P. (2013). *Neural Representations of Contextual Guidance in Visual Search of Real-World Scenes. Journal of Neuroscience*, *33*(18), 7846-7855. https://doi.org/10.1523/JNEUROSCI.5840-12.2013

Puigcerver, A., Martinez-Selva, J. M., Garcia-Sanchez, F. A., & Gomez-Amor, J. (1989). Individual differences in psychophysiological and subjective correlates of speech anxiety. *Journal of Psychophysiology*, *3*(1), 75–81.

Putnam, P. T., & Gothard, K. M. (2019). Multidimensional neural selectivity in the primate amygdala. *ENeuro*, *6*(5). https://doi.org/10.1523/ENEURO.0153-19.2019

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K): *Diagnostica*, *51*(4), 195–206. https://doi.org/10.1026/0012-1924.51.4.195

Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, *35*(8), 741–756. https://doi.org/10.1016/S0005-7967(97)00022-3

Rauthmann, J. F., Seubert, C. T., Sachse, P., & Furtner, M. R. (2012). Eyes as windows to the soul: Gazing behavior is related to personality. *Journal of Research in Personality*, *46*(2), 147–156. https://doi.org/10.1016/j.jrp.2011.12.010

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Reeves, A., & McLellan, J. S. (2020). The “anti-shift”: Shifting attention opposite to a saccade. *Vision Research*, *167*, 31–38. https://doi.org/10.1016/j.visres.2019.06.009

Reichenberger, J., Schwarz, M., König, D., Wilhelm, F. H., Voderholzer, U., Hillert, A., & Blechert, J. (2016). Angst vor negativer sozialer Bewertung: Übersetzung und Validierung der Furcht vor negativer Evaluation–Kurzskala (FNE-K). *Diagnostica*, *62*(3), 169–181. https://doi.org/10.1026/0012-1924/a000148

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The Human Fetus Preferentially Engages with Face-like Visual Stimuli. *Current Biology*, 27(12), 1825-1828.e3. https://doi.org/10.1016/J.CUB.2017.05.044

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*(3), 703–714. https://doi.org/10.1016/S0896-6273(00)81206-4

Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., & Clifford, C. W. (2005). Adaptation and face perception: How aftereffects implicate norm-based coding of faces. *Fitting the mind to the world: Adaptation and after-effects in high-level vision*, 213-240.

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

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual’s gaze. *NeuroReport*, *13*(17), 2259–2264. https://doi.org/10.1097/00001756-200212030-00018

Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, *73*(2), 291–296. https://doi.org/10.3758/s13414-010-0042-1

Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6*, 143. https://doi.org/10.3389/fnhum.2012.00143

Risko, E. F., Richardson, D., & Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science : Real World Social Attention and the Dual Function of Gaze. https://doi.org/10.14288/1.0305737

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

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, 24(3), 715-718.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1 PART 1), 31–40. https://doi.org/10.1016/0028-3932(87)90041-8

Ro, T., Russell, C., & Lavie, N. (2001). Changing Faces: A Detection Advantage in the Flicker Paradigm. *Psychological Science*, *12*(1), 94–99. https://doi.org/10.1111/1467-9280.00317

Rodríguez-Liñares, L., Vila, X., Méndez, A. J., Lado, M. J., & Olivieri, D. (2008). RHRV: An R-based software package for heart rate variability analysis of ECG recordings. In *Iberian Conference on Information Systems and Technologies, CISTI*. http://www.mathworks.com

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Ed.), *The handbook of research synthesis.* (pp. 231–244). Russell Sage Foundation.

Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLOS ONE*, *12*(7), e0182037. https://doi.org/10.1371/journal.pone.0182037

Rosset, D. B., Rondan, C., Da Fonseca, D., Santos, A., Assouline, B., & Deruelle, C. (2008). Typical Emotion Processing for Cartoon but not for Real Faces in Children with Autistic Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *38*(5), 919–925. https://doi.org/10.1007/s10803-007-0465-2

Rosset, D. B., Santos, A., Da Fonseca, D., Poinso, F., O'Connor, K., & Deruelle, C. (2010). Do children perceive features of real and cartoon faces in the same way? Evidence from typical development and autism. *Journal of Clinical and Experimental Neuropsychology*, *32*(2), 212–218. https://doi.org/10.1080/13803390902971123

Rossi, A., Parada, F. J., Latinus, M., & Puce, A. (2015). Photographic but not line-drawn faces show early perceptual neural sensitivity to eye gaze direction. *Frontiers in Human Neuroscience*, *9*(185). https://doi.org/10.3389/fnhum.2015.00185

Rossion, B., Curran, T., & Gauthier, I. (2002). A defense of the subordinate-level expertise account for the N170 component. *Cognition*, *85*(2), 189–196. https://doi.org/10.1016/S0010-0277(02)00101-4

Rubo, M., & Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Scientific Reports*, *8*(1), 3804. https://doi.org/10.1038/s41598-018-22127-w

Rubo, M., Huestegge, L., & Gamer, M. (2019). Social anxiety modulates visual exploration in real life – but not in the laboratory. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12396

Russchen, F. T., Amaral, D. G., & Price, J. L. (1985). The afferent connections of the substantia innominata in the monkey,Macaca fascicularis. *The Journal of Comparative Neurology*, *242*(1), 1–27. https://doi.org/10.1002/cne.902420102

