Social Attention in the Laboratory, in Real Life and in Virtual Reality

Soziale Aufmerksamkeit im Labor, in vivo und in virtueller Realität

DISSERTATION

For a Doctoral Degree (Dr. rer. nat)

Faculty of Human Sciences

Julius-Maximilians-Universität Würzburg

Submitted by

MARIUS RUBO

Born in Worms

Würzburg, 2019

MEMBERS OF THE COMMITEE:

Chairperson:	Prof. Dr. Paul Pauli
Primary Supervisor:	Prof. Dr. Matthias Gamer
Supervisor (Second):	Prof. Dr. Lynn Huestegge

Date of Public Defense:	______________________
Date of Receipt of Certificates:	______________________

Submitted On:	______________________
	Office Stamp

ACKNOWLEDGEMENTS

Several individuals deserve recognition for their help in this work. First, I would like to thank Prof. Dr. Matthias Gamer for his guiding throughout this project and for also allowing me to unfold my research passions and familiarize myself with a range of methods and approaches. I also owe a great deal to Albert End, who joined our overarching research project before me and did some of the heavy lifting that I and others could later capitalise on. I am grateful to Aleya Flechsenhar and Lara Rösler for their continuous exchange and aid in some of the tricky details which surrounded my research topic. Many of my other colleagues in our work group also deserve recognition for their help and support. Dr. Kristina Suchotzki, Dr. Mario Reutter and Dr. Lea Hildebrandt all played an invaluable role in introducing me to the world of science and the specifics one needs to know to find one's way. I also want to acknowledge the many enlightening conversations I had with Daniel Gromer, Dr. Ivo Käthner and Jonas Großekathöfer. All my other friends and colleagues at our institute deserve recognition for continually offering support which made my time here easier and more enjoyable. I would also like thank my committee for providing insightful feedback and orientation in maturing as a researcher. Finally, Würzburg University as an institution deserves recognition for providing an inspiring environment and many stimulating scientific events throughout the year.

Summary

Social attention is a ubiquitous, but also enigmatic and sometimes elusive phenomenon. We direct our gaze at other human beings to see what they are doing and to guess their intentions, but we may also absorb social events en passant as they unfold in the corner of the eye. We use our gaze as a discrete communication channel, sometimes conveying pieces of information which would be difficult to explicate, but we may also find ourselves avoiding eye-contact with others in moments when self-disclosure is fear-laden. We experience our gaze as the most genuine expression of our will, but research also suggests considerable levels of predictability and automaticity in our gaze behavior. The phenomenon's complexity has hindered researchers from developing a unified framework which can conclusively accommodate all of its aspects, or from even agreeing on the most promising research methodologies.

The present work follows a multi-methods approach, taking on several aspects of the phenomenon from various directions. Participants in study 1 viewed dynamic social scenes on a computer screen. Here, low-level physical saliency (i.e. color, contrast, or motion) and human heads both attracted gaze to a similar extent, providing a comparison of two vastly different classes of gaze predictors in direct juxtaposition. In study 2, participants with varying degrees of social anxiety walked in a public train station while their eye movements were tracked. With increasing levels of social anxiety, participants showed a relative avoidance of gaze at near compared to distant people. When replicating the experiment in a laboratory situation with a matched participant group, social anxiety did not modulate gaze behavior, fueling the debate around appropriate experimental designs in the field. Study 3 employed virtual reality (VR) to investigate social gaze in a complex and immersive, but still highly controlled situation. In this situation, participants exhibited a gaze behavior which may be more typical for real-life compared to laboratory situations as they avoided gaze contact with a virtual conspecific unless she gazed at them. This study provided important insights into gaze behavior in virtual social situations, helping to better estimate the possible benefits of this new research approach. Throughout all three experiments, participants showed consistent inter-individual differences in their gaze behavior. However, the present work could not resolve if these differences are linked to psychologically meaningful traits or if they instead have an epiphenomenal character.

ZUSAMMENFASSUNG

Soziale Aufmerksamkeit ist ein allgegenwärtiges, aber auch ein rätselhaftes Phänomen, das mitunter schwierig zu fassen ist. Wir richten unseren Blick auf andere Menschen, um ihr Tun zu verfolgen und um ihre Absichten einzuschätzen, aber manchmal verfolgen wir soziale Ereignisse auch ganz beiläufig aus dem Augenwinkel heraus. Wir setzen unseren Blick als ein eigenes Kommunikationsmedium ein und übertragen mit ihm teilweise Botschaften, die nur schwer zu beschreiben sind, aber wir weichen mitunter dem Blickkontakt mit anderen auch aus, wenn wir Angst davor haben, zu viel von uns preiszugeben. Unser Blick stellt sich für uns als eine ureigene Äußerung unseres Willens dar, aber die Forschung hat auch gezeigt, dass unser Blickverhalten in beträchtlichem Maße vorhersehbar und automatisch abläuft. In der Vielschichtigkeit des Phänomens liegt für Forscher eine Hürde bei dem Versuch, alle seine Aspekte schlüssig in ein umfassendes Bezugssystem einzuordnen, oder sich auch nur auf die vielversprechendsten Forschungsmethoden zu einigen.

Die vorliegende Arbeit verbindet den Einsatz unterschiedlicher Methoden, um sich mehreren Aspekten des Phänomens aus verschiedenen Blickrichtungen zu nähern. Die Versuchspersonen in Studie 1 sahen dynamische soziale Szenen, die ihnen auf einem Computerbildschirm dargeboten wurden. Hierbei wurde ihr Blick in ähnlichem Maße von physikalischer Salienz (z.B. Farbe, Kontrast oder Bewegung) angezogen wie von menschlichen Köpfen, wodurch zwei ganz unterschiedliche Gruppen von Prädiktoren für Blickverhalten in direkter Gegenüberstellung verglichen wurden. In Studie 2 bewegten sich Versuchspersonen mit unterschiedlich ausgeprägter sozialer Ängstlichkeit zu Fuß in einem öffentlichen Bahnhof, während ihre Augenbewegungen erfasst wurden. Mit zunehmender sozialer Ängstlichkeit neigten Versuchspersonen dazu, nahe Personen im Gegensatz entfernteren Personen im Verhältnis weniger anzuschauen. Als das Experiment mit einer gematchten Gruppe von Versuchspersonen in einer Laborsituation wiederholt wurde, zeigte sich kein Einfluss der sozialen Ängstlichkeit auf das Blickverhalten, was der Diskussion um angemessene experimentelle Designs in diesem Forschungsbereich einen weiteren Impuls verlieh. In Studie 3 wurde Virtuelle Realität (VR) eingesetzt, um das Blickverhalten in einer komplexen und immersiven, aber dennoch streng kontrollierten Umgebung zu untersuchen. In dieser Situation zeigten Probanden ein Blickverhalten, das eher dem in echten Situationen als dem im Labor entspricht, indem sie direkten Blickkontakt mit einer virtuellen Person mieden, so lange diese sie nicht anschaute. Durch diese Studie konnten wichtige Erkenntnisse über das Blickverhalten in sozialen virtuellen Situationen gewonnen werden, wodurch der mögliche Nutzen dieses neuen Forschungsansatzes besser beurteilt werden kann. In allen drei Experimenten zeigten Versuchspersonen konsistente inter-individuelle Unterschiede in ihrem Blickverhalten. Es konnte jedoch im Rahmen der vorliegenden Arbeit nicht geklärt werden, ob diese Unterschiede psychologisch bedeutsame Eigenschaften oder eher Epiphänomene darstellen.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADHD	Hyperactivity Disorder
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
ASD	Autism Spectrum Disorder
AQ	Autism Spectrum Quotient
BDI	Beck's Depression Inventory
BFI	Big Five Inventory
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalography
FASD	Fetal Alcohol Spectrum Disorder
FFA	Fusiform Face Area
fMRI	Functional Magnetic Resonance Imaging
GBVS	Graph-Based Visual Saliency
GLMM	Generalized Linear Mixed Model
HR	Heart Rate
IAPS	International Affective Picture System
IAT	Implicit Association Test
PD	Parkinson's Disease
REML	Restricted Maximum Likelihood
ROI	Region of Interest
SAD	Social Anxiety Disorder
SEM	Standard Error of the Mean
SIAS	Social Interaction Anxiety Scale
SPAI	Social Phobia and Anxiety Inventory
STAI	State-Trait Anxiety Inventory
TPJ	Temporoparietal Junction
V1	Primary Visual Cortex
V2	Secondary Visual Cortex
V3	Third Visual Complex
VR	Virtual Reality

CHAPTER 1

INTRODUCTION

Basic Concepts in Social Attention

Humans guide their attention, and find their attention to be guided by the external world, in every waking moment. Several debates exist around the questions how precisely shifts in attention are orchestrated, and if attention to social information bears characteristics which are qualitatively different from attention to other objects around us.

Bottom-Up and Top-Down Attention

Attention is the process of selecting relevant pieces of information from the wealth of sensory inputs which our brains, as well as the brains of our relatives in the animal kingdom, are confronted with at each point in time. Traditional models of attention distinguish four processes which are seen as fundamental: bottom-up processing, competitive selection, top-down sensitivity control and working memory (Knudsen, 2007). In these models, sensory input is first filtered for salient stimuli using bottom-up processes, which operate automatically and are not modulated by higher cognitive functions. Among these initial neural representations of the outside world, processes of competitive selection determine which pieces of information arrive in the working memory. At this stage, however, the working memory is capable of influencing or posing restrictions to the competitive selection of information via top-down controlling of the sensitivity of different information channels. The pieces of information which have survived this process, i.e. the objects of attention, are now represented in the working memory and ready to be processed in further ways, some of which involve higher cognitive functioning.

The model thus formulates a clear dichotomy between bottom-up processes which are determined mostly by stimulus characteristics and act automatically, and top-down processes which are goal-directed and may at least to some degree be under voluntary control. This dichotomy is in line with most people's intuitive understanding of attention and is easily depicted in examples from everyday situations: For instance, if one is looking for a particular type of product in a supermarket shelf, this product is usually more likely to attract attention than if it is not being looked for. Here, processes of top-down attentional regulation are at play. On the other hand, a loud noise or a flashing light will also capture one's attention, even though one was not waiting for or looking for such a stimulus to occur. In such a situation, a bottom-up regulation of attention can be observed. Of course, top-down and bottom-up control of attention also occur and interact at more subtle levels than in these examples. For instance, it was established that physical properties like contrast or motion increase the probability of the attentional system to guide attention towards a specific area in the field of view (see below under *visual saliency*). During the past years, the conceptual separation of bottom-up and top-down mechanisms remained largely unchallenged (although some findings led to the suggestion that a repetition of past selections should be seen as a unique process which needs to be differentiated both from bottom-up and top-down mechanisms (Awh, Belopolsky, & Theeuwes, 2012)). The dichotomy is also reflected in neural models of attentional selection, which suggest that a dorsal network regulates attention under top-down control (also *endogenous* attention), while a ventral network mediates bottom-up attentional processes (Corbetta, Patel, & Shulman, 2008; Corbetta & Shulman, 2002). The ventral system, which is strongly lateralized to the right hemisphere, is thought to consist of the temporoparietal junction (*TPJ*) and some parts of middle frontal gyrus, the inferior frontal gyrus as well as the anterior insula. The dorsal system, on the other hands, shows no clear lateralization and consists of the dorsal parietal cortex and dorsal frontal cortex, including the frontal eye fields.

Is Social Attention an Instance of Top-Down Attention?

In the past, models of attention were often mainly informed by experiments using non-social stimuli, and it was (sometimes implicitly) assumed that attention towards social stimuli represents just another instance of top-down attentional control. This view implies that social information is detected in the brain after the visual signal has been modulated by the faster bottom-up processes. However, several studies found saccades towards regions containing social information very shortly after stimulus onset (Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Rösler, End, & Gamer, 2017; Scheller, Büchel, & Gamer, 2012) and it was argued that they may be best described as reflexive eye movements. It was proposed that processing of socially relevant regions may be partially driven by subcortical routes involving the amygdala (Benuzzi et al., 2007; Gamer & Büchel, 2009; Gamer, Schmitz, Tittgemeyer, & Schilbach, 2013), and not (entirely) by top-down information processing along cortical routes (Knudsen, 2007). Additionally, the TPJ, which has been argued to mediate attention towards social information (Nardo, Santangelo, & Macaluso, 2011), lacks the retinotopical organization required to directly control saccades towards social features, and it was argued that an additional coupling with other networks makes the fast reaction times for saccades towards social features difficult to explain. Instead, it was proposed that a fast subcortical route involving the amygdala may be crucial in evaluating the personal relevance of external stimuli (Pessoa, 2010; Pessoa & Adolphs, 2010; Sander, Grafman, & Zalla, 2003), which may include a preference for social stimuli. Although the amygdala was originally associated with fear responses, i.e. with connecting external stimuli to defense responses (LeDoux, 2003), several findings support the view that the amygdala is furthermore involved in the processing of social stimuli. Adolphs et al. (2005) found a patient with bilateral amygdala damage to show a reduced amount of spontaneous fixations to the eyes when viewing images of faces. Along these lines, Taubert et al. (2018) reported that rhesus monkeys lost their otherwise strong preference for looking at real and illusory faces after bilateral amygdala lesions. In another study by Gamer & Büchel (2009), amygdala activation predicted gazing towards the eye region in fearful faces. Altogether, there is now accumulating evidence for an involvement of the amygdala in the regulation of social attention (Adolphs, 2010b), substantiating the view that social stimuli are processed in a qualitatively different manner compared to other stimuli of interest.

Face Processing

Besides the fast spontaneous saccades towards social stimuli which are not easily explained by cortical processing, other lines of research have also proposed that the processing of human faces is special and not merely a learned subcategory of more general object processing functionalities. One piece of evidence comes from the existence of a neuropsychological condition called *prosopagnosia*, in which the ability to recognize even very familiar faces may be impaired, while object recognition is often normal (Barton, Press, Keenan, & Connor, 2002). Secondly, it was found that recognition of faces in healthy populations suffers more from inverting an image upside down compared to the recognition of objects under the same circumstances (Kanwisher, Tong, & Nakayama, 1998). Hershler & Hochstein (2005) found the reaction time for finding a face amidst several distractors to be relatively independent of the number of distractors (i.e. to *pop out*), an effect which could not be documented in the reverse search (finding an object amongst several faces). Additionally, fMRI brain scans could identify a brain region which appears to be rather specifically involved in the detection of faces, the fusiform face area (*FFA*, Ishai, Ungerleider, Martin, & Haxby, 2000; Kanwisher, McDermott, & Chun, 1997).

In many of these studies, the interpretation was not entirely unambiguous. In the case of reaction times towards face-like stimuli, it was discussed whether low-level factors such as the Fourier amplitude information in images may explain the findings better than assumptions about a *holistic* mode of face processing (Hershler

& Hochstein, 2006; VanRullen, 2006). The functional specificity of the FFA towards processing of faces was also questioned when it was shown that expertise in a specific domain (e.g. cars or birds) can also lead to higher activation in the FFA when an instance of that domain is seen (Gauthier, Skudlarski, Gore, & Anderson, 2000; Tarr & Gauthier, 2000). This finding opens up the possible interpretation that the FFA mediates expertise, and is merely triggered by faces since most persons have acquired a high expertise in face recognition. This interpretation, on the other hand, is somewhat restrained by the observation that FFA activation was still larger for face stimuli compared to other well-learned stimuli. Additionally, it was pointed out that some studies may not have properly accounted for the fact that faces are more geometrically eccentric than other stimuli (Hasson, Levy, Behrmann, Hendler, & Malach, 2002), and that this eccentricity may have partially driven some of the observed effects.

In sum, although several individual studies were critizised for not excluding all possible alternative explanations, it seems fair to say that the overall picture suggests that face processing in humans involves mechanisms which are at least partially domain specific and separate from more general top-down processing mechanisms (Duchaine & Yovel, 2015).

Gaze Following

A behavioral phenomenon within the domain of social attention which we encounter frequently in our daily lives is *gaze following*, a shift in one's visuospatial attention towards an area gazed at by a conspecific. Take the example of buying vegetables on a market: when we notice another person turn her head or gaze towards an area in space we have not paid attention to yet, we are likely to orient our own gaze on the same spot to see if we find a vegetable crop which is interesting for us. Although ubiquitous in the daily lives of many people, this behavior as well as its statistical properties and temporal arrangement with other behaviors have barely been directly addressed in psychological research. Instead, research attempted to identify and describe basic mechanisms that underlie this behavior. A seminal study by Friesen & Kingstone (1998) asked participants to respond to a letter displayed at the left or right edge of a computer screen, while a face presented in the screen's center was either looking to the right, to the left or straight ahead. Although the presented gaze direction was not predictive for the appearance of letters, and participants were informed about this fact, response times were faster in trials in which gaze was directed towards compared to away from the later appearing target stimulus, indicating a shift in attention as a consequence of another person's gaze direction.

This finding spawned a debate on whether such effects generated using social stimuli need to necessarily involve processes of social attention, or can instead be better described in terms of more general attentional processes. Critically, a similar effect can be produced just by employing arrows as substitutes for gaze direction in presented faces (Eimer, 1997; Posner, 1980; Ristic, Friesen, & Kingstone, 2002; Tipples, 2002), although the effects were found to be more robust for social cues (Birmingham & Kingstone, 2009; Frischen, Bayliss, & Tipper, 2007). Social cues, unlike arrows, were found to be predictive for attentional shifts even if they were counter-predictive for the appearance of a target cue (Friesen, Ristic, & Kingstone, 2004). Likewise, gaze cueing was found to be robust against effects of cue-target color contingencies, while arrow cueing was not (Ristic, Wright, & Kingstone, 2007). Additionally, gaze cueing and arrow cueing were found to activate distinct structures in the brain (Hietanen, Nummenmaa, Nyman, Parkkola, & Hämäläinen, 2006): While gaze-cued orienting was associated with activation in the left inferior occipital gyrus and the right medial and inferior occipital gyri, arrow-cued orienting was found to elicit enhanced activation in a wider range of structures, such as areas in the medial/inferior occipital gyri and the medial temporal gyri as well as in the left intraparietal area.

Additionally, it was argued that faces and arrows share more similarities in laboratory situations compared to real-world situations, leading to blurred distinctions between the two types of cues when used in the laboratory (Gibson & Kingstone, 2006): While word cues like *above*, *below*, *left* or *right* represent projective relations which must be interpreted with respect to a reference frame centered around a reference object, arrows as well as a person's gaze represent deictic relations which can be understood immediately with little interpretation effort. Since the deictic properties of many cues are not as clearly preserved in real-life situations as other person's gaze - most cues do not suddenly appear at the hemisphere of interest and arrows often do not point precisely at a point of interest - the distinction should become clearer as one leaves the laboratory and enters real-world situations. However, there has not yet been a rigorous attempt to compare attentional shifts following gaze or arrows in real-world scenarios, which would be an important backup for such theoretical considerations. Taken together, evidence cumulates to the view that while attentional processes exist that share basic mechanisms with gaze following, the phenomenon seems to also possess properties which are uniquely social and cannot be investigated using abstract signals.

Interestingly, gaze following is also found in several primate species (Emery, 2000; Tomasello, Call, & Hare, 1998), although the cognitive processes involved in this behavior appear to differ substantially (Rosati & Hare, 2009): For instance, while some apes like chimpanzees and bonobos do, to a lesser extent than humans, react to changes in gaze directions in conspecifics (Tomasello, Hare, Lehmann, & Call, 2007), other species like rhesus monkeys (Emery, Lorincz, Perrett, Oram, & Baker, 1997) and capuchin monkeys (Vick & Anderson, 2000) were only observed to react to shifts in a conspecific's head direction. Note that gaze direction in humans is more easily detectable than in any other primate species due to the large exposed sclera in our eye outline (Kobayashi & Kohshima, 1997, 2001), which may point to a co-evolution of social attentional mechanisms and physical properties in humans.

Development of Social Attention and Cognition

By the time most humans reach adulthood, and usually long before, they have acquired an expertise in understanding and handling social situations that goes beyond what current theories in social cognition are able to model, and that is unrivalled by current attempts to reproduce social cognition using artificial intelligence. So just how do humans become experts in social situations, and does a deep understanding of these learning processes help to identify how and why persons suffering from psychiatric conditions like autism spectrum disorder or social anxiety disorder sometimes have trouble in managing social situations?

As a foundation for learning from social cues, humans possess a congenital propensity to scan their environments for social information which is documented even in newborns (Haaf & Bell, 1967; Haith, Bergman, & Moore, 1977; Muir, Humphrey, & Humphrey, 1994). For instance, one study (Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000) found newborn babies to spend more time looking at a photograph depicting a person with her eyes open compared to the same photograph with the person having her eyes closed. Farroni, Csibra, Simion, & Johnson (2002) found 2-to 5-day old newborns to preferentially look at persons showing direct gaze towards them compared to showing averted gaze, and moreover were found to have an enhanced N170 response in the EEG of 11 out of 12 occipitally placed electrodes when gaze was directed towards them. In another study by Farroni, Massaccesi, Pividori, & Johnson (2004), 2-to 5-day old newborns were furthermore faster at gazing at objects presented in their periphery after they had observed a real or schematic face move its gaze to the same direction; that is, they showed a simple form of gaze following behavior. One study (Bushnell, 2001) even found a subtle but significant preference in babies as young as two to seven hours to gaze at their mothers compared to a female stranger, although this finding requires further backup.

By 14 weeks of age, infants consistently prioritize internal components of faces like the mouth and the eyes over external features like hair and the outline (Hunnius & Geuze, 2004). From 3 to 9 months, an increasing preference to look at faces from their own ethnic group was demonstrated in Caucasian (Kelly et al., 2007) as well as Chinese infants (Kelly et al., 2009). This finding backs up the idea that our perceptual systems fine-tune their sensitivity to optimally detect stimuli which are frequently encountered in one's own cultural environment (Kelly et al., 2005). The refinement of social attention and cognition during childhood is documented for several of its aspects (Corkum & Moore, 1998). For instance, infants as young as 6 to 18 months will follow adults' gaze direction, but neglect the possibility that the adult may be referring to a point of interest located beyond the child's field of view (Butterworth & Cochran, 1980) — a cognitive capability which develops later. To accommodate both social gaze in newborns and alterations during development, it was suggested that infants have an intrinsic and congenital preference to attend to face-like stimuli, which, in normal environments, results in a high probability that faces actually are gazed at, and consequently in an abundance of opportunities to learn about the visual characteristics of faces (Morton & Johnson, 1991).

Of course, while attention towards other humans lays the foundation for social development, it is higher cognitive functions that essentially constitute social life. Such higher processes, for which basic attentional mechanisms constitute an essential premise, undergo a parallel developmental process which could be demonstrated in several intriguing studies. In a study by Zeifman, Delaney, & Blass (1996), one month-old infants were calmed by sweet taste, but only if they simultaneously received eye contact. Hains & Muir (1996) found that three-month old infants were more likely to smile after eye contact with an adult. In children aged 6 months, gaze following behavior can be initiated by first establishing mutual gaze (Senju & Csibra, 2008), and at about 9 months of age, children themselves begin to point towards objects to initiate joint attention (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998). The present work, however, does not aim at covering the entire abundance of phenomena that make up the social world and the vast scientific field around it, but will mainly focus on basic mechanisms of social attention.

Alterations of Social Attention in Clinical Conditions

While there is undoubtedly a great variability in what can be called *normal* social behavior and development, research and experience of practitioners suggests that this area of living is quite often altered in a variety of psychiatric conditions, sometimes becoming an own source of psychological burden and grief. Two conditions in which alterations of social functioning have been described as especially pronounced and characterizing are social anxiety disorder (SAD) and autism spectrum disorder (ASD).

Social Anxiety

Social anxiety, the nervousness in or fear of social situations, is common in persons diagnosed with anxiety disorders or substance use disorders (Schneier, Blanco, Antia, & Liebowitz, 2002; Schneier, Johnson, Hornig, Liebowitz, & Weissman, 1992), but also in the general population (Stein, Walker, & Forde, 1994). In severe cases, when symptoms of anxiety in social situations — such as speaking in public or to persons in authority — cause considerable distress and impediments in daily life, the person concerned may receive a discrete diagnosis of social anxiety disorder (SAD) or social phobia (American Psychiatric Association, 2013). SAD becomes more frequent with increasing age (Rapee & Spence, 2004). In Europe, lifetime prevalence was estimated between 7% and 13% (Fehm, Pelissolo, Furmark, & Wittchen, 2005), making it the third most prevalent mental disorder after depression and substance abuse (Kessler et al., 2005; Wittchen & Fehm, 2004).

SAD has been linked to impairments in developing intimate and peer relationships, but also in academic achievements, employment opportunities as well as financial independence (Stein & Stein, 2008). Importantly, SAD was furthermore found to precede additional psychiatric conditions such as substance use disorders and depression (Ruscio et al., 2007). In epidemiological studies, females were shown to be at a higher risk of meeting criteria for SAD (Essau, Conradt, & Petermann, 1999), albeit consequences seem to differ between men and women. In studies in Western societies in the 1980s and 1990s (Caspi, Elder, & Bem, 1988; Rapee, 1995), socially anxious men tended to achieve less occupational and marital stability, whereas socially anxious women did not experience less marital stability, but tended to follow conventional life concepts and a role in homemaking instead of pursuing a professional career. The finding that social anxiety is more prevalent in the general population in women compared to men furthermore stands at a contrast with observations of roughly equal proportions of both genders in clinical settings, and has been linked to the possibly more severe impact on life quality in men (Rapee, 1995). Such findings of course need to be established individually for each culture and time in history, as they may interfere with role models and their alterations in different societies.

The development of SAD was found to be partly linked to genetic roots, with heritability estimated between 39% and 64% in an identical twin adoption study (Kendler, Karkowski, & Prescott, 1999). The onset of social anxiety was reported to be before the age of 18 (mean: 10 to 13 years) in the majority of cases (Nelson et al., 2000; Otto et al., 2001), with an earlier onset linked to the development of the more severe generalized subtype later in life (Wittchen, Stein, & Kessler, 1999). Moreover, an onset of SAD in adult life that is not linked to previous social anxiety during childhood or adolescence is actually very rare, and the few cases usually appear in the context of a major depression or panic disorder (Neufeld, Swartz, Bienvenu, Eaton, & Cai, 1999). If untreated, SAD was found to be a highly stable condition, often developing a chronic course (Fehm et al., 2005).

Biases in Processing of Social Information in Social Anxiety

Several studies found a tendency in socially anxious persons to interpret ambiguous social information negatively (Huppert, Pasupuleti, Foa, & Mathews, 2007; Mellings & Alden, 2000; Rapee & Lim, 1992; Veljaca & Rapee, 1998). Moreover, socially anxious persons felt to be watched by a conspecific within a larger range of gaze (e.g. even when not precisely hitting the socially anxious person (Gamer, Hecht, Seipp, & Hiller, 2011; Schulze, Renneberg, & Lobmaier, 2013)). These tendencies to confirm to oneself negative, but possibly unfounded assumptions about the social world are believed to maintain the disorder (Stein, 2006).

Several lines of empirical research have attempted to more precisely pin-point deviations in attention and cognitive functions in general in socially anxious persons. Similar to persons with other anxiety disorders, socially anxious individuals were found to exhibit an initial attentional bias towards threatening cues (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & IJzendoorn, 2007). For instance, social phobics show increased vigilance towards angry faces (Mogg, Philippot, & Bradley, 2004; Seefeldt, Krämer, Tuschen-Caffier, & Heinrichs, 2014) and threatening words (Spector, Pecknold, & Libman, 2003), but also to the eyes of faces in general (Boll, Bartholomaeus, Peter, Lupke, & Gamer, 2016). This phase of hypervigilance may then be followed by an avoidance of such stimuli (Amir, Foa, & Coles, 1998; Bögels & Mansell, 2004; Garner, Mogg, & Bradley, 2006; Schofield, Inhoff, & Coles, 2013; Wieser, Pauli, Alpers, & Mühlberger, 2009), although evidence for avoidance is not entirely consistent (Boll et al., 2016; Gamble & Rapee, 2010; Seefeldt et al., 2014). Information processing in socially anxious is, however, altered not only at the level of attention, but also in their memory, as shown in a preference to memorize negative social cues (Lundh & Öst, 1996).

Interestingly, gaze behavior of socially anxious persons was not consistently found to resemble clinical reports when studied in laboratory situations. While persons high on social anxiety are often described to fear and avoid gaze contact in clinical settings (Baker & Edelmann, 2002; Schneier, Rodebaugh, Blanco, Lewin, & Liebowitz, 2011), only some studies examining socially anxious persons' gaze behavior when confronted with images of persons found reduced amounts of gaze on depicted faces (Moukheiber et al., 2010; Weeks, Howell, & Goldin, 2013). Other studies, on the other hand, even found enlarged amounts of gaze on faces in socially anxious persons (Brunet, Heisz, Mondloch, Shore, & Schmidt, 2009; Wieser et al., 2009).

While basic research on interpretation styles of social information often utilizes every-day tasks like talking to other persons or remembering actual social situations, studies on attentional biases in social anxiety mostly rely on passively viewing impoverished images of persons. Only a small number of studies investigated viewing behavior in naturalistic situations, with mixed results. Baker & Edelmann (2002) found decreased eye contact in social phobics as well as subclinically socially anxious persons compared to controls, whereas Hofmann, Gerlach, Wender, & Roth (1997) found no relation between anxiety and gaze behavior. In these studies, however, viewing behavior was evaluated by manually extracting estimated gaze direction from video recordings instead of using more precise eye tracking methodology. Using a mobile eye-tracker in a real dyadic situation, Hessels, Gijs A. Holleman, Cornelissen, Hooge, & Kemner (2017) found persons with SAD traits to show similar gaze behavior as in experiments employing pictures or videos, i.e. hyperscanning followed by gaze aversion. However, data on viewing pattern of persons with social anxiety traits or SAD in real-world-situations remains scarce.