Rutishauser, U., Mamelak, A. N., & Adolphs, R. (2015). The primate amygdala in social perception – insights from electrophysiological recordings and stimulation. *Trends in Neurosciences*, *38*(5), 295–306. https://doi.org/10.1016/J.TINS.2015.03.001

Sagiv, N., & Bentin, S. (2001). Structural Encoding of Human and Schematic Faces: Holistic and Part-Based Processes. *Journal of Cognitive Neuroscience*, *13*(7), 937–951. https://doi.org/10.1162/089892901753165854

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. K., O'Reilly, J. X. O., Jbabdi, S., Croxson, P. L., Jenkinson, M., Miller, K. L., & Rushworth, M. F. S. (2011). Social Network Size Affects Neural Circuits in Macaques. *Science*, *334*(6056), 697–700. https://doi.org/10.1126/science.1210027

Sanghera, M. ., Rolls, E. ., & Roper-Hall, A. (1979). Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Experimental Neurology*, *63*(3), 610–626. https://doi.org/10.1016/0014-4886(79)90175-4

Santangelo, V. (2015). Forced to remember: When memory is biased by salient information. *Behavioural Brain Research*, 283, 1–10. https://doi.org/10.1016/j.bbr.2015.01.013

Santangelo, V., di Francesco, S. A., Mastroberardino, S., & Macaluso, E. (2015). Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory. *Human Brain Mapping*, *36*(12), 5003–5017. https://doi.org/10.1002/hbm.22984

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. https://doi.org/10.1038/253265a0

Scheel, A. M., Ritchie, S. J., Brown, N. J. L., & Jacques, S. L. (2018). Methodological problems in a study of fetal visual perception. *Current Biology*, *28*(10), R594–R596. https://doi.org/10.1016/J.CUB.2018.03.047

Scheller, E., Büchel, C., & Gamer, M. (2012). Diagnostic Features of Emotional Expressions Are Processed Preferentially. *PLoS ONE*, 7(7). https://doi.org/10.1371/journal.pone.0041792

Schofield, C. A., Inhoff, A. W., & Coles, M. E. (2013). Time-course of attention biases in social phobia. *Journal of Anxiety Disorders*, 27(7), 661–669. https://doi.org/10.1016/J.JANXDIS.2013.07.006

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 of London. Series B: Biological Sciences*, *358*(1430), 415–427. https://doi.org/10.1098/rstb.2002.1208

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D. G., Poldrack, R. A., & Bookheimer, S. Y. (2010). Reward processing in autism. *Autism Research*, *3*(2), https://doi.org/10.1002/aur.122

Seefeldt, W. L., Krämer, M., Tuschen-Caffier, B., & Heinrichs, N. (2014). Hypervigilance and avoidance in visual attention in children with social phobia. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(1), 105–112. https://doi.org/10.1016/J.JBTEP.2013.09.004

Senju, A., & Johnson, M. H. (2009). Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1204–1214. https://doi.org/10.1016/J.NEUBIOREV.2009.06.001

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*(1), 38–45. https://doi.org/10.1016/j.tics.2005.11.008

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *32*(4), 811–830. https://doi.org/10.1016/j.neubiorev.2007.12.002

Shechner, T., Jarcho, J. M., Britton, J. C., Leibenluft, E., Pine, D. S., & Nelson, E. E. (2013). Attention bias of anxious youth during extended exposure of emotional face Pairs: An eye-tracking study. *Depression and Anxiety*, *30*(1), 14–21. https://doi.org/10.1002/da.21986

Shechner, T., Jarcho, J. M., Wong, S., Leibenluft, E., Pine, D. S., & Nelson, E. E. (2017). Threats, rewards, and attention deployment in anxious youth and adults: An eye tracking study. *Biological Psychology*, *122*, 121–129. https://doi.org/10.1016/J.BIOPSYCHO.2015.10.004

Shelley-Tremblay, J., & Mack, A. (1999). Metacontrast Masking and Attention. *Psychological Science*, *10*(6), 508–515. https://doi.org/10.1111/1467-9280.00197

Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, 4(5). https://doi.org/10.3389/fnint.2010.00005

Shic, F., Bradshaw, J., Klin, A., Scassellati, B., & Chawarska, K. (2011). Limited activity monitoring in toddlers with autism spectrum disorder. *Brain research*, *1380*, 246-254.

Shic, F., Macari, S., & Chawarska, K. (2014). Speech Disturbs Face Scanning in 6-Month-Old Infants Who Develop Autism Spectrum Disorder. *Biological Psychiatry*, 75(3), 231–237. https://doi.org/10.1016/J.BIOPSYCH.2013.07.009

Shimizu, M., Seery, M. D., Weisbuch, M., & Lupien, S. P. (2011). Trait Social Anxiety and Physiological Activation: Cardiovascular Threat During Social Interaction. *Personality and Social Psychology Bulletin*, *37*(1), 94–106. https://doi.org/10.1177/0146167210391674

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. https://doi.org/10.1038/nn1150

Shioiri, S., & Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception*, *18*(3), 347–361. https://doi.org/10.1068/p180347

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in*

Cognitive Sciences, *13*(11), 488–495. https://doi.org/10.1016/J.TICS.2009.08.005

Silverstein, D. N., & Ingvar, M. (2015). A multi-pathway hypothesis for human visual fear signaling. *Frontiers in Systems Neuroscience*, *9*, 101. https://doi.org/10.3389/fnsys.2015.00101

Singh, J. S., Capozzoli, M. C., Dodd, M. D., & Hope, D. A. (2015). The Effects of Social Anxiety and State Anxiety on Visual Attention: Testing the Vigilance–Avoidance Hypothesis. *Cognitive Behaviour Therapy*, *44*(5), 377–388. https://doi.org/10.1080/16506073.2015.1016447

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/10.1016/J.BRAINRES.2005.12.090

Smit, A. C., Van Gisbergen, J. A. M., & Cools, A. R. (1987). A parametric analysis of human saccades in different experimental paradigms. *Vision Research*, 27(10), 1745–1762. https://doi.org/10.1016/0042-6989(87)90104-0

Smith, E. A., & Bliege Bird, R. L. (2000). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and Human Behavior*, *21*(4), 245–261. https://doi.org/10.1016/S1090-5138(00)00031-3

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science*, *16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801.x

Speer, L. L., Cook, A. E., McMahon, W. M., & Clark, E. (2007). Face processing in children with autism. *Autism*, *11*(3), 265–277. https://doi.org/10.1177/1362361307076925

Sperry, S. H., Kwapil, T. R., Eddington, K. M., & Silvia, P. J. (2018). Psychopathology, everyday behaviors, and autonomic activity in daily life: An ambulatory impedance cardiography study of depression, anxiety, and hypomanic traits. *International Journal of Psychophysiology*, *129*, 67–75. https://doi.org/10.1016/j.ijpsycho.2018.04.008

Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007). Abnormal Use of Facial Information in High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *37*(5), 929–939. https://doi.org/10.1007/s10803-006-0232-9

Spielberger, C. D. (1983). State-trait anxiety inventory for adults.

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–1887. https://doi.org/10.1038/nn.3574

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., & Hoyer, J. (1999). Die Erfassung sozialer Phobie durch die Social Interaction Anxiety Scale (SIAS) und die Social Phobia Scale (SPS). *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *28*(1), 28–36. https://doi.org/10.1026//0084-5345.28.1.28

Stavropoulos, K. K. M., & Carver, L. J. (2013). Research review: Social motivation and oxytocin in autism--implications for joint attention development and intervention. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(6), 603–618. https://doi.org/10.1111/jcpp.12061

Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2014). Rapid fear detection relies on high spatial frequencies. *Psychological Science*, *25*(2), 566–574. https://doi.org/10.1177/0956797613512509

Stevens, S., Rist, F., & Gerlach, A. L. (2011). Eye movement assessment in individuals with social phobia: Differential usefulness for varying presentation times? *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(2), 219–224. https://doi.org/10.1016/J.JBTEP.2010.11.001

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences*, *10*(10), 471–476. https://doi.org/10.1016/J.TICS.2006.08.006

Striano, T., & Stahl, D. (2005). Sensitivity to triadic attention in early infancy. *Developmental Science*, *8*(4), 333–343. https://doi.org/10.1111/j.1467-7687.2005.00421.x

Suda, Y., & Kitazawa, S. (2015). A model of face selection in viewing video stories. *Scientific Reports*, *5*, 7666. https://doi.org/10.1038/srep07666

Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., Drew, A., Charman, T., Rees, L., & Wheelwright, S. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(5), 747–753. https://doi.org/10.1017/S0021963098002595

Sylvester, C. M., Yu, Q., Srivastava, A. B., Marek, S., Zheng, A., Alexopoulos, D., Smyser, C. D., Shimony, J. S., Ortega, M., Dierker, D. L., Patel, G. H., Nelson, S. M., Gilmore, A. W., McDermott, K. B., Berg, J. J., Drysdale, A. T., Perino, M. T., Snyder, A. Z., Raut, R. V., … Dosenbach, N. U. F. (2020). Individual-specific functional connectivity of the amygdala: A substrate for precision psychiatry. *Proceedings of the National Academy of Sciences*, 201910842. https://doi.org/10.1073/pnas.1910842117

Symons, L. A., Hains, S. M. J., & Muir, D. W. (1998). Look at me: five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. *Infant Behavior and Development*, *21*(3), 531–536. https://doi.org/10.1016/S0163-6383(98)90026-1

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and Communication in Autism. In *Handbook of autism and pervasive developmental disorders,* (pp. 335–364).