There is also evidence suggesting that socially anxious do not only show biases in attention, making them vulnerable to illusory perceptions of scrutiny, but often do show tangible deficits in social competence, and are sometimes perceived as behaving somewhat oddly in social interactions (Baker & Edelmann, 2002; Lange, Rinck, & Becker, 2014; Moscovitch & Hofmann, 2007; Schneider & Turk, 2014). Socially phobic children were rated to be less socially competent not only by themselves, but also by their parents and peers, and, possibly as a consequence, received positive reactions from peers less frequently than other children (Spence, Donovan, & Brechman-Toussaint, 1999).

Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a syndrome composing deficits in social behavior, language, restrictive interests, and repetitive behaviors (Geschwind, 2009). Prevalence rates have been subject to debate, with current estimates ranging from 0.6% (French, Bertone, Hyde, & Fombonne, 2013) to 2.6% (Kim et al., 2011), and a male-to-female ratio estimated to be about 4:1 (French et al., 2013). Autistic traits are also believed to be distributed among the general population, with a continuous transition between sub-clinical and clinically relevant manifestations, and slightly higher average expression in men than in women (Baron-Cohen, Hoekstra, Knickmeyer, & Wheelwright, 2006). It is assumed that the condition is congenital, livelong and highly heritable (Constantino et al., 2012). To date, this heritability could not be attributed to individual genes (instead, several hundred genes were found to be associated with ASD by a small margin), which may in part be explained by the phenotypical heterogeneity of the syndrome (State & Sestan, 2012).

As part of the deficits in social behavior, reduced eye contact has long been seen as a defining feature of the condition (Kanner, 1943; Volkmar, Lord, Bailey, Schultz, & Klin, 2004). This phenomenon could partly be reproduced in passive-viewing laboratory studies (Senju & Johnson, 2009), with some studies, on the other hand, finding no difference between persons with and without ASD (Rutherford & Krysko, 2008). Using a dual eye-tracking setup, Hessels et al. (2017) found that persons high on autistic traits avoided direct eye contact during a real conversation. Likewise, when Freeth & Bugembe (2018) tracked gaze in persons with ASD as well as neurotypical controls during a face-to-face conversation, they found a relative avoidance of gaze at the conspecific mostly while she was looking at the participant, i.e. an avoidance of mutual eye contact. Hayes & Henderson (2018) could trace even more subtle differences between persons of high and low autistic traits with a machine learning algorithm which explained a third of the variance in autistic traits based on gaze data of passive viewing of real-world scenes.

Differences between persons with ASD and neurotypical persons could also be identified early in childhood. While attention to eyes is robustly present in most 2-year-old infants, it was found to be significantly reduced in infants of the same age with ASD (Jones, Carr, & Klin, 2008), and lack of gaze towards the eye region was found to predict a later diagnosis of ASD even at the age of 6 months (Jones & Klin, 2013). Among children diagnosed with ASD, Murias et al. (2017) found that time spent looking at the actor in a video was positively correlated with several measures of social communication skills. Summing up, there is ample evidence for reduced eye-contact, a subdivision of overt social attention, in persons with ASD, which is likely involved in deficits in higher social cognitive functioning.

The Human Visual System

The present work will mostly focus on basic attentional phenomena rather than higher cognitive functions, and will touch upon the relationship between such basic behavioral measures and persons' traits. While individuals may vary on how they distribute their (visual) attention in social situations, their visual systems nonetheless share certain basic characteristics which one should bear in mind even when making comparisons across groups.

The Ventral Stream and the Dorsal Stream

Research on visual attention needs to be informed by the general architecture of the human visual system, which, perhaps counter-intuitively, is organized along two largely separate routes of information processing (Goodale & Milner, 1992; Milner & Goodale, 2008; Schneider, 1969): When exiting the occipital lobe, visual information is projected onto two separate pathways: a comparatively slow ventral pathway in which items are identified and processed semantically ("what"-pathway) and a faster dorsal pathway processing an object's spatial location and possibilities of interaction ("where"-pathway). Consequently, information processing in the dorsal pathway builds on binocular visual input which is required for precise representation of the three-dimensional configuration of one's surroundings in the near field. Although the simple separation of visual functioning into two streams has been challenged by more complex accounts (Cardoso-Leite & Gorea, 2010), its basic principle is still widely accepted and exemplified by a range of stunning findings. For instance, some patients with damage to the parietal lobe are no longer able to reach towards visual targets with their hands, although they still report seeing them (Damasio & Benton, 1979; Perenin & Vighetto, 1988). In reverse cases, a patient with damage in the lateral occipital region was still able to accurately move their hands and fingers towards an object, but failed to describe some of its central properties (Goodale, Milner, Jakobson, & Carey, 1991). A reassessment of her abilities confirmed this dissociation, although it found it to be smaller than originally thought (Himmelbach, Boehme, & Karnath, 2012).

More evidence for the existence of two separate streams of information processing for visual perception and action comes from experimental data with healthy persons. In an intriguing experimental setup, Aglioti, DeSouza, & Goodale (1995) recorded grasping movements towards circles in the Titchener circles' illusion. In these size-contrast-ambiguous scenes, target circles are surrounded by additional circles of varying sizes. Although subjects reported target circles of equal size to be smaller with increasing size of its surrounding circles — a typical reaction to this optical illusion — calibration of the grasp movement was refractory to the perceptual illusion and instead determined by the circle's true size. Dyde & Milner (2002) expanded this finding by showing that the simultaneous tilt illusion, which is generated in earlier visual regions V1 and V2 and before the visual stream is divided into the dorsal and the lateral pathway, affected both action and perception. The rod-and-frame illusion, which is thought to be generated in the ventral pathway, on the other hand, only affected perception. More detailed analyses questioned the extent to which spatial representations for vision-for-action and vision-for-perception can be seen as separate. Interestingly, vision-for-action was found to be more robust against the Müller-Lyer perceptual illusion when participants were required to point towards an object (Bruno, Bernardis, & Gentilucci, 2008) compared to when grasping it (Bruno & Franz, 2009), hinting to selective interconnections between the two visual systems.

Priority Maps

Although a body of evidence has helped to outline which brain regions constitute the visual system and how its different components can be characterized functionally, one may additionally ask just how information is processed at the individual stages. Several authors have proposed more refined models how the human brain selects the most relevant pieces of visual information and drops or partly neglects less relevant information to arrive at a meaningful interpretation of the visual world (Baluch & Itti, 2011; Fecteau & Munoz, 2006; Serences & Yantis, 2006). As a general theme, most theories encompass the idea of a priority map, which represents the visual field at more or less abstract levels and highlights areas which are to be given prioritized access to subsequent information processing.

Several brain regions have been linked to the creation of such a priority map after neural activity in these regions were found to dynamically mirror the locus of visual attention: the frontal eye field (Serences & Yantis, 2006), the posterior intraparietal sulcus and the precentral sulcus (Jerde, Merriam, Riggall, Hedges, & Curtis, 2012), the lateral intraparietal cortex (Bisley & Goldberg, 2010; Gottlieb, Kusunoki, & Goldberg, 1998), but also regions located earlier in the visual system such as V4 (Mazer & Gallant, 2003) and V1 (Sprague & Serences, 2013). With such a variety of brain regions associated with structuring visual input into more and less relevant pieces, a major task lies in the description of how each region takes part in this process. It is assumed that the earlier visual system primarily reflects the physical, low-level saliency of stimuli, wheras higher-order areas increasingly target at identifying the behavioral relevance of visual input (Baluch & Itti, 2011; Fecteau & Munoz, 2006). With regards to the term *priority map*, the reservation must be made that studies in the field typically do not actually map the processing of individual stimuli or individual parts in the visual field to subregions in the brain regions under investigation (i.e. construct a topographic profile describing the receptive fields of individual populations of neurons), but usually document cohesions between attention and neural activity in a region as a whole. This lack of detail in current data is due to relatively low spatial and temporal resolution of brain imaging techniques such as fMRI, and the limited spatial window provided by single cell studies (Fecteau & Munoz, 2006). One attempt to more precisely map the neural activity in individual voxels to changes in attention found that stimulus representation in higher visual areas is enhanced, but the size of receptive fields is not modulated by attention (Sprague & Serences, 2013). Mo, He, & Fang (2017) succeeded in constructing a topographic representation of complex natural stimuli (images of faces) in the early and still retinotopic primary and extrastriate visual cortices (V1, V2 and V3) by modelling the differences in first saccadic eye movements after stimulus onset of correctly displayed versus scrambled faces, and predicting these differences with voxelwise brain activation data. Taken together, the hypothesized priority maps of visual attention could be traced in several brain regions representing several layers of information processing in the visual system, but data on the precise topographic profile of individual populations of neurons remains scarce.

Visual Saliency

It was proposed that the bottom-up processes, which collectively embody one aspect of the hypothesized priority maps, encompass the detection of a variety of physical features in the field of view, among them color, orientation, luminance and motion, and bias attention towards regions in which these features are most prominent (Veale, Hafed, & Yoshida, 2017). As these features, unlike many determinants of top-down processes, can be described computationally in a comparably straightforward manner, attempts were made to explicitly model bottom-up processes via algorithms which are to some extent functionally equivalent to corresponding processes in the brain. Such models typically extract a variety of low-level features from images, condense these features into a *saliency map* and allow to use these maps as direct predictors of gaze behavior in humans watching these images. Over the last twenty-five years, a plethora of algorithms was developed to perform this task, and their performance in predicting human gaze data is constantly pitted against each other (the perhaps most catchy synopsis of this competition can be found at http://saliency.mit.edu/results_mit300.html). Some of the algorithms literally detect contrasts in feature channels like color, orientation, luminance and motion, combine these features to create the saliency map, and consequently allow for a palpable interpretation of the map's signification (e.g. Itti, Koch, & Niebur, 1998; Harel, Koch, & Perona, 2007; Koch & Ullman, 1987). The descriptive accuracy of saliency maps is then sometimes enhanced by including a person's tendency to avoid gazing at regions which were previously gazed at (*inhibition of return*, (Itti & Koch, 2001; Klein, 2000)). Some of the newer algorithms adopt more flexible *deep neural network* algorithms which employ convolutional filters trained on object recognition tasks (e.g. Kummerer, Wallis, Gatys, & Bethge, 2017; Kruthiventi, Ayush, & Babu, 2017). These algorithms represent some of the best-performing instances of saliency detection in terms of gaze data prediction. However, since the algorithms are not manually constructed, but trained with no or very little supervision, the exact procedure employed to arrive at a saliency map can no longer be easily identified; in fact, they may even contain fragments of object or face recognition processes at an extent which is not trivial to quantify, hampering the claim that they represent models of bottom-up visual processes.

All in all, while even the metrics to determine the predictive accuracy of saliency models are a matter of debate (Kümmerer, Wallis, & Bethge, 2015), the field can be characterized as a successful endeavor: Over the past twenty-five years, gaze data prediction has constantly improved, and several algorithms now predict novel gaze data better than the gaze behavior of another, randomly selected person. However, to date, no algorithm outperforms the prediction derived from aggregating gaze data obtained from many human beings.

It is still an ongoing debate how bottom-up and top-down processes interplay in guiding visual attention, and which processes are more dominant in which situations. Anderson, Ort, Kruijne, Meeter, & Donk (2015) found that when viewing complex images, the first saccade after stimulus onset is more likely to target an area characterized by high visual saliency the sooner it occurs, substantiating the idea that the processing of visual saliency is a fast route to oculomotor control, which may be overridden by additional processes at later stages.

More generally, however, although authors of saliency algorithms typically do not claim that modelling gaze behavior via saliency algorithms can provide an exhaustive description of human visual attention, several contributions have observed a tendency in this research community to exaggerate the influence of bottom-up processes and neglect or downplay the influence of top-down processes (Birmingham et al., 2009b). In fact, several studies could significantly improve gaze prediction by including predictors at the object-level (Torralba, Oliva, Castelhano, & Henderson, 2006) or the semantic level (Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014), or by including faces as predictors (Cerf, Frady, & Koch, 2009; Cerf, Harel, Einhäuser, & Koch, 2008; Parks, Borji, & Itti, 2015). Moreover, it was documented that areas in an image containing social information (Birmingham et al., 2009b; End & Gamer, 2017a) or, more generally, semantic richness or *meaning* (Henderson & Hayes, 2017), predicted gaze to a greater extent than visual saliency. To sum up, while modern saliency algorithms do predict gaze behavior with a decent accuracy in some situations (especially complex, but non-social images and videos), their performance in other, especially social situations continues to be a subject of debate.

Ecological Validity

A central request in psychological methodology is to generate testing conditions which allow for justified generalizations towards a phenomenon *in general* as opposed to its instantiation within a specific experimental setup (in the current diction, the experiment should be *ecologically valid*). Although a seemingly obvious requirement, there exist ongoing debates around its precise definition and importance. The matter may be especially important when studying social attention, as this process often takes place in interaction with others and can not easily be dissociated from its context.

An Historical Account

Since the beginnings of the psychological sciences, the idea of ecological validity and its precise conceptualization have been the cause for heated debates, some confusion around related terms and even hurt feelings among spokespersons of neighboring disciplines. Over the years, several lines of research investigating psychological phenomena in laboratory conditions were criticised for overly prioritizing *internal validity* at the cost of ecological validity. Although, of course, no researcher will object the idea that internal validity represents another requirement for good research, its one-sided preference was linked to a failure in understanding a behavior's meaning in its natural habitat or in eliciting the full range of psychological phenomena as they naturally occur. For instance, Brunswik (1943) argued that the behavior of organisms is so fundamentally defined by the complex environments which set the stage for any meaningful action that the search for uniform laws defining behavior at a general level may be an insentient one. Note that although Brunswik coined the term *ecological validity*, it was originally used to express the degree of relation between two cues within an experiment, and strictly kept apart from the idea of experimental designs which are *representative* of or *generalizable* towards other situations. Nonetheless, researchers from the 1970s until today commonly use the term *ecological validity* to refer to the latter (Bandura, 1978; Bronfenbrenner, 1977), and the present work will follow this perhaps somewhat inept change in naming convention. Hull (1943), in a direct debate carried out in *Psychological Review*, argued against Brunswick's view, and advocated the possibility of isolating individual laws in behavior when properly accounting for environmental factors. These two opposing views were reenacted, differentiated and fleshed out with descriptive examples in the decades to come. As one instance, Jenkins (1974) went so far as to characterizing the time in his earlier career, when he believed that complex behavior could generally be understood as an assembly of individual building blocks, as a time when he was “caught in a metatheoretical trap”. Elms (1975) noted that many findings in social psychology failed to be reproduced in conceptual replications, since they were bound to variables which are specific to a certain, well-controlled experimental setting — an objection that may remind the reader of a current debate (Gilbert, King, Pettigrew, & Wilson, 2016; Open Science Collaboration, 2015). Cronbach (1975) argued for a more detailed analysis of *who* is being tested in an experiment, and how persons’ aptitudes may interact with the testing condition to produce the observed outcomes. This point was later extended by Henrich, Heine, & Norenzayan (2010) to suggest that experimental psychology should attempt to represent the entire variability of humans in their participant groups to justify broad generalizations.

Along the same lines, a discussion among the newly crafted field of *cognitive science* involved the question of whether and how studies on culture should be incorporated into the overall picture which was to be drawn of human cognition. Norman (1980) argued that a person’s belief system, or culture, constantly colors the internal cognitive processes and therefore cannot be reasonably omitted. Nonetheless, Gardner (1987) reports a marginalization of culture, history and context in the investigation of cognitive processes, and complains that relatively little work dealt with the influence of these boundary conditions on cognition. Hutchins (1995a) pointed out that cognitive processes need not be understood as something happening inside an individual; rather, it is sometimes necessary to choose an entire system consisting of a person and the environment as the unit of analysis in order to adequately retrace the outcome of a situation (an idea which has thrived among clinical practitioners at least since the 1930s (Gurman & Fraenkel, 2002)).

As one example, Hutchins (1995b) describes the situation in an airplane cockpit in which two persons and several technological devices need to interlace their information processing to purposefully store relevant information and to, literally, arrive at navigating towards a destination. Gigerenzer & Brighton (2009) argue that only a careful description of environmental factors allows to normatively evaluate a behavior or strategy as *rational*, and use the term *ecological rationality* to establish a border between their environmentalist approach and other, environment-agnostic definitions of rationality.

Many of these examples originate from different times in history and fields of study and target different aspects of scientific methodology. Nonetheless, they represent reiterations of a general theme which runs through the history of the behavioral sciences as a red thread: the claim that some attempts to obtain generalizable knowledge simply by holding variables constant do not work, but that researchers must instead investigate cognitive phenomena in a variety of contexts and take into account the variability in human aptitudes and their interaction with the environment. This point is perhaps most laconically stated by Beller, Bender, & Medin (2012), p.346: "[...] by virtue of being cognitive scientists, many researchers feel that they may justly be interested in generalizable findings only and hence delegate any exploration of diversity to the fringes. But generality must be demonstrated, not assumed." The present work generally follows this view, and seeks ways to explore mechanisms of social attention under naturalistic conditions. Principles of strict standardization, as evident in studies employing highly impoverished stimuli with little resemblance to what persons encounter in their everyday life, are loosened for the benefit of higher representativeness. Note that this rationale is further supported by methodological considerations that standardizing experimental situations beyond the variables of interest may result in heightened sensitivity, but degraded reproducibility (Richter, Garner, Auer, Kunert, & Würbel, 2010; Richter, Garner, & Würbel, 2009; Würbel, 2000). Along these lines, Dhami, Hertwig, & Hoffrage (2004) summarize how phenomena like the overconfidence in one's judgements can be elicited using a set of standardized but unrepresentative questions, but may fail to replicate when a different or more representative set of questions is used.

Ecological Validity in Social Attention Research

Although applicable to all behavioral sciences (and science in general, for that matter), the debate about adequate testing conditions or ecological validity has also prominently culminated in social attention research in recent years. The majority of research on social attention employs static images which are shown to participants in isolation, sometimes sitting passively and sometimes with an instruction to react to certain stimuli with a button press or mouse click. While research carried out in such an environment has undoubtedly helped to understand important basic mechanisms in human attention including social attention, several researchers have bluntly added that "social cognition is fundamentally different when we are in interaction with others rather than merely observing them" (quote from Schilbach et al. (2013), p. 393, but similar points made by Gobel, Kim, & Richardson (2015) and Risko, Richardson, & Kingstone (2016)). More specifically, it is argued that our eyes serve two main functions in social situations — collecting information *and* signaling one's own mental states and intentions to conspecifics. Importantly, mechanisms modulating gaze direction in order to give consideration to the latter are beyond the scope of research in which stimuli — e.g. static images — cannot react.

Although this idea is by no means new and was discussed at detail in the decades before (Argyle & Cook, 1976; Goffman, 1963; Kleinke, 1986), it has only relatively recently made its way into biopsychological experiments employing eye-tracking or measures of the central or peripheral nervous system. And indeed, research investigating human social behavior *in the wild* has revealed interesting mechanisms which seem to operate below the radar of standard laboratory paradigms. Participants in a study by Kuhn, Teszka, Tenaw, & Kingstone (2016) either watched a video of a magician performing a trick or saw the same scene live in front of them. Here, participants spent more time looking at the magician's face when watching the video compared to seeing the person in reality. Foulsham, Walker, & Kingstone (2011) tracked gaze of participants who were either walking on a University campus or viewing the same scene on a monitor, and found participants in real-life to avoid looking at near pedestrians compared to participants in the laboratory. Laidlaw, Foulsham, Kuhn, & Kingstone (2011) compared a situation in which participants had the opportunity to gaze at a confederate who was present in the same room with a situation in which the same participant could be seen on a video monitor, and found that participants frequently looked at the participant when shown on a monitor, but avoided looking at her when present in the same room. Laidlaw, Rothwell, & Kingstone (2016) found that participants were less likely to look at a stranger who said "Hey" into a phone in a public situation while moving her hand in a greeting wave compared to the same stranger waving her hand and saying "Hey" without holding a phone. Here, the presence of a cell phone constitutes a social cue which allows to differentiate between a situation in which the conspecific's utterance is directed towards the participant from a situation in which the utterance is directed elsewhere. Since the conspecific's utterance occurred at a point in time when she was not gazed at by the participant, the authors conclude that participants must employ a form of covert attention to classify the situation into a public one (in which the conspecific has no phone and might be addressing the participant) or a private one (in which the conspecific is most likely to address another person via a phone call). Moreover, the authors assume that covert attention may be employed to guide social looking behavior in socially accepted ways — a claim that can only reasonably addressed in real social situations. Pointing towards another mechanism to regulate interpersonal closeness, several studies found participants to literally increase personal space towards a conspecific while engaged in mutual direct gaze (Hayduk, 1981; Patterson, 1976, 1982).

At the same time, research employing "standard" laboratory settings rather than observations in real-life have also been able to further illuminate the importance of context as well as culture in several studies by adopting elaborate experimental setups. Yin, Xu, Duan, & Shen (2018) found that viewing two hands in an orientation suitable for a handshake — but not the same hands in a different orientation — are perceived as one entity, with cueing effects transferring from one hand to another, indicating the the social meaning of a configuration exerts a top-down influence on basic attentional mechanisms. As an example for cultural modulation of social attention, Cohen, Sasaki, German, & Kim (2015) displayed images of several faces, one in the foreground and several in the background. European Americans showed a gaze cueing effect only towards the face in the foreground, while participants from East Asia additionally showed a gaze cueing effect towards faces in the background, corrobating the idea of a wider scope of social attention in more interdependent cultures (represented in several Asian societies) compared to more independent cultures (represented in Western societies). Likewise, Lee, Greene, Tsai, & Chou (2016) showed Taiwanese and American participants the image of a face which was flanked by other faces, and found Taiwanese participants to spend more time looking at the flanking faces compared to the American participants, suggesting that persons from the two cultures employ different strategies for attention allocation. Gobel et al. (2015) had participants watch videos of people of higher or lower social rank, with one group believing that the depicted people would later be watching them in return, and a second group believing they would not be watched by the targets at a later point. In the former group, participants avoided looking at the eyes of higher ranked compared to the lower ranked individuals, while the pattern was reverse in the latter group, indicating that social norms may become active even at the prospect of being watched and judged by other persons.

Moreover, it should also be noted that the idea of ecological validity does not imply that situations which are not sampled from a representative distribution of real-life situations necessarily do not allow for broader generalizations. Instead, the notion merely states that the legitimacy of generalizations must first be demonstrated. Importantly, several attentional mechanisms relevant to social cognition may be relatively situation-agnostic and operate in a similar manner both in real-life and when seeing images of persons. For instance, Peterson, Lin, Zaun, & Kanwisher (2016) found that a preference to first look at the eyes or the mouth when seeing a face — a preference shown to be quite stable within individuals when presented different images of faces on a screen — is similar between a laboratory and a real-life situation, substantiating the claim that face processing may occur in naturalistic manner even when viewing images of instead of real human faces.

To sum up, although research employing static images or videos of social scenarios does allow to address interesting phenomena in the field of social attention (and should not be rejected prematurely as *ecologically invalid*), the present work follows the idea that investigating phenomena in natural, real-life situations represents the gold standard in social attention and social cognition research. The following studies strive to make experimental setups more and more ecologically valid while tweaking the level of internal validity, and hope to contribute to the question of what parameters are essential in establishing ecological validity in this field.

Virtual Reality in Social Attention Research

Virtual Reality (VR) is now used as a tool in various fields of psychological research and is becoming increasingly popular and easy to adopt. Proponents of this technology argue that it may solve the conflict between internal and ecological validity by immersing participants in realistic, yet highly controlled virtual situations. For social attention and social cognition research, however, researchers are confronted with the methodological challenge of creating believable virtual humanoid agents.

Virtual Reality in Psychological Research

The use of virtual reality in psychological research has seen a meteoric rise throughout the last two decades and has been employed for such diverse tasks as sensorimotor training in neurorehabilitation (Adamovich, Fluet, Tunik, & Merians, 2009), exposure therapies for anxiety disorders (Opriş et al., 2012) or as a research tool to study body perception and bodily consciousness (Blanke, 2012; Slater, Spanlang, Sanchez-Vives, & Blanke, 2010), motor control (Patton, Dawe, Scharver, Mussa-Ivaldi, & Kenyon, 2006) and spatial cognition (Pine et al., 2002). Until recently, performing VR research required purchasing costly hardware and specialized software. In 2016, the market entry of novel VR headsets such as the Oculus Rift or HTC Vive provoked a dramatic price drop, while at the same time improving performance in terms of image resolution, field of view and time lag between head movement and display reaction. Concurrently, major 3D engines such as Unity 3D (Unity Technologies) and Unreal Engine (Epic Games, Inc.) now natively support VR functionalities, making the development of VR experiments feasible for smaller research teams. Nonetheless, technical hurdles still abound: researchers are required to learn the basics of 3D modelling as well as engine logic and scripting, often still preventing this method from a more widespread use (Parsons, 2015). Proponents of the use of virtual reality in psychological research often present the sense of *presence* — the feeling of being located inside of a scenario as opposed to watching it from the outside — as an important benefit over watching scenes (or even interactive simulations) on a computer screen. At the same time, the use of virtual reality brings about novel challenges such as a sense of *simulator sickness*, which some participants may suffer from.

Presence

Although rather intuitive to most, various definitions flourish around the concept of *presence*. Most of them center around the idea of a "subjective experience of being in one place or environment, even when one is physically situated in another" (Witmer & Singer, 1998, p. 225), or, put simply, of "being there" (Skarbez, Frederick P. Brooks, & Whitton, 2017, p. 1). Presence, which is an inherently subjective measure, is usually contrasted against *immersion*, which describes the objective properties of a VR system in recreating sensory stimuli (Slater, 1999). In VR research, some degree of immersion is usually seen as a prerequisite for the installment of presence, although many other factors play a role in establishing presence and there need not even be a clear relationship between the two (and of course, humans are also capable of experiencing presence in their dreams or imaginations, when there is no immersive system involved). Although there has been some disagreement about the term *immersion* and its distinction from *presence* (Biocca & Delaney, 1995; Lombard et al., 2000; Witmer & Singer, 1998), the present work will follow the definition stated above.

A large part of the work on presence focused on establishing explicit measures of self-report (mostly questionnaires) to represent the concept and different subcategories (e.g. Witmer & Singer (1998), Usoh, Catena, Arman, & Slater (2000) and Schubert, Friedmann, & Regenbrecht (2001); for an overview and comparison see Skarbez et al. (2017)). Some studies, on the other hand, have attempted to detect physiological or neural correlates of presence. The advantage to this approach is that, if successful, it would allow for real-time measurements along the course of an experiment with no need to repeatedly pose a battery of questions to participants. Several studies found that presence led to higher levels of physiological arousal if some sort of threat was presented in the virtual scenario (Meehan, Insko, Whitton, & Brooks, 2002; Slater, Brogni, & Steed, 2003; Slater et al., 2006). However, another study (Dillon, Keogh, & Freeman, 2002) employing less arousing stimuli did not find a correlation between subjective presence and physiological metrics.

Likewise, when virtual reality was used to create the illusion of having a virtual body, the strength of the illusion was associated with higher skin conductance reaction towards a threat of the fake body (Kilteni, Normand, Sanchez-Vives, & Slater, 2012; Petkova & Ehrsson, 2008). Unfortunately, techniques incorporating threat can barely be seen as a continuous measurement of presence since they require to disrupt the ongoing experience in the virtual scenario. More importantly even, there exist ethical objections towards threatening participants during psychological experiments, even if the threat is only targeted towards a virtual body which the participant has taken ownership of (Madary & Metzinger, 2016). Attempts to find neural correlates of presence using fMRI have been inconclusive, with one study reporting increased activity in the dorsolateral prefrontal cortex (Baumgartner, 2008), and another reporting increased activity in the parahippocampus (Bouchard et al., 2012). Critically, however, current fMRI technology poses heavy restrictions on participants' body movements and even body posture, making this technology an unfeasible candidate for implicit presence measurements for the majority of use cases.

Another promising approach to retrace presence in participants as they are involved in a virtual scenario lies in the detection of behavioral markers. Although some authors speculate that behavioral signals might be found in "nearly any virtual scenario" (Skarbez et al., 2017, p. 32), attempts in this regard have been sparse. As one exception, Slater, Usoh, & Chrysanthou (1995) asked participants to point in the direction of a radio which they had seen in different places in the real and the virtual scene. Here, a relative bias for pointing towards the virtual versus real radio was found when the virtual scene was depicted at a higher visual acuity, rendering the claim plausible that such a behavioral marker constitutes a measure of presence.

Simulator Sickness

A problematic side-effect of experiences in virtual environments is the tendency in a considerable percentage of users to develop symptoms of simulator sickness (Johnson, 2005), e.g. nausea, oculomotor difficulties, vertigo or blurred vision. While sickness occuring in real situations like on boats or in rollercoasters may be explained by the sheer novelty of the motion cues (*postural instability theory*), sickness in simulators more probably originates from the discrepancy between visual and vestibular stimulation (*sensory conflict theory*; Cobb, Nichols, Ramsey, & Wilson, 1999; Johnson, 2005).

The problem should be tackled carefully, since it can be a severely negative experience for participants sometimes lasting for hours after the immersion in a VR scenario and has unknown side-effects on the physiological and behavioral level. For instance, the onset of subjective symptoms of sickness is often preceded by postural sway during stance (Stoffregen, Hettinger, Haas, Roe, & Smart, 2000), which constitutes a problematic confound if movement patterns are analyzed and reported in an experiment. One method to reduce the risk of simulator sickness is to harmonize optical flow patterns with the participants' own body movements, or, technically speaking, to always move the position of the participant's perspective in the virtual world in accordance to the actual head movement. Although simulator sickness has many facets, a single questionnaire named *simulator sickness questionnaire*, which covers the variety of symptoms quite exhaustively, prevails the literature as the standard instrument (Kennedy, Lane, Berbaum, & Lilienthal, 1993) and is used in the present work as well.