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, *11*(10), 697–709. https://doi.org/10.1038/nrn2889

Tamietto, M., Pullens, P., de Gelder, B., Weiskrantz, L., & Goebel, R. (2012). Subcortical Connections to Human Amygdala and Changes following Destruction of the Visual Cortex. *Current Biology*, *22*(15), 1449–1455. https://doi.org/10.1016/J.CUB.2012.06.006

Tanaka, J. W., & Curran, T. (2001). A Neural Basis for Expert Object Recognition. *Psychological Science*, *12*(1), 43–47. https://doi.org/10.1111/1467-9280.00308

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5–5. https://doi.org/10.1167/11.5.5

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

Tatler, B.W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision*, *11*(5), 5. https://doi.org/10.1167/11.5.5

Tennie, C., Frith, U., & Frith, C. D. (2010). Reputation management in the age of the world-wide web. *Trends in Cognitive Sciences*, *14*(11), 482–488. https://doi.org/10.1016/j.tics.2010.07.003

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

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*(6), 657–665. https://doi.org/10.1080/13506280500410949

Tipper, C. M., Handy, T. C., Giesbrecht, B., & Kingstone, A. (2008). Brain Responses to Biological Relevance. *Journal of Cognitive Neuroscience*, *20*(5), 879–891. https://doi.org/10.1162/jocn.2008.20510

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

Tsao, D. Y., & Freiwald, W. A. (2006). What's so special about the average face? *Trends in cognitive sciences*, *10*(9), 391-393.

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, *43*(3), 333–346.

https://doi.org/10.1016/S0042-6989(02)00498-4

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Ungerleider, S. K. and L. G. (2000). Mechanisms of Visual Attention in the Human Cortex. *Annual Review of Neuroscience*, *23*(1), 315–341. https://doi.org/10.1146/annurev.neuro.23.1.315

Van Belle, G., De Graef, P., Verfaillie, K., Busigny, T., & Rossion, B. (2010). Whole not hole: Expert face recognition requires holistic perception. *Neuropsychologia*, *48*(9), 2620–2629. https://doi.org/10.1016/j.neuropsychologia.2010.04.034

Vecera, S. P., & Rizzo, M. (2004). What are you looking at?: Impaired 'social attention' following frontal-lobe damage. *Neuropsychologia*, *42*(12), 1657–1665. https://doi.org/10.1016/j.neuropsychologia.2004.04.009

Velden, M., & Wolk, C. (1987). Depicting cardiac activity over real time: a proposal for standardization. *Journal of Psychophysiology*, *1*(2), 173–175. http://psycnet.apa.org/record/1989-04195-001

von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders*, *3*(2), 556–569. https://doi.org/10.1016/J.RASD.2008.12.003

Vuilleumier, P. (2000). Faces call for attention: evidence from patients with visual extinction. *Neuropsychologia*, *38*(5), 693–700. https://doi.org/10.1016/S0028-3932(99)00107-4

Wachsmuth, E., Oram, M. W., & Perrett, D. I. (1994). Recognition of Objects and Their Component Parts: Responses of Single Units in the Temporal Cortex of the Macaque. *Cerebral Cortex*, *4*(5), 509–522. https://doi.org/10.1093/cercor/4.5.509

Walters, K. S., & Hope, D. A. (1998). Analysis of social behavior in individuals with social phobia and nonanxious participants using a psychobiological model. *Behavior Therapy*, *29*(3), 387–407. https://doi.org/10.1016/S0005-7894(98)80039-7

Wang, L., Baumgartner, F., Kaule, F. R., Hanke, M., & Pollmann, S. (2019). Individual face- and house-related eye movement patterns distinctively activate FFA and PPA. *Nature Communications*, *10*(1), 5532. https://doi.org/10.1038/s41467-019-13541-3

Wang, S. (2019). Brief Report: Atypical Visual Exploration in Autism Spectrum Disorder Cannot be Attributed to the Amygdala. *Journal of Autism and Developmental Disorders*, *49*(6), 2605–2611. https://doi.org/10.1007/s10803-019-04009-w

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

Webber, C. L., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In *Tutorials in contemporary nonlinear methods for the Behavioral Sciences Web Book*. https://doi.org/10.1007/s00213-012-2719-8

Weeks, J. W., Heimberg, R. G., & Heuer, R. (2011). Exploring the Role of Behavioral Submissiveness in Social Anxiety. *Journal of Social and Clinical Psychology*, *30*(3), 217–249. https://doi.org/10.1521/jscp.2011.30.3.217

Weeks, J. W., Howell, A. N., & Goldin, P. R. (2013). Gaze avoidance in social anxiety disorder. *Depression and Anxiety*, *30*(8), 749–756. https://doi.org/10.1002/da.22146

Wells, A., Clark, D. M., Salkovskis, P., Ludgate, J., Hackmann, A., & Gelder, M. (1995). Social phobia: The role of in-situation safety behaviors in maintaining anxiety and negative beliefs. *Behavior Therapy*, *26*(1), 153-161.

Wessa, M., Kanske, P., Neumeister, P., Bode, K., Heissler, J., & Schönfelder, S. (2010). EmoPics: Subjektive und psychophysiologische Evaluation neuen Bildmaterials für die klinisch-biopsychologische Forschung. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *39*(1), 77.

Whalen, Paul J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge. *Journal of Neuroscience*, *18*(1), 411–418. https://doi.org/10.1523/JNEUROSCI.18-01-00411.1998

Wicker, B., Michel, F., Henaff, M.-A., & Decety, J. (1998). Brain Regions Involved in the Perception of Gaze: A PET Study. *NeuroImage*, *8*(2), 221–227. https://doi.org/10.1006/NIMG.1998.0357

Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. Journal of Anxiety Disorders, 23(1), 93–103. https://doi.org/10.1016/J.JANXDIS.2008.04.004

Wieser, M. J., Pauli, P., Weyers, P., Alpers, G. W., & Mühlberger, A. (2009). Fear of negative evaluation and the hypervigilance-avoidance hypothesis: an eye-tracking study. *Journal of Neural Transmission*, *116*(6), 717–723. https://doi.org/10.1007/s00702-008-0101-0

Williams, J. H. G., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *NeuroImage*, *25*(1), 133–140. https://doi.org/10.1016/J.NEUROIMAGE.2004.10.047

Wilming, N., Betz, T., Kietzmann, T. C., & König, P. (2011). Measures and limits of models of fixation selection. *PloS one*, *6*(9).

Wilson, C. E., Brock, J., & Palermo, R. (2010). Attention to social stimuli and facial identity recognition skills in autism spectrum disorder. *Journal of Intellectual Disability Research*, *54*(12), 1104–1115. https://doi.org/10.1111/j.1365-2788.2010.01340.x

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, *9*(1), 11–29. https://doi.org/10.1007/BF01531288

Wittchen, H. U., Stein, M. B., & Kessler, R. C. (1999). Social fears and social phobia in a community sample of adolescents and young adults: Prevalence, risk factors and co-morbidity. *Psychological Medicine*, *29*(2). https://doi.org/10.1017/S0033291798008174

Woodbury-Smith, M., & Scherer, S. W. (2018). Progress in the genetics of autism spectrum disorder. *Developmental Medicine & Child Neurology*, *60*(5), 445–451. https://doi.org/10.1111/dmcn.13717

Wu, D. W. L., Bischof, W. F., Anderson, N. C., Jakobsen, T., & Kingstone, A. (2014). The influence of personality on social attention. *Personality and Individual Differences*, *60*, 25–29. https://doi.org/10.1016/j.paid.2013.11.017

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

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In *Eye Movements and Vision* (pp. 171–211). Springer US. https://doi.org/10.1007/978-1-4899-5379-7_8

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

Zajonc, R. B. (1965). Social Facilitation. *Science*, *149*(7), 269–274. American Association for the Advancement of Science. https://doi.org/10.1641/0006-3568(2006)56[620:sf]2.0.co;2

Zhang, B., Noble, P. L., Winslow, J. T., Pine, D. S., & Nelson, E. E. (2012). Amygdala volume predicts patterns of eye fixation in rhesus monkeys. *Behavioural Brain Research*, *229*(2), 433–437.

https://doi.org/10.1016/J.BBR.2012.01.009

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zwickel, J., & Vo, M. L. H. (2010). How the presence of persons biases eye movements. *Psychonomic Bulletin & Review*, *17*(2), 257–262. https://doi.org/10.3758/PBR.17.2.257

LIST OF ABBREVIATIONS

ASD	Autism spectrum disorder
AUC	Area under the receiver operating characteristic curve
BOLD	Blood-oxygen-level dependent
CORM	Center of recurrent mass
DGP	German Psychological Society
D_{KL}	Kullback-Leibler divergence
DLPFC	Dorsolateral prefrontal cortex
DSM	Diagnostics and Statistical Manual of Mental Disorders
EEG	Electroencephalography
ERP	Event-related potential
FEF	Frontal eye fields
FFA	Fusiform face area
FV	Free-viewing
fMRI	Functional magnetic resonance imaging
GBVS	Graph-based visual saliency
GC	Gaze-contingent
GLMM	Generalized Linear Mixed Model
HF-HRV	Heart rate variability in the high frequency band
HR	Heart rate
IPS	Inferior parietal sulcus
ICD	International Classification of Diseases
MT	Middle temporal visual area
mPFC	Medial prefrontal cortex
PET	Positron-emission tomography
RDoC	Research Domain Criteria
ROI	Regions of interest
RQA	Recurrent quantification analysis
STS	Superior temporal sulcus
STG	Superior temporal gyrus
TPJ	Temporoparietal junction
V1	Human striate visual cortex
V2	Human prestriate visual cortex
V3	Third visual complex with visual area V3
V4	Human extrastriate cortex
V5	Human middle temporal visual area
VLPFC	Ventrolateral prefrontal cortex
VR	Virtual Reality

CURRICULUM VITAE

Lara Rösler

@ rosler.lara@gmail.com
ℹ German

Professional Experience

Current 03/2020	Postdoctoral Researcher, Sleep and Cognition Group, Netherlands Institute for Neuroscience › Investigating which mechanisms of insomnia are linked to anxiety and depression vulnerability. `fMRI` `EEG` `Psychophysiology` `Matlab` `R`
02/2020 04/2016	PhD Student, Department of Psychology, University of Würzburg › Investigating the neural and behavioral correlates of social attention under the supervision of Prof. Dr. Matthias Gamer. `Eye-Tracking` `fMRI` `MVPA` `GLMM` `Matlab` `R`
10/2019 02/2019	PhD Exchange and Research Assistant, Department of Psychology, University of Cambridge › Investigating dissimilarities of tonotopic auditory maps in wakefulness versus sleep under supervision of Dr. Anat Arzi. `fMRI` `EEG` `Decoding Analyses` `fmriprep` `BIDS` `Linux`
04/2016 10/2015	Research Assistant, Brain Imaging Center, University Hospital Frankfurt am Main › Investigating the neurophysiological underpinnings of attention and working memory dysfunction in patients with schizophrenia and patients with bipolar disorder `fMRI` `Matlab`
01/2015 07/2014	Research Assistant, 7-Tesla-MRI Flow-Manager, Psychiatry, University Hospital Utrecht › Planning, execution and supervision of all scans at the 7Tesla MRI scanner.
12/2014 01/2014	Research Intern, STAR Lab, Brain Center Rudolf Magnus, Utrecht › Investigating behavioral, structural, and metabolic correlates of agency dysfunction in schizophrenia under the supervision of Prof. Dr. Bas Neggers and Dr. Katharine Thakkar. `1H-MRS` `Eye-Tracking` `Matlab`