Virtual Humans as a Proxy for Real Persons

While virtual reality software has reached a level of maturity in many fields and allowed researchers to adequately address their research questions, its use in social attention research is often considered to be still in its infancy (Pan & Hamilton, 2018). For virtual reality to be used as an adequate model of reality, it can be argued that experimental software should confront participants with believable virtual humans — a task arguably more intricate than designing, say, believable virtual interiors or even mapping a person's body posture onto a virtual avatar.

As a theoretical ideal in this respect, one might suggest that virtual characters should be believable enough to be indistinguishable from real humans. This idea is probably most famously formulated in Alan Turing's *imitation game* (now mostly referred to simply as *Turing Test* (French, 2000; Turing, 1950)). Turing rephrased the question "Can machines think?" with the question of whether a software can communicate via text in such a way that it becomes impossible to identify it as software rather than a human being. Specifically, Turing proposed to test software for its capability to *think* by having an interrogator asking both the software and a human being questions in written natural language, with no limits on the subjects. Both the human and the software will answer these questions, also in written natural language, for five minutes each before the interrogator decides which of the two is the human being and which is the software. If the software reliably deceives interrogators into falsely choosing it as the presumed human being in more than 30% of the cases (a proportion Turing seems to have chosen rather arbitrarily), it has won the imitation game and, according to Turing, may be called *intelligent*. The exact conditions for passing the test have been adapted in various ways over the last decades, but the simple idea at its core has remained a hinge point in heated debates in the philosophy of mind (Block, 1981; Searle, 1980) and artificial intelligence (Moor, 2001), with an additional controversy around the question of whether or not working towards passing the Turing test is a fruitful endeavor for artificial intelligence in the first place (Harnad, 1992; Hayes & Ford, 1995).

Importantly for the goal of designing virtual characters which are indistinguishable from real humans, it must be noted that to date, no software has passed the Turing test. Moreover: Even if it had, there would still be a need to create embodied agents which are indistinguishable from real humans in their entire physical appearance, as opposed to software which merely produces text-based utterances in natural language (Ortiz, 2016). On the other hand, it may be argued that the theoretical ideal to create virtual agents which are indistinguishable from real humans is not only unattainable, but also unnecessary. Instead, it may just be sufficient to design humanoid stimuli which are capable of eliciting natural modes of perception and action in human participants. A central question in this respect is which kinds of stimuli do, and which do not stimulate such natural modes of information processing in participants. Although the effect of virtual agents on humans have only recently been studied experimentally, several studies have already pointed to the idea that confronting participants with animated virtual agents displayed in 3-dimensional scenarios may elicit perception and behavior similar to that in real-world situations. In one study (Sanz, Olivier, Bruder, Pettré, & Lécuyer, 2015) participants maintained interpersonal distance towards virtual avatars but not towards inanimate objects, seemingly conforming to social norms as they are usually observed among real persons. This effect was found to be enlarged in participants high on social anxiety (Rinck et al., 2010). In another study (Pan, Gillies, Barker, Clark, & Slater, 2012), participants showed an increase in skin conductance when approached by a virtual character, an observation which is also found when participants are approached by real persons.

The Present Work

The present work encompasses three individual studies, each attempting to add knowledge to particular aspects of social attention where open questions still prevail. Study 1 tests the predictive power of physical saliency and social information on gaze allocation when viewing dynamic videos, and investigates the stability of viewing preferences in this regard. Here, experimental standardization is loosened for the sake of higher generalizability by employing videos which range on an abundance of variables such as context, number of depicted persons or lighting conditions. A relatively novel statistical approach using generalized linear mixed models (GLMM) is used to handle gaze data in such a complex environment and carve out the contribution of individual feature channels to gaze allocation. We furthermore link gaze behavior to each video's emotional valence which were validated both by means of self-reports and physiological measures. This study has been published in a peer-reviewed scientific journal:

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

Study 2 takes gaze measurement to a real-life situation, surveying how social anxiety is associated with alterations in gaze behavior when walking in a public train station. In such an environment, experimental standardization is even more difficult to ensure, as each participant may be confronted with a different set of unplanned encounters. We rectified comparisons between individuals by means of a control condition in which matched participants viewed identical scenes on a computer monitor. This study has been accepted in a peer-reviewed scientific journal:

Rubo, M., Huestegge, L. & Gamer, M. (*in press*). Social anxiety modulates visual exploration in real-life — but not in the laboratory. *British Journal of Psychology.*

Study 3 attempts to balance experimental standardization and ecological validity by confronting participants with a social situation in virtual reality. Here, it was possible to manipulate individual variables (e.g. a virtual character's facial experession) while all other variables within a complex stream of events are held constant. Moreover, virtual reality technology allows to more precisely link behavioral data to events in the virtual social scene. For instance, it is possible to describe gaze behavior in relation to the interpersonal distance to an agent — a type of analysis which was performed in Study 2 using an approach which was more laborious, but somewhat less precise and objective compared to the approach used in study 3. This study is currently under review in a peer-reviewing journal:

Rubo, M., & Gamer, M. (2019). Natural Gaze Behavior and Stable Intra-individual Gaze Preferences in a Social Virtual Scenario with a Reactive Virtual Agent. Manuscript submitted for publication.

CHAPTER 2

Social Attention in the Laboratory

Social Content and Emotional Valence modulate Gaze Fixations in Dynamic Scenes

Marius Rubo, Matthias Gamer

Julius-Maximilians-Universität Würzburg

Corresponding Author: Marius Rubo

Email adress: marius.rubo@uni-wuerzburg.de

Published 2018 in *Scientific Reports, 8(1).* doi:10.1038/s41598-018-22127

Abstract

Previous research has shown that low-level visual features (i.e., low-level visual saliency) as well as socially relevant information predict gaze allocation in free viewing conditions. However, these studies mainly used static and highly controlled stimulus material, thus revealing little about the robustness of attentional processes across diverging situations. Secondly, the influence of affective stimulus characteristics on visual exploration patterns remains poorly understood. Participants in the present study freely viewed a set of naturalistic, contextually rich video clips from a variety of settings that were capable of eliciting different moods. Using recordings of eye movements, we quantified to what degree social information, emotional valence and low-level visual features influenced gaze allocation using generalized linear mixed models. We found substantial and similarly large regression weights for low-level saliency and social information, affirming the importance of both predictor classes under ecologically more valid dynamic stimulation conditions. Differences in predictor strength between individuals were large and highly stable across videos. Additionally, low-level saliency was less important for fixation selection in videos containing persons than in videos not containing persons, and less important for videos perceived as negative. We discuss the generalizability of these findings and the feasibility of applying this research paradigm to patient groups.

Introduction

Like most vertebrates, humans can only obtain a part of their visual field at a high acuity and therefore repeatedly move their eyes in order to construct a representation of their environment with sufficiently high resolution (Land & Fernald, 1992). Controlling gaze along with retrieving and filtering relevant signals from the environment is a central task of the attentional system (Desimone & Duncan, 1995). In the past, various lines of research have addressed the mechanisms driving such attentional control.

As sociability is one of human's key features (Adolphs, 2010a), a large body of research has assessed how we gather social information in order to infer other persons' intentions and feelings. For instance, it was shown that socially relevant features like human heads and eyes (Birmingham, Bischof, & Kingstone, 2008; Yarbus, 1967), gaze direction of depicted people (Borji, Parks, & Itti, 2014), people who are talking (Coutrot & Guyader, 2014) and people with high social status (Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010) attract attention when freely viewing images or dynamic scenes. However, non-social cues like text (Cerf et al., 2009; Ross & Kowler, 2013) and the center of the screen (Le Meur, Le Callet, & Barba, 2007; Tatler, 2007; Tseng, Carmi, Cameron, Munoz, & Itti, 2009) can also serve as predictors for gaze behavior.

Another line of research has focused on the predictive value of low-level image features such as contrast, color, edge density and, for dynamic scenes, motion. A range of algorithms exists to extract these features in images and videos and condense them into one low-level saliency value between 0 and 1 for each pixel, resulting in topographic low-level saliency maps(Kümmerer et al., 2015). Low-level saliency has been shown to explain fixation patterns for a variety of naturalistic and abstract images (Harel et al., 2007; Parkhurst, Law, & Niebur, 2002), as well as naturalistic videos (Carmi & Itti, 2006; Itti, 2005; Le Meur et al., 2007) and has been argued to be a biologicallyplausible model of early visual processing (Itti et al., 1998).

The influence of social stimuli and visual low-level saliency on eye movements have only recently been studied within the same datasets, and rarely in direct juxtaposition. During face perception, it was shown that facial regions diagnostic for emotional expressions received enhanced attention irrespective of their physical low-level saliency (Scheller et al., 2012). Birmingham and colleagues found social areas in an image to be a better predictor for fixation behavior than low-level saliency (Birmingham et al., 2009b; End & Gamer, 2017a). Other studies found faces to outperform low-level saliency on gaze prediction in dynamic scenes showing conversations between persons7 and documented higher predictive power for faces than for low-level saliency for adult participants watching a comic clip, although faces were not controlled for low-level saliency in this particular analysis (Frank, Vul, & Johnson, 2009). Several studies reported an improvement of low-level saliency-based models by including faces as predictors (Cerf et al., 2009, 2008; Parks et al., 2015).Xu and colleagues included a variety of predictors at pixel level (color, intensity, orientation), object-level (e.g., size, solidity) and semantic level (e.g., face, gazed-at objects, text) and found higher weights for the combined predictors at the semantic level than at pixel- and object-level (Xu et al., 2014).

Despite recent recommendations of increasing the ecological validity in social attention research (Kingstone, 2009), several studies utilized impoverished stimuli such as schematic depictions of faces that are typically stripped of context or background information (Scheller et al., 2012). While this research strategy can illuminate basic attentional principles, its results may not easily extrapolate to real-world attentional phenomena, where faces are only one feature among many competing for an observer's attention. Furthermore, most studies that do attempt to study social attention using contextually rich scenes typically do so using static images (Birmingham et al., 2008; Cerf et al., 2009; Fletcher-Watson et al., 2008; Xu et al., 2014). However, as motion is ubiquitously present in virtually all everyday situations and has been shown to be the strongest single predictor for gaze allocation (Itti, 2005; Mital, Smith, Hill, & Henderson, 2011), video stimuli seem advantageous when investigating social attention compared to static stimuli. Moreover, it was demonstrated that participants show more consistent eye movement patterns when viewing videos compared to static images (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Smith & Mital, 2013), thus indicating a potentially higher predictive value of basic stimulus properties on visual exploration.

In order to address these issues, the current study followed the cognitive ethology approach mentioned earlier (Birmingham et al., 2009a; Birmingham & Kingstone, 2009; Kingstone, 2009). We used uncut, dynamic scenes showing naturalistic situations with no artistic ambition. By incorporating both low-level features such as motion and social information into one analysis, we aim at further illuminating the determinants of visual attention under ecologically more valid conditions. Importantly, gaze data was analyzed using a generalized linear mixed model (GLMM) approach. This allows for estimating several features' unique contribution to fixation selection even in cases of co-variations between predictors. Specifically, this approach allows for crystallizing the effect of social information on gaze allocation even when, as it naturally occurs in real situations, depicted persons move or become visually salient in other respects. We hypothesized the performance of low-level saliency-based models to be poorer in social scenes compared to non-social scenes (End & Gamer, 2017a). Furthermore, we expected social information to be a significant predictor for gaze behavior, even when controlling for low-level saliency and centrality.

A second rationale behind employing contextually rich video stimuli is linked to, but partly independent from the concept of ecological validity: by deliberately omitting standardization of stimuli on many dimensions, we intended to identify only robust attentional effects which are independent of idiosyncrasies in the experimental setup. Several intriguing experiments have demonstrated heightened sensitivity, but degraded reproducibility as a result of strict experimental standardization in animal research (Richter et al., 2010, 2009), and the theoretical considerations employed to explain these findings (Würbel, 2000) seamlessly extend to human behavioral sciences. This idea is not entirely new to psychological experimentation, as documented by a general acceptance to leave plenty surrounding conditions unstandardized (e.g., time of the day, day of the week, participant's mood and appetite, weather, room temperature, room smell, experimenter mood, air pressure), even though they may be expected to produce effects in some circumstances. The video stimuli used in the present experiment extended this rationale by varying on a large number of dimensions: general semantics of the scene, composition, brightness, lighting, color, amount of movements, type of movements, direction of movements, camera movement, appearance or disappearance of objects or persons, attractiveness of persons, to name just a few. To our opinion, adopting a cognitive ethology approach not only encompasses investigating social attention under naturalistic conditions, but also assessing whether social attentional mechanisms become tangible when considerable amounts of external variance are at play, as one would expect in naturalistic situations. In order to estimate the robustness of attentional effects, we will not only estimate the predictive value of social attention and low-level saliency throughout the entire dataset, but also examine their intra-individual consistency along the various video stimuli.

As an additional experimental manipulation, we compiled the stimulus material such that video clips differed in their affective quality. It is a well-established finding that threatening stimuli (Öhman, Flykt, & Esteves, 2001; Wieser, McTeague, & Keil, 2011), but also emotional stimuli in general (Vuilleumier & Huang, 2009; Yiend, 2010) attract attention and are processed preferentially. The majority of studies in this field employed static stimuli with drastic differences in valence like images selected from the International Affective Picture System (IAPS) (Lang, Bradley, & Cuthbert, 1997). By contrast, and again along the idea of a cognitive ethology approach, the current study aimed at investigating whether emotional quality affects gaze allocation when viewing naturalistic videos, in which differences in perceived valence are within the range of what persons typically encounter in their lives. Recordings of autonomic nervous system activity were additionally obtained to confirm the affective quality ratings. We hypothesized social features that contribute to the affective quality of the stimulus to gain weight in predicting gaze allocation at the expense of the influence of low-level visual features. To the best of our knowledge, social attention has not been studied before within such a setup of naturalistic affective videos whilst statistically controlling for low-level physical low-level saliency.

Materials and Methods

Participants

Thirty-two participants (M = 27.84 years, SD = 7.46 years, 7 males, 23 students) took part in this study. The sample size was determined a priori to detect a medium effect size of $d = 0.50$ in a one-tailed paired comparison with a power of at least 0.85. No participant reported a history of psychiatric or neurological illness or taking centrally-acting medication. All participants had normal or corrected-to normal vision. The study was approved by the ethics committee of the German Psychological Society (DGPs) and conducted in accordance with the Declaration of Helsinki.