Publications

In Prep	**Rösler, L.** & Gamer, M. Early visual cortex and amygdala activity is modulated by reflexive social attention.
Under Review	**Rösler, L.**, Göhring, S., Strunz, M. & Gamer, M. Social anxiety is associated with heart rate but not gaze in a real social interaction. *Psychiatry Research.* Preprint available from : https ://psyarxiv.com/gps3h/
2019	**Rösler, L.**, Rubo, M. & Gamer, M. Artificial faces predict gaze allocation in complex scenes. *Frontiers in Psychology.*
2019	**Rösler, L.** & Gamer, M. . Freezing of gaze during action preparation under threat imminence. *Scientific Reports.* Preprint available from : https ://psyarxiv.com/a2pfh/
2019	Yao, B., Neggers, S.F.W., Rolfs, M., **Rösler, L.**, ... & Thakkar, K.N. Structural thalamo-frontal hypoconnectivity is related to oculomotor corollary discharge dysfunction in schizophrenia. *Journal of Neuroscience.*
2018	Flechsenhar*, A., **Rösler*, L.** & Gamer, M. Investigating the dynamics of social attention with a gaze-contingent display and recurrence quantification analysis. *Scientific Reports.* *Both authors contributed equally.
2017	**Rösler, L.**, End, A., & Gamer, M. Orienting towards social features in naturalistic scenes is reflexive. *PloS ONE.*
2017	Thakkar, K.N., **Rösler, L.**, Wijnen, J.P., Boer, V.O., Klomp, D.W.O., Cahn, W., Kahn, R.S. & Neggers, S.F.W. 7T proton magnetic resonance spectroscopy of GABA, glutamate, and glutamine reveals altered concentrations in schizophrenia patients and healthy siblings. *Biological Psychiatry.*
2015	**Rösler, L.**, Rolfs, M., Van der Stigchel, S., Neggers, S. F.W., Cahn, W., Kahn, R. S., & Thakkar, K. N.. Failure to use corollary discharge to remap visual target locations is associated with psychotic symptom severity in schizophrenia. *Journal of Neurophysiology.*

Education

2013 – 2015	Master of Science, Neuroscience and Cognition, Utrecht University, *Cum Laude, highest distinction*
2010 – 2013	Bachelor of Liberal Arts, Philosophy and Cognitive Neurosciences, University College Utrecht, *Magna Cum Laude*

Conference Contributions

2019	**Rösler, L.** & Gamer, M. Freezing of gaze supports action preparation under threat imminence. *Poster presented at WASAD Congress, Würzburg, Germany.*
2018	**Rösler, L.** & Gamer, M. Early visual cortex and amygdala activity is modulated by reflexive social attention. *Poster presented at Gentner Symposium "Understanding Others : From Psychological Concepts to Neural Mechanisms", Jerusalem, Israel.*
2018	**Rösler, L.** & Gamer, M. The influence of social anxiety on a real-life social interaction. *Talk presented at Science Conference of Zentrum für Psychische Gesundheit, Würzburg, Germany.*
2018	**Rösler, L.** & Gamer, M. Behavioral and neural mechanisms of reflexive social attention. *Talk presented at Psychologie und Gehirn Conference, Giessen, Germany.*
2018	**Rösler, L.** & Gamer, M. Neural mechanisms of reflexive social attention : a combined eye-tracking and fMRI study. *Poster presented at the Annual Meeting of the Cognitive Neuroscience Society, Boston, United States.*
2017	**Rösler, L.**, End, A., & Gamer, M. Orienting towards social features in naturalistic scenes is reflexive. *Poster presented at the International Conference for Cognitive Neuroscience, Amsterdam, The Netherlands.*
2016	**Rösler, L.**, Thakkar, K.N., Wijnen, J.P., Boer, V.O., Klomp, D.J.W., Cahn, W., Kahn, R.S. & Neggers, S.F.W. Altered Glutamine, Glutamate and GABA Levels in Schizophrenia Patients and Their Healthy First-Degree Relatives : A 1H-MRS Study at 7T. *Talk presented at the biennial meeting of the Schizophrenia International Research Society, Florence, Italy.*

Teaching and Mentoring

2019 Clinical Psychology Seminar, Bachelor of Psychology, University of Würzburg
- Discussing emotional disorders in depth with a focus on network theory

2019
2016 Biopsychology Seminar, Bachelor of Psychology, University of Würzburg
- Teaching biopsychological methods in a seminar setting
- Acquainting students with the study of scientific literature

June 2018 Biopsychology Lecture, Bachelor of Psychology, University of Würzburg
- Two lectures on functional magnetic resonance imaging physics, analyses and applications

2019
2016 Thesis Supervision, B.Sc. and M.Sc., University of Würzburg
- Sandrine Michelberger, Master thesis : Does social anxiety modulate social attention ? An eye-tracking study.
- Teresa Wörsdörfer, Master thesis : The neural correlates of social attention.
- Stefan Göring, Master thesis : The influence of social anxiety on gaze during a real social interaction.
- Michael Strunz, Master thesis : How social anxiety influence a real social interaction - An eye-tracking and psychophysiology study.
- Julia Jochim, Bachelor thesis : The influence of fear on visual attention.
- Alexandra Vogt, Bachelor thesis : Intraindividual stability of visual exploration patterns - from the lab to the real life.

Honors and Awards

2019	Poster Award, World Association for Stress Related and Anxiety Disorders (WASAD) Congress 2019
2017	Neuroscience Research Award, Graduate School of Life Sciences, Würzburg
2017	Travel Grant from the Graduate School of Life Sciences Würzburg travel award to attend the Statistical Parametric Mapping (SPM) workshop, London
2016	Travel Grant from the University Hospital Frankfurt to attend the Schizophrenia International Research Society Conference, Florence
2014-2015	U/Select Honors Program, Utrecht University

Languages

German	●●●●●
English	●●●●●
Dutch	●●●●○

Programming

- Matlab
- R
- Neurobehavioral Systems Presentation
- Bash

STATEMENT OF INDIVIDUAL AUTHOR CONTRIBUTIONS

Statement of individual author contributions and of legal second publication rights

Dissertation Based on Several Published Manuscripts

Publication (complete reference): Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. PloS one, 12(7), e0182037.					
Participated in	Author Initials, Responsibility decreasing from left to right				
Study Design Methods Development	M.G. (40%) L.R. (40%)	A.E. (20%)			
Data Collection	L.R. (90%)	M.G. (10%)			
Data Analysis and Interpretation	L.R. (70%)	M.G. (25%)	A.E. (5%)		
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 L.R. (80%) L.R. (70%) L.R. (80%) L.R. (90%)	 M.G. (15%) M.G. (25%) M.G. (15%) M.G. (10%)	 A.E. (5%) A.E. (5%) A.E. (5%)		

Explanations

L. Rösler: conceptualization, data curation, formal analysis, visualization, writing original draft as well as writing review and editing.

A. End: conceptualization, review and editing.

M. Gamer: supervision, funding acquisition, conceptualization, validation, writing review and editing.

Publication (complete reference): Flechsenhar, A., Rösler, L. & Gamer, M. (2018). Attentional Selection of Social Features Persists Despite Restricted Bottom-Up Information and Affects Temporal Viewing Dynamics. Scientific Reports; 8(1), 12555. http://dx.doi.org/10.1038/s41598-018-30736-8.					
Participated in	Author Initials, Responsibility decreasing from left to right				
Study Design Methods Development	M.G.(33%) L.R. (33%) & A.F. (33%)				
Data Collection	L.R. (45%) & A.F. (45%)	M.G. (10%)			
Data Analysis and Interpretation	L.R. (40%) & A.F. (40%)	M.G. (20%)			
Manuscript Writing					
Writing of Introduction	L.R. (40%) & A.F. (40%)	M.G. (20%)			
Writing of Materials & Methods	L.R. (40%) & A.F. (40%)	M.G. (20%)			
Writing of Discussion	L.R. (40%) & A.F. (40%)	M.G. (20%)			
Writing of First Draft	L.R. (45%) & A.F. (45%)	M.G. (10%)			

Explanations

L. Rösler and A. Flechsenhar share authorship of this publication.

L. Rösler & A. Flechsenhar: conceptualization, data curation, formal analysis, visualization, writing original draft as well as writing review and editing.

M. Gamer: supervision, funding acquisition, conceptualization, validation, writing review and editing.

Publication (complete reference): Rösler, L., Rubo, M., & Gamer, M. (2019). Artificial Faces Predict Gaze Allocation in Complex Dynamic Scenes. Frontiers in Psychology, 10, 2877.					
Participated in	Author Initials, Responsibility decreasing from left to right				
Study Design Methods Development	M.G. (50%) M.R. (50%)				
Data Collection	M.R. (90%)	M.G. (10%)			
Data Analysis and Interpretation	L.R. (40%) & M.R.(40%)	M.G. (20%)			
Manuscript Writing					
Writing of Introduction	L.R. (70%)	M.R. (20%)	M.G. (10%)		
Writing of Materials & Methods	L.R. (50%)	M.R. (30%)	M.G. (20%)		
Writing of Discussion	L.R. (70%)	M.R. (20%)	M.G. (10%)		
Writing of First Draft	L.R. (90%)	M.G. (10%)			