Stimuli

The participants viewed, in a randomized order, 90 complex naturalistic video clips of a duration of 20s each, depicting a variety of indoor (e.g., private homes, public buildings, public transport) and outdoor (e.g., streets, countryside, beach) scenes (for a description of some of these videos, see online supplement). Participants were not given any task or external motivation, but were instructed to freely view the scenes as though they were watching television. Sound was turned off in all videos. Forty-five of the video clips contained human faces and typically other body parts and were categorized as "social" (e.g., people walking in the streets or playing a ball game), while the remaining 45 clips did not show human beings (e.g., a train driving by, a scene in a forest). All videos were either obtained from publicly available online streaming services (e.g., www.youtube.com) or filmed by ourselves. We made sure not to use popular videos in order to reduce the risk of displaying a video to a participant who has viewed it before. No participant reported having seen any of the videos before when asked to disclose what had drawn their attention. Videos were required to depict situations that one could encounter in real life, as opposed to scenes that are primarily filmed for their artistic value. Moreover, we made sure that the persons appearing in the videos were unknown (i.e., no famous persons). Unlike impoverished stimuli sets often used, our video clips included a variety of visual information both in the back- and foreground, depicting a complex set of human actions and natural events. They were filmed with unpretentious camera movements and no cut. For the social as well as the non-social scenes, we made sure to include positive, negative and rather neutral clips. However, this a priori selection was only done to ensure sufficient variation in affective quality and the analyses were calculated using individual affective ratings of each participant. All clips had a resolution of 1280 × 720 pixels and a frame rate of 30 frames/s.

Apparatus

Video clips were presented centrally on a 24-inch LCD monitor (LG 24MB65PY-B, physical display size of 516.9 × 323.1 mm, resolution of 1920 × 1200 pixels). Viewing distance amounted to approximately 50 cm, resulting in a visual angle for the videos of 38.03° horizontally × 21.94° vertically. Eye movement data were recorded from the right eye using an EyeLink 1000plus system (SR Research, Ontario, Canada) with a sampling rate of 250 Hz. Head location was fixed using a chin rest and a forehead bar. Autonomic responses were continuously recorded at a sampling rate of 500 Hz during stimulation using a Biopac MP150 device (Biopac Systems, Inc.). Skin conductance was measured at the thenar and hypothenar eminences of the participant's non-dominant hand by a constant voltage system (0.5 V) using a bipolar recording with two Hellige Ag/AgCl electrodes (1 cm diameter) filled with 0.05 M NaCl electrolyte. An electrocardiogram (ECG) was recorded using mediware Ag/AgCl electrodes (servoprax, Wesel, Germany) attached to the manubrium sterni and the left lower rib cage. The reference electrode was placed at the right lower rib cage. Stimulus presentation and data collection were controlled using the Psychophysics Toolbox (Brainard & Vision, 1997) on MATLAB R2011b (MathWorks, Natick, MA, USA), and the EyeLink Toolbox (Cornelissen, Peters, & Palmer, 2002).

Procedure

Participants were invited to the laboratory individually and informed about the purpose of the study. Upon completing an informed consent form and a sociodemographic questionnaire, they were connected to the measurement instruments and given a detailed verbal explanation of the experiment. The 90 video clips were randomly sorted for each participant and presented in three blocks containing 30 clips each. Participants were asked to hold their heads still during the blocks, but allowed to sit comfortably or stand up between the blocks. The eye tracking system was calibrated and validated before each block using a 9-point calibration grid. Furthermore, a central fixation cross was presented for a randomly selected time interval between 5 and 9s before each video clip, and participants were asked to fixate it. The participants were given the instruction to watch and freely explore the video clips similar to watching a television program. Heart rate and skin conductance were recorded only during this part of the experiment. Subsequently, participants watched the clips for a second time in the same order as before and rated them for arousal and valence using the Self-Assessment Manikin (Bradley & Lang, 1994) on a scale from 1 to 9. For the participants, about 45 minutes passed between watching a video a first and a second time. The Self-Assessment Manikin, which is routinely used in psychological research on emotional processing, involves a numerical scale which is accompanied by simplified drawings of a person in order to illustrate the concepts of valence and arousal with facial expressions and other comic-style visualization techniques. Additionally, we constructed a 9-point personal relevance scale by adopting non-verbal, graphic representations similar to those used for arousal and valence. Participants were asked to state, in a broad sense, to what degree each depicted scene had a personal relevance to them. To illustrate the abstract idea of relevance, the manikins were color coded using various shades of grey (darker colors = higher relevance). Finally, participants filled out several psychometric tests and questionnaires which are not part of this study.

Data Processing

Image processing was performed using MATLAB R2011b (The MathWorks, Natick, MA,USA). We computed low-level saliency maps for each video frame using the GBVS algorithm (Harel et al., 2007). The channels "DKL Color", "Intensity", "Orientation" and "Flicker" were integrated into the maps with equal weights. In order to reduce the impact of strong changes in the low-level saliency distribution between successive video frames, we applied Gaussian blurring along the temporal dimension of the video data with a standard deviation of 2 frames. This technique aimed at better harmonizing the temporal reactivity of low-level saliency distributions with that of the human visual system, which cannot perform an entire action-perception cycle within the duration of one video frame. Each low-level saliency map was then normalized by dividing values for each pixel by the mean of the image, ensuring an average low-level saliency of 1 while preserving differences in low-level saliency variation between video frames.

Gaze raw data were analyzed using R for statistical computing (version 3.2; R Development Core Team, 2015). Gaze data during the first 150 ms after stimulus onset were excluded from the analysis to account for a minimum reaction time to leave the central fixation cross presented immediately before (Rösler et al., 2017). Data of each trial were corrected to account for drifts in head position. This was done using the mean valid gaze positions of the last 300 ms before stimulus onset as baseline. A recursive outlier removal algorithm was adopted to avoid correcting for drifts based on faulty gaze data (e.g., when participants did not fixate the fixation cross at some point during the last 300 ms of its appearance): Separately for x and y baseline coordinates, the lowest and highest values were both removed from the distribution, individually compared to the distribution of the remaining data and entered again if they were located within 3 standard deviations from the mean. This process was recursively applied to the remaining data until both the highest and the lowest data point met the criterion to be re-entered to be distribution. Subsequently, baseline position data from trials containing blinks or a discarded x or y component (M = 8.19% of all trials per participant, SD = 8.74%) were replaced by the mean of all valid trials, and baselines were subtracted from gaze data in each trial.

Since we preselected the videos with respect to their emotional valence, we primarily analyzed the influence of valence on attentional exploration and used arousal and relevance ratings to ensure comparability of video sets. Subjective valence ratings were expressed by the participants on a scale from 1 (very negative) to 9 (very positive). Intraclass correlation coefficients revealed varying interindividual consistency for valence (ICC = 0.65, 95% CI = [0.58, 0.72]), arousal (ICC = 0.47, 95% CI = [0.40, 0.56]) and relevance ratings (ICC = 0.19, 95% CI = [0.15, 0.26]). As a rule of thumb, coefficients between 0.60 and 0.75 are considered good, results between 0.40 and 0.59 are considered fair and results below 0.40 are considered low regarding interindividual consistency (Cicchetti, 1994). On the one hand, we directly used these ratings as a predictor in the GLMMs, on the other hand, we reclassified the videos into positive, neutral, and negative clips for additional analyses and manipulation checks. The thresholds between these three categories were adjusted individually for each participant to align the frequency with which each valence category was selected. For instance, if a participant tended to disregard the extremes of the rating scheme while showing a positivity bias, her ratings 6 and 7 may be relabeled as neutral (instead of 4 to 6 as one would define a priori). Specifically, an algorithm compared all possible permutations of the two thresholds and selected the combination that exhibited the smallest total difference in category size. As a result, M = 27.59 (SD = 7.12) videos were classified as negative, M = 32.09 (SD = 5.66) as neutral and M = 30.31 (SD = 5.16) as positive.

Autonomic responses were analyzed using the R software package as well. Skin conductance (or electrodermal activity, EDA) at trial start was subtracted from all data points within each trial, and data for each trial were averaged for further analyses. Heart rate (HR) data were calculated from the ECG recordings. First, R-waves were detected using a semi-automatic procedure. R-R-intervals were then converted to HR (in beats per minute) and a second-by-second sampling was applied (Velden & Wölk, 1987). The last second prior to stimulus onset served as prestimulus baseline and the corresponding HR value was subtracted from all values during stimulation (i.e., 20s). As for EDA, data were then averaged across each trial.

Personality Assessment

In an additional and exploratory analysis, we characterized participants along several self-report measures for personality traits (NEO-FFI; Costa & McCrae, 1989; Körner, Geyer, & Brähler, 2002), general trait anxiety (State-Trait Anxiety Inventory; STAI; Spielberger, 1983), anxiety in social interactions (Social Interaction Anxiety Scale; SIAS; Mattick & Clarke, 1998) and autistic traits (short version of the Autism-Spectrum Quotient; AQ-K; Baron-Cohen et al., 2006; Freitag et al., 2007), and investigated the cohesion between these measures and gaze behavior. All of these analyses are described in Appendix A.

Results

Arousal, Relevance and Autonomic Responses as Function of Presence of Persons and Valence.

In order to confirm the expected modulation of autonomic responses by differences in perceived valence as well as the presence of persons, we first examined the influence of valence and presence of persons on arousal ratings and autonomic measures using 2 $\times$ 3 repeated measures ANOVAs with video category (social vs. non-social) and emotional valence ratings (individually reclassified as positive, neutral, and negative) as within-subject factors. In all statistical analyses, α was set to 0.05. For ANOVAs and regression models, η_p^2 and R^2 are reported as effect size estimates, respectively. For all ANOVAs, degrees of freedom were adjusted using the Greenhouse-Geisser correction to account for possible violations in sphericity, and corresponding ϵ values are reported. Post-hoc pairwise comparisons were performed using Tukey's HSD test.

Arousal ratings Figure 2.1 were affected by valence ($F(2,\ 62) = 62.66$, $p < 0.001$, $\eta_p^2 = 0.67$, $\epsilon = 0.680$) and presence of persons (F(1, 31) = 22.80, $p < 0.001$. $\eta_p^2 = 0.42$) but not by a valence $\times$ presence of persons interaction ($F(2,\ 62) = 2.76$, $p = 0.085$, $\eta_p^2 = 0.08$, $\epsilon = 0.788$). Specifically, arousal ratings were higher for social compared to non-social videos, and higher for negative compared to neutral ($p < 0.001$) and neutral compared to positive videos ($p = 0.002$).

Relevance ratings 2.1 were affected by valence ($F(2,\ 62) = 44.73$, $p < 0.001$, $\eta_p^2 = 0.59$, $\epsilon = 0.870$) and by a valence $\times$ presence of persons interaction ($F(2,\ 62) = 9.68$, $p < 0.001$, $\eta_p^2 = 0.24$, $\epsilon = 0.929$), but not by presence of persons alone ($F(1,\ 31) = 2.55$, $p = 0.120$, $\eta_p^2 = 0.08$). Specifically, relevance ratings were higher for positive compared to negative ($p < 0.001$) and for negative compared to neutral videos ($p < 0.001$), resulting in a skewed U-shaped relation between valence and relevance. Relevance ratings for positive videos were furthermore higher than for negative videos ($p < 0.001$). For videos rated as positive, non-social videos were rated as more relevant ($p = 0.001$), whereas for videos rated as negative, social videos were rated as marginally more relevant ($p = 0.071$). There was no difference in relevance ratings between social and non-social videos rated as neutral ($p = 0.376$).

Heart rate and skin conductance were measured as manipulation checks for differences in perceived valence 2.2. We found a larger heart rate deceleration in social compared to non-social scenes ($F(1,\ 31) = 7.47$,$p = 0.010$, $\eta_p^2 = 0.19$) and an effect of video valence on heart rate changes ($F(2,\ 62) = 4.12$, $p = 0.029$, $\eta_p^2 = 0.12$, $\epsilon = 0.826$), but no interaction of the two factors ($F(2,\ 62) = 0.66$, $p = 0.521$, $\eta_p^2 = 0.02$, $\epsilon = 0.929$). Specifically, negative videos resulted in a stronger heart rate deceleration compared to positive videos ($p = 0.020$), while there was

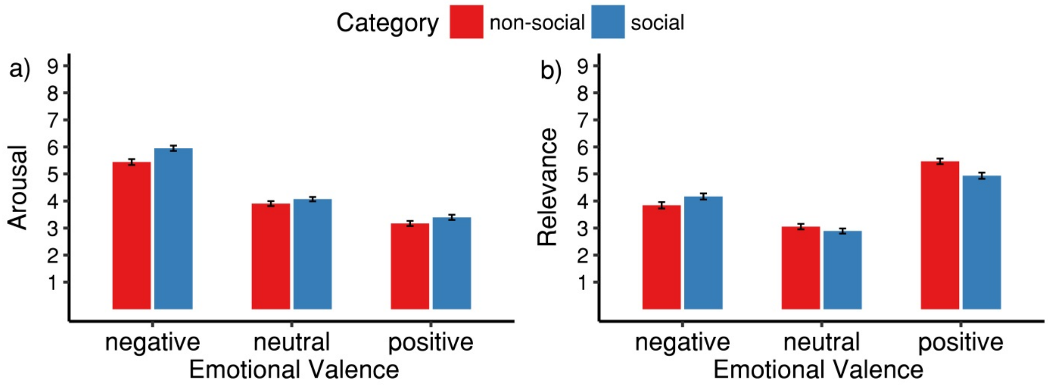

Figure 2.1: Effects of valence and presence of persons in videos on arousal and relevance ratings. Error bars indicate SEM.

no statistically significant difference between negative and neutral ($p = 0.109$) or neutral and positive ($p = 0.756$) videos. Skin conductance was affected by valence ($F(2,\ 62) = 3.81$, $p = 0.027$, $\eta_p^2 = 0.11$, $\epsilon = 0.916$), but not by presence of persons ($F(1,\ 31) = 1.18$, $p = 0.286$, $\eta_p^2 = 0.04$) or an interaction ($F(2,\ 62) = 0.26$, $p = 0.774$, $\eta_p^2 = 0.01$, $\epsilon = 0.979$). Specifically, skin conductance was lower for negative than for positive videos ($p = 0.021$), but there was no statistically significant difference between negative and neutral ($p = 0.312$) or neutral and positive ($p = 0.407$) videos.