Explanations

L. Rösler and M. Rubo share authorship of this publication.

L. Rösler: formal analysis, visualization, writing original draft as well as writing review and editing.

M. Rubo: conceptualization, data curation, formal analysis, review and editing.

M. Gamer: supervision, funding acquisition, conceptualization, partial analysis, validation, writing review and editing.

Publication (complete reference): Rösler, L., Göhring, S., Strunz, M., & Gamer, M. (Under Review). Social anxiety is associated with heart rate but not gaze in a real social interaction. Psychiatry Research.					
Participated in	Author Initials, Responsibility decreasing from left to right				
Study Design Methods Development	M.G. (40%) & L.R. (40%)	S.G. (10%) & M.S. (10%)			
Data Collection	S.G. (40%) & M.S. (40%)	L.R. (15%)	M.G. (5%)		
Data Analysis and Interpretation	L.R. (75%)	M.G. (25%)			
Manuscript Writing					
Writing of Introduction	L.R. (80%)	M.G. (20%)			
Writing of Materials & Methods	L.R. (70%)	M.G. (20%)	S.G. (5%) & M.S. (5%)		
Writing of Discussion	L.R. (80%)	M.G. (20%)			
Writing of First Draft	L.R. (90%)	M.G. (10%)			

Explanations

L. Rösler: conceptualization, formal analysis, visualization, writing original draft as well as writing review and editing.

S. Göhring & M. Strunz: conceptualization, data curation, review and editing.

M. Gamer: supervision, funding acquisition, conceptualization, partial analysis, validation, writing review and editing.

The doctoral researcher confirms that she/he has obtained permission from both the publishers and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the abovementioned assessment.

Lara Rösler		Würzburg	
Doctoral Researcher’s Name	Date	Place	Signature

Prof. Dr. Matthias Gamer		Würzburg	
Primary Supervisor’s Name	Date	Place	Signature

Statement of individual author contributions to figures/tables/chapters included in the manuscripts

Dissertation Based on Several Published Manuscripts

<table>
<tr><td colspan="6">Publication (complete reference):
Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. PloS one, 12(7), e0182037.</td></tr>
<tr><td>Figure</td><td colspan="5">Author Initials, Responsibility decreasing from left to right</td></tr>
<tr><td>1</td><td>L.R.</td><td>M.G.</td><td>A.E.</td><td></td><td></td></tr>
<tr><td>2</td><td>L.R.</td><td>M.G.</td><td>A.E.</td><td></td><td></td></tr>
<tr><td>3</td><td>L.R.</td><td>M.G.</td><td>A.E.</td><td></td><td></td></tr>
<tr><td>Table</td><td colspan="5">Author Initials, Responsibility decreasing from left to right</td></tr>
<tr><td>1</td><td>L.R.</td><td>M.G.</td><td>A.E.</td><td></td><td></td></tr>
</table>

<table>
<tr><td colspan="6">Publication (complete reference):
Flechsenhar, A., Rösler, L. & Gamer, M. (2018). Attentional Selection of Social Features Persists Despite Restricted Bottom-Up Information and Affects Temporal Viewing Dynamics. Scientific Reports; 8(1), 12555. http://dx.doi.org/10.1038/s41598-018-30736-8</td></tr>
<tr><td>Figure</td><td colspan="5">Author Initials, Responsibility decreasing from left to right</td></tr>
<tr><td>1</td><td>L.R. & A.F.</td><td>M.G.</td><td></td><td></td><td></td></tr>
<tr><td>2</td><td>L.R. & A.F.</td><td>M.G.</td><td></td><td></td><td></td></tr>
<tr><td>3</td><td>L.R. & A.F.</td><td>M.G.</td><td></td><td></td><td></td></tr>
<tr><td>4</td><td>L.R. & A.F.</td><td>M.G.</td><td></td><td></td><td></td></tr>
<tr><td>5</td><td>L.R. & A.F.</td><td>M.G.</td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><td colspan="6">Publication (complete reference):
Rösler, L., Rubo, M., & Gamer, M. (2019). Artificial Faces Predict Gaze Allocation in Complex Dynamic Scenes. Frontiers in Psychology, 10, 2877.</td></tr>
<tr><td>Figure</td><td colspan="5">Author Initials, Responsibility decreasing from left to right</td></tr>
<tr><td>1</td><td>M.G.</td><td>L.R.</td><td>M.R.</td><td></td><td></td></tr>
<tr><td>2</td><td>L.R.</td><td>M.R.</td><td>M.G.</td><td></td><td></td></tr>
<tr><td>Table</td><td colspan="5">Author Initials, Responsibility decreasing from left to right</td></tr>
<tr><td>1</td><td>L.R.</td><td>M.R.</td><td>M.G.</td><td></td><td></td></tr>
</table>

Publication (complete reference): Rösler, L., Göhring, S., Strunz, M., & Gamer, M. (Under Review). Social anxiety is associated with heart rate but not gaze in a real social interaction. Psychiatry Research.					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	S.G.	L.R.	M.S.	M.G.	
2	L.R.	S.G.	M.S.	M.G.	
3	L.R.	M.G.			
4	L.R.	M.G.			
5	L.R.	M.G.			
Table	**Author Initials,** Responsibility decreasing from left to right				
1	L.R.	S.G.	M.S.	M.G.	

I also confirm my primary supervisor’s acceptance.

Lara Rösler Würzburg

Doctoral Researcher’s Name Date Place Signature

AFFIDAVIT

I hereby confirm that my thesis entitled *Behavioral and Neural Mechanisms of Social Attention* is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date	Signature

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich an Eides statt, die Dissertation *Behaviorale und Neuronale Mechanismen der Sozialen Aufmerksamkeit* eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum

Unterschrift