Low-level Saliency of looked-at Pixels in different Video Categories

Next, we investigated the effect of valence and social content on the tendency to look at visually salient regions. To this end, we compared mean low-level saliency of all looked-at pixels by means of a 2×3 repeated measures ANOVA using video category (social vs. non-social) and emotional valence (positive, neutral, negative) as within-subject factors. Low-level saliency of looked-at pixels 2.3 was affected both by presence of persons ($F(1,\ 31) = 93.29$, $p < 0.001$, $\eta_p^2 = 0.75$) and valence ($F(2,\ 62) = 7.01$, $p = 0.002$, $\eta_p^2 = 0.18$, $\epsilon = 0.993$). The interaction of both factors did not reach statistical significance ($F(2,\ 62) = 1.61$, $p = 0.208$, $\eta_p^2 = 0.05$, $\epsilon = 0.863$). Specifically, low-level saliency of looked-at pixels was lower in videos with social information compared to videos without social information. Low-level

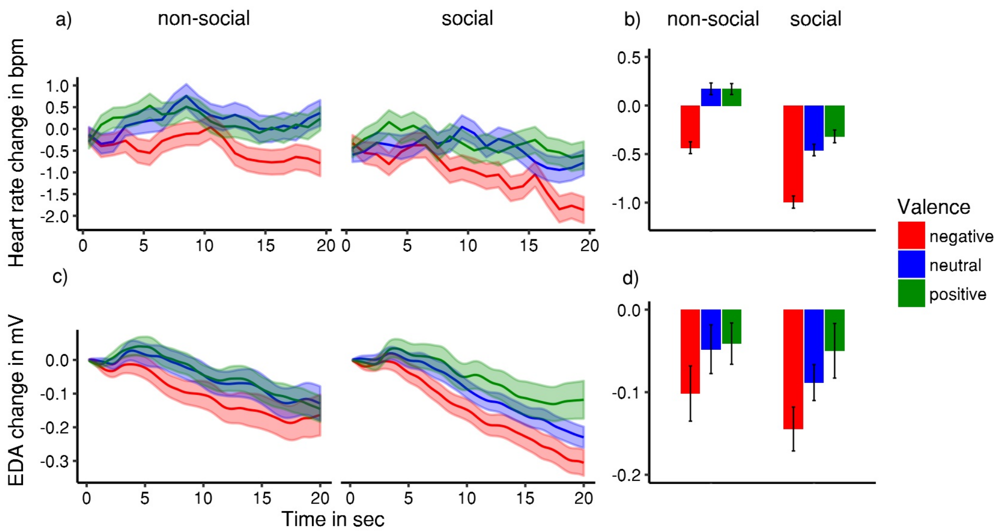

Figure 2.2: Physiological responses to different video categories. (a) Baseline-corrected heart rate change over time for non-social vs. social videos as a function of valence. (b) Heart rate change data aggregated across each trial. (c) Baseline-corrected change in electrodermal activity over time for social vs. non-social videos as a function of valence. (d) Electrodermal activity change aggregated across each trial. Ribbons and error bars indicate SEM.

saliency of looked-at pixels was also lower for negative than for neutral ($p = 0.005$) and positive ($p = 0.006$) videos, while there was no such difference between neutral and positive videos ($p = 0.997$). In all conditions, low-level saliency of looked-at pixels was higher than 1 - the value expected for a viewing behavior not guided by low-level saliency.

Directly Predicting Gaze using GLMMs

While the analysis described above suggests a reduced influence of low-level saliency on visual exploration in the presence of social features, it cannot describe the relative contribution of both factors directly. Furthermore, it is susceptible to correlations between low-level saliency, social information and other potential predictors such as centrality. We therefore set up various generalized linear mixed models (GLMM) to directly describe the influence of centrality, low-level saliency, social information and valence on gaze behavior in the social videos.

This approach was adapted from Nuthmann and Einhäuser (Nuthmann &

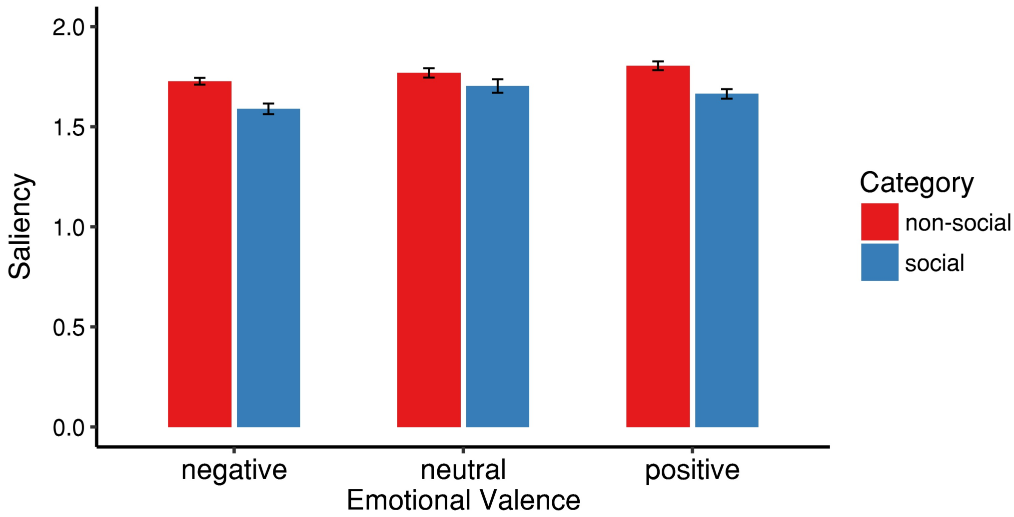

Figure 2.3: Mean low-level saliency of looked-at pixels in videos with and without presence of persons and for all three emotional valence subgroups. Error bars indicate SEM.

Einhäuser, 2015). Social information in the videos was defined in a conservative manner, marking the human heads in each video frame with circular regions of interest (ROI). Analogous to the low-level saliency maps, these ROI maps consisted of ones representing pixels on heads and zeros representing pixels elsewhere. Centrality for each pixel was defined as inverse Euclidean distance to the center of the video. Predictor maps for each video frame were then divided into a 32×18 grid and data were collapsed within each of these 40×40 pixels grid cells. The size of the grid cells ($2.5° \times 2.5°$ of visual angle) approximated the functional field of the human fovea centralis. Values for low-level saliency, social ROI and centrality were then z-standardized to make resulting beta coefficients comparable.

In the GLMM, we implemented centrality, low-level saliency, social ROI as well as valence as fixed effects, low-level saliency $\times$ ROI as well as a low-level saliency $\times$ valence as interaction terms and participant ID and video ID as random effects. The response variable was binary and stated whether a given grid cell was fixated in a particular video frame or not. It was made accessible to linear modelling using the probit link function. In order to ascribe the same importance to both of the dependent variable's states (looked-at vs. not looked-at) during modeling, we did not include all of the 575 grid cells per video frame which were not looked at. Instead, only one grid cell which was not looked at was randomly selected in addition to the looked-at grid cell. The resulting observation matrix consisted of approximately 1.73 million entries (32 participants × 45 videos × 600 frames × 2 grid cells). To compute regression weights for each predictor, we used the glmer function of the lme4 package (Bates, Maechler, Bolker, Walker, & others, 2014) and the bobyqa optimizer. Since estimating goodness of fit is intricate for linear mixed models, we computed an analogue to the coefficient of determination, R^2, but maintained this naming convention. This was accomplished by calculating the square of the correlation between observed data and data predicted by the model (Byrnes & Stachowicz, 2009; Cameron & Windmeijer, 1996).

We adopted an incremental model building strategy in order to assess the lower bound of each predictor's contribution to explained variance. In the first model, we only included centrality as a fixed term, as this simple cue provides the most frugal gaze prediction. Second, we further included low-level saliency, a more complex but still bottom-up information channel. In a third model, the predictor social ROI, defined as depicted people's heads, was included. In a fourth and fifth model, we further included a low-level saliency × ROI interaction term and a low-level saliency × valence interaction term.

Adopting an incremental model building strategy resulted in five models comprising an increasing number of predictors, with each model being nested in the consecutive one. This procedure allowed to estimate the lowest bound of a predictor's contribution to explained variance, even with correlations among the predictors. Since each model implemented random selections of only one out of 575 not looked-at-cells per video frame, we repeated the entire process 100 times and report averages and 95% confidence intervals of models' characteristics in 2.1. Explained variance profited from the inclusion of low-level saliency and social ROI, which can be seen in the rising and non-overlapping 9% confidence intervals of explained variance in these models. Explained variance did not profit from further adding a low-level saliency × ROI interaction and a low-level saliency × valence interaction. The gain in explained variance when including social ROI in addition to centrality and low-level saliency was 6.62%, marking the most conservative amount of explained variance that can be attributed to social ROI alone. The three predictors centrality, low-level saliency and social ROI collectively explained 31.82% of the variance in gaze data.

Table 2.1: Results of hierarchical generalized linear mixed models (GLMMs) examining the contribution of different predictors for fixation selection. Standardized regression weights and explained variance (R^2) for models comprising an increasing number of predictors. Models are nested and include predictors in models shown above. 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 2.5^{th} and 97.5^{th} percentile rank as an unbiased estimate of the 95% confidence interval.

		Centrality	Saliency	ROI	Saliency × ROI	Saliency × Valence	R^2
1	Centrality	0.554 [0.553, 0.555]					0.154 [0.153, 0.155]
2	+Saliency	0.266 [0.265, 0.268]	0.574 [0.572, 0.577]				0.252 [0.252, 0.253]
3	+ROI	0.288 [0.286, 0.289]	0.544 [0.542, 0.547]	0.506 [0.502, 0.509]			0.318 [0.317, 0.319]
4	+Saliency × ROI	0.287 [0.285, 0.289]	0.526 [0.524, 0.529]	0.509 [0.505, 0.512]	-0.103 [-0.107, -0.100]		0.318 [0.318, 0.319]
5	+Saliency × Valence	0.288 [0.286, 0.290]	0.528 [0.525, 0.530]	0.510 [0.507, 0.514]	-0.104 [-0.108, -0.100]	0.002 [-0.001, 0.004]	0.320 [0.319, 0.321]

Internal Consistency of Predictors

In this study, participants viewed a variety of video clips which varied on a number of dimensions. This stimulus set was thus very different from the well-standardized sets of stimuli that were used in many studies in the field, but aims at mapping the diversity and richness of every-day experiences. One may therefore object that models based on viewing behavior might not reflect general patterns in attentional allocation, but rather reflect idiosyncrasies of the individual video clips used. To our opinion, this concern can be refuted by demonstrating an intraindividual stability of viewing behavior across the different video clips. We therefore assessed the consistency of interindividual differences in viewing patterns across this diverse set of video stimuli. To this end, we computed generalized linear models as described above, but individually for each social video for each participant, each time describing the influence of centrality, low-level saliency and social information on gaze allocation (32 participants $\times$ 45 videos = 1440 GLMMs). The entire procedure was again repeated 100 times to account for influences of the random selection of not looked-at cells.

On average among the 100 bootstrapping draws, 87.1 out of 1440 models (6.05%, range: 68-100, 4.72-6.94%) could not converge. Beta weights in these models were replaced using a multiple imputations technique, Predictive Mean Matching (Van Buuren, 2018) (PMM). We created five imputed datasets for each iteration, resulting in a total number 500 datasets of predictor weights. Predictor weights were z-standardized along the video dimension to exclude effects due to general differences in the videos (e.g., flashing lighting, sudden appearance of fast objects or persons), but maintain the order and distances between predictor strengths for each participant. Resulting values were then tested for consistency across the whole set of videos using Cronbach's α. Cronbach's α is commonly used to quantify, on a scale from 0 to 1, the extent to which different items (e.g., from a questionnaire) are intraindividually consistent with each other, or, figuratively speaking, point into the same direction (Cortina, 1993). Internal consistency was

$\alpha = 0.88$ (95% CI = [0.87, 0.89]) for the predictor central bias, $\alpha = 0.75$ (95% CI = [0.70, 0.79]) for the predictor low-level saliency and $\alpha = 0.87$ (95% CI = [0.85, 0.89]) for social ROIs. These values indicate high intraindividual stability in the attentional preferences across the stimulus set. Interestingly, internal consistencies above 0.90 have been argued to indicate redundancy rather than consistency for personality questionnaires (Streiner, 2003). The currently observed values for a rich and ecologically valid set of videos can hardly be called redundant with regards to the video content and suggest high stability of attentional exploration patterns.

Discussion

In the present study, we assessed how social information and affective quality of naturalistic video scenes affect gaze allocation in addition to low-level image features such as physical saliency and centrality. Low-level saliency and social information both had substantial and similarly large effects on gaze behavior. Additionally, participants exhibited consistent differences in their viewing behavior in terms of the predictive value of centrality, low-level saliency and social ROIs in the rich set of video stimuli used. This demonstrates that attentional mechanisms driven by centrality, low-level saliency and social information exert a similar influence across a wide range of situations and do not depend on subtle changes in the experimental setup. To our opinion, this finding provides backup for the assumption that comparisons of viewing behavior along different video categories (social vs. non-social, positive vs. neutral vs. negative) are informative and valid, even when standardization was reduced in the current study in favor of external validity.

Valence variation between videos, although arguably more subtle compared to standard image databases like the IAPS (Lang et al., 1997), could be affirmed by a heart rate deceleration for negative and for social videos. These findings are in line with other studies that report heart rate deceleration in persons viewing negative compared to positive or neutral images, and a stronger heart rate deceleration for images containing human attacks compared to images containing animal attacks (Bradley, Codispoti, Cuthbert, & Lang, 2001). Variation in video valence and arousal was, however, not underpinned by lower skin conductance levels during viewing neutral as compared to negative or positive videos. It must be noted that, although autonomic measures are an established tool to quantify emotional reactions, findings differ on subgroups of emotions (Kreibig, 2010). For instance, one study (Bernat, Patrick, Benning, & Tellegen, 2006) found an enhanced skin conductance response to threatening pictures, but not to pictures that were negative in other respects. We cannot rule out the possibility that the videos used in the present study elicited specific subgroups of emotions that we did not inquire in the questionnaires. Moreover, most studies on autonomic responses to affective stimulation used pictorial material (Bradley et al., 2001; Lang, Greenwald, Bradley, & Hamm, 1993) and it is therefore unclear to what degree these findings translate to dynamic scenes such as the video clips used here. Finally, although arousal ratings were generally higher for negative video clips, no such increased arousal was evident for positive videos and overall arousal ratings were rather moderate. Interestingly, a U-shaped distribution was found for relevance ratings with emotionally charged video clips receiving higher ratings than neutral stimuli. Since participants viewed each video twice, modulations of perceived valence due to a mere exposure effect cannot be ruled out, although we expect such an effect, if present, to be subtle and not specific to individual videos (Zajonc, 2001).

One line of gaze data analysis showed that participants looked at less salient areas in social as compared to non-social scenes, and at less salient areas in negative compared to positive and neutral scenes. This finding is in line with the concept of a default attention system that directs gaze towards visually salient objects, but is partly overridden by top-down processes such as the search for social or aversive information (Cerf et al., 2008; Nyström & Holmqvist, 2008). This pattern is comparable to arousal ratings where we also observed higher ratings for social than for non-social video clips but does not directly correspond to relevance ratings that showed an interaction between emotional valence and the presence of persons. However, the class of analysis used here does not allow for directly assessing the relationship between social information and gaze behavior, and is susceptible to correlations between social information and other information channels. Moreover, for the videos used in this study, arousal levels were not only higher for social videos, but also for negative scenes in general, thus potentially distorting comparisons between the different videos categories.

We therefore computed several generalized linear mixed models encompassing various cues to predict gaze behavior in social scenes. The best-performing model included centrality, low-level saliency and social information as predictors. Crucially, even though social information was defined conservatively as comprising only human heads, it yielded a regression weight nearly as large as low-level saliency and explained at least 6.62% of variance in gaze data in addition to centrality and low-level saliency. The negative low-level saliency × social information interaction may be interpreted as a ceiling effect in attentional allocation: when a scene area is both visually salient and exhibits social information, the resulting interest in this area is large, but smaller than would be expected if both attentional mechanisms were merely added, as assumed in a GLMM. However, it must be noted that the gain in explained variance due to a low-level saliency × social information interaction was not significant. A low-level saliency × affective quality interaction did not contribute to explained variance in this analysis. This finding may seem surprising considering that mean low-level saliency of looked-at pixels was lower in negative compared to neutral or positive videos. However, in the GLMM, variance can be allocated to the factors centrality as well as directly to the social regions of interest, possibly suppressing variance allocation for certain interactions found in other analyses. This finding highlights the complementary nature of the two gaze analyses we performed - comparing low-level saliency of looked-at pixels and directly predicting gaze location.

In the present study, low-level saliency was defined as a summation of feature maps in the GBVS algorithm (Harel et al., 2007) with equal weights for each channel. Future studies may test the robustness of our approach by comparing models using several of the abundance of low-level saliency models that have been proposed (Kümmerer et al., 2015). Likewise, although summation of feature channels in low-level saliency algorithms is still widespread (Cerf et al., 2009; Kümmerer et al., 2015), future research should test whether optimizing feature weights using one of several proposed approaches (Borji, 2012; Coutrot & Guyader, 2017; Zhao & Koch, 2011) can even increase the amount of variance explained by low-level saliency. However, since simple summation of feature weights has been shown to outperform weight optimization techniques in several domains where large amounts of uncontrolled variance are at play (DeMiguel, Garlappi, & Uppal, 2007; Gigerenzer & Brighton, 2009), simple feature weight summation appears to be a reasonable default strategy.

Operationalizing social information as only the heads of depicted persons may be seen not only as a conservative but even as an impoverished definition. For instance, two studies (Birmingham et al., 2008; End & Gamer, 2017a) found not only heads to be fixated more often than other objects, but also - though less so - human bodies. A similar attentional bias was found for objects which are gazed at by depicted persons (Castelhano, Wieth, & Henderson, 2007; Parks et al., 2015; Xu et al., 2014). A comprehensive definition of social information would therefore need to include these and perhaps even more features. As incorporating more predictors into the model would increase the amount of variance explained, this further highlights that the importance of social information for fixation selection is still underestimated in the present study.

The material used in this study was informed by a cognitive ethology approach (Birmingham et al., 2009a). We avoided artificially impoverished stimuli such as images of faces shown in isolation, and instead presented participants a large variety of complex, dynamic and contextually rich video clips. By these means, we intended to elicit more natural andrepresentative viewing behavior in our participants. The use of generalized linear mixed models allowed to guard the effect of social attention against possible confounds, thus serving as a counterpart for an experimental setup in which variables are not held constant between experimental conditions.

However, it should also be noted that the testing environment itself still significantly deviates from field conditions, since participants were asked to continuously attend to a video screen placed in front of them and were unable to interact with the persons and situations presented to them. Some authors (Tatler, Hayhoe, Land, & Ballard, 2011) argue that in real-world situations, fixation selection is often guided by an expectation to interact with an object. Furthermore, it was asserted that gaze behavior in real social situations is often guided by the knowledge that conspecifics may detect, and possibly reciprocate, one's gaze (Foulsham et al., 2011). Since these possibilities are disrupted in passive-viewing tasks with photos or videos presented on a computer screen, viewing behavior might systematically deviate from that found in everyday situations. One technical solution which has been argued to simultaneously excel at both ecological validity and experimental control is virtual reality (Parsons, 2015). With realistic forms of interaction implemented, this technology promises to close a gap between complex field studies and well-controlled laboratory experiments.

While this study demonstrated the relevance of social information for attracting gaze allocation, an open question is to what extent this form of attention must be seen as deliberate or automatic. Over the course of a 20-second-video, fixation selection can evidently not be entirely automatic. However, there are hints that saccades towards social stimuli may be reflexive in a time period right after the appearance of such stimuli. Several studies (Fletcher-Watson et al., 2008; Rösler et al., 2017; Scheller et al., 2012) found saccades toward socially relevant regions so shortly after stimulus onset that they are not well explained by cortical routes of top-down information processing (Knudsen, 2007). Instead, it has been proposed that faces or eyes are also processed in subcortical circuits involving the amygdala (Benuzzi et al., 2007) and might drive reflexive attentional capture via this route (Gamer & Büchel, 2009; Gamer et al., 2013). An interesting question arising in this context is whether naturalistic video clips contain identifiable key moments that elicit reflexive saccades towards social features.

A promising application of our paradigm may be the investigation of attentional mechanisms in patient groups. One clinical condition that typically entails altered face processing is social anxiety disorder. Patients with this disorder show an initial hypervigilance for social threat cues (Boll et al., 2016; Mogg et al., 2004; Seefeldt et al., 2014), but avoid looking at the eyes region when presented with images for an extended period of time (Moukheiber et al., 2010; Weeks et al., 2013). Patients with autism spectrum disorder were found to orient their gaze more towards salient areas and less towards faces, objects indicated by other persons' gaze (Wang et al., 2015) and eyes (Spezio, Adolphs, Hurley, & Piven, 2006) when viewing naturalistic images, as well as less towards faces and more towards letters when viewing dynamic scenes (Nakano et al., 2010), although findings were not entirely consistent (Rutherford & Krysko, 2008). With healthy observers, our GLMM-analysis of naturalistic videos yielded robust results while posing little cognitive demands to the participants. Together with the simplicity and naturalness of the task, this approach may be informative as well as feasible in a variety of patient groups for whom alterations in social attention are debated. Crucially, analyses based on GLMMs allow for detailed comparisons of model weights between individuals, stimulus material and their interactions (Nuthmann & Einhäuser, 2015; Nuthmann, Einhäuser, & Schütz, 2017).

The present study gauged the importance of social information on gaze behavior when viewing naturalistic, contextually rich dynamic scenes, while at the same time controlling for the low-level information channels centrality and low-level saliency. With a conservative definition of social information, we found its influence on viewing behavior similarly large as low-level saliency. We furthermore argue that our research paradigm shows promise for investigations of social attention under a variety of circumstances, such as in clinical populations.

Data availability The datasets generated during and/or analyzed during the current study are available at https://osf.io/943qb/.

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

Author Contributions M.R. and M.G. designed the experiment. M.R. collected and analysed the data. M.G. supervised data analysis. M.R. and M.G. wrote and reviewed the manuscript.

CHAPTER 3

Social Attention in Real-Life

Social Anxiety modulates Visual Exploration in Real-Life — but not in the Laboratory

Marius Rubo, Lynn Huestegge & Matthias Gamer

Julius-Maximilians-Universität Würzburg

Corresponding Author: Marius Rubo

Email adress: marius.rubo@uni-wuerzburg.de

Published 2019 in *British Journal of Psychology.*

Abstract

In clinical reports, individuals high on social anxiety are often described to avoid gaze at other people whereas several experimental studies employing images of persons yielded conflicting results. Here we show that gaze avoidance crucially depends on the possibility of social interactions. We examined gaze behavior in individuals with varying degrees of social anxiety in real-life and in a second group of participants using a closely matched laboratory condition. In the real-life situation, individuals with a higher degree of social anxiety had a reduced bias to look at near persons compared to individuals with a lower degree of social anxiety, while gaze behavior in the laboratory group was not modulated by social anxiety. This effect was specific to social attention since there was no corresponding effect regarding fixations on objects. The presence of anxiety effects in real-life but not in the laboratory condition, where participants do not expect to be evaluated by gazed-at conspecifics, points to critical deficits of current laboratory research paradigms in eliciting authentic social attentional mechanisms, possibly leading to spurious results.

Introduction

A large proportion of people in the community exhibit signs of social anxiety such as fear of speaking to an audience or talking to persons in authority (Stein et al., 1994), and a subgroup of these individuals suffers so severely from their fear of other persons' scrutiny that they may receive a diagnosis of social anxiety disorder (SAD), or social phobia (American Psychiatric Association, 2013). Cognitive models of SAD (Rapee & Heimberg, 1997) propose biases in evaluating social information that provoke anxiety in social situations and contribute to the etiology and maintenance of the disorder. For instance, individuals high on social anxiety show a tendency to interpret ambiguous social information negatively (Huppert et al., 2007) and preferentially memorize negative social cues (Lundh & Öst, 1996).

Clinical reports of individuals with SAD furthermore emphasize an avoidance of other people's gaze (Schneier et al., 2011). This specific effect, however, could not be robustly replicated in laboratory settings, with some studies reporting reduced (Moukheiber et al., 2010; Weeks et al., 2013), but other studies reporting heightened amounts of fixation on faces or eyes in social phobic adults and shy children (Boll et al., 2016; Brunet et al., 2009; Wieser et al., 2009).

Note that such laboratory-based studies on attentional biases in social anxiety typically rely on passively viewing images of persons and do not incorporate any interaction with real people. This approach is problematic since gaze behavior is known to be modulated by the presence of other people, even in the general population. For instance, Laidlaw et al. (2011) found participants to frequently look at the videotape of a confederate, but to avoid looking at a live confederate when she was present in the room. Along these lines, Gobel et al. (2015) found participants to avoid looking at the eyes of higher ranked individuals in a videotape when they believed the depicted person would in return later see a videotape of them. These effects of the (real or imagined) presence of another person on one's gaze behavior is typically explained by a dual function of gaze, that is by the fact that orienting one's gaze serves both to redirect overt attention and to signal one's intentions to others (Gobel et al., 2015; Risko et al., 2016). Additionally, well-established effects like the facilitation or impediment of certain tasks in social situations (Guerin, 2010) highlight the important distinction between viewing an image of another person and locating oneself in his or her presence. We therefore believe that the distinction between real-life and laboratory situations, which is beginning to be acknowledged in the literature on social attention in the general population, might explain conflicting findings in social anxiety research and should be considered. The present study is therefore informed by a cognitive ethology approach (Kingstone, Smilek, & Eastwood, 2008), where behavior is first investigated in a situation in which it naturally occurs, and only then transferred to a laboratory situation.

In the current study, a first group of participants varying in pre-screened social anxiety traits walked a specified path in a public train station while their gaze was tracked. A second group matched for gender and social anxiety viewed video recordings of these itineraries in the laboratory. In both conditions, we quantified the time in which gaze was directed towards another person and object, or the path and furthermore registered when gaze was directed towards the vicinity or the distance (Foulsham et al., 2011). This approach allows us to investigate how the presence of other people influences social gaze in participants, and how participants' social anxiety modulates this effect. The real-life situation in the present study poses few restrictions on participants' behavior and employs no scripted situations (e.g. carried out by a confederate of the experimenter), allowing for an ecologically more valid assessment of behavior. We attempted to control for the higher variance which is naturally introduced in such paradigms by implementing a side-by-side comparison with a tallied laboratory situation involving a matched participant group. This enables us to more clearly carve out the effect of the physical presence of other people on participants' gaze behavior. We expected social anxiety to influence gaze on other individuals in the real-life situation more strongly as compared to the laboratory group.

Materials and Methods

Participants

Among several hundred persons who completed an online pre-screening, we recruited an a priori defined number of 60 participants based on medium to high social anxiety scores (M = 26.58 years, SD = 6.82 years, 43 females, 49 students, see Supplemental Methods for further details). Participants were either assigned to a real-life group (30 participants) or a laboratory condition (30 participants) while ensuring a matching between groups regarding gender and social anxiety pre-screening scores (matching for gender was not possible in one pair). Groups did not differ significantly in age, depression, general or social anxiety see Table 3.1. Note that in the present study, each participant in the laboratory group viewed scenes recorded from the perspective of their partner in a real-life situation. Matching partners on gender and social anxiety served to reduce systematic effects of moving behavior on comparisons of gaze behavior between the two groups. For instance, if gender or social anxiety systematically influence walking speed, straightness of walking or any other relevant variable, such effects will be equally present in the real-life and the laboratory situation and will therefore not distort comparisons between groups. All participants had normal or corrected-to normal vision and did not wear glasses in everyday situations. The study conformed to the principles expressed in the Declaration of Helsinki and was approved by the local ethics committee.

Following data acquisition, one participant in the real-life group (together with the corresponding partner in the laboratory group) was excluded from the analysis due to substantial discrepancies between prescreening and detailed assessment of social anxiety (for details see Supplemental Methods), which made it impossible to estimate the quality of the matching between the partners. For the remaining 29 pairs, prescreening scores were highly correlated ($r(29) = .82$, $p < .001$) thus indicating successful matching. Three further participants, all in the real-life group, were excluded due to technical problems with the mobile eye-tracking device resulting in less than 75% valid data. Altogether, 26 participants (20 females) in the real-life group and 29 participants (21 females) in the laboratory group remained in the analysis. These participants were characterized regarding symptoms of depression (Beck Depression Inventory, BDI (Beck, 1961)), trait anxiety (trait part of the State Trait Anxiety Inventory, STAI (Spielberger, 1983)) and social anxiety (Social Interaction Anxiety Scale, SIAS, (Mattick & Clarke, 1998)), and the Social Phobia Anxiety Inventory, SPAI (Turner, Beidel, Dancu, & Stanley, 1989)) by means of self-report questionnaires that were completed after the experiment (see Supplemental Methods). The SPAI aims at representing the entire continuum of socially anxious concerns and serves as the most comprehensive account of social

Table 3.1: Sociodemographic and questionnaire data of participants in the real-life and the laboratory group. Prescreening: Social anxiety assessment. BDI: Beck's Depression Inventory; STAI-T: State-Trait Anxiety Inventory; SIAS: Social Interaction Anxiety Scale; SPAI: Social Phobia and Anxiety Inventory.

	Real-Life			Laboratory			
	Range	*M*	*SD*	Range	*M*	*SD*	Group comparison
Age	[19 46]	26.04	6.69	[18 48]	26.21	5.97	$t(53) = 0.10$, $p = .92$
Prescreening	[3.0 4.6]	3.53	0.38	[3.0 5.0]	3.63	0.59	$t(53) = 0.77$, $p = .45$
BDI	[2 35]	11.46	8.93	[1 40]	11.34	9.62	$t(53) = 0.05$, $p = .96$
STAI-T	[29 77]	50.54	13.43	[28 74]	46.69	12.15	$t(53) = 1.12$, $p = .27$
SIAS	[10 52]	29.92	11.48	[13 61]	35.69	13.99	$t(53) = 1.66$, $p = .10$
SPAI	[76 169]	119.82	27.37	[77 177]	133.25	30.98	$t(53) = 1.70$, $p = .10$

anxiety in the present study. Therefore, the SPAI was defined as the primary measure for characterizing social anxiety traits and all analyses rely on these scores. The SIAS specifically, and more briefly, focusses on fear of interacting with other people. Exploratory analyses on this measure are reported in the Supplemental Results.

Apparatus and Stimuli

For the real-life group, we recorded gaze of both eyes using SMI Eye Tracking Glasses 2.1 (SensoMotoric Instruments, October 2014) with iViewETG software at a sampling rate of 60 Hz. A video capturing the participants' field of view was recorded at 30 Hz with a resolution of 960 × 720 pixels. For the laboratory group, video clips were presented centrally on a 24-inch LCD monitor (LG 24MB65PY-B, resolution of 1920 × 1200 pixels). Viewing distance amounted to approximately 50 cm, resulting in a visual angle for the videos of 28.98° horizontally × 21.94° vertically. In the laboratory, eye movement data were recorded from the right eye using an EyeLink 1000 Plus system (SR Research, Ontario, Canada) at a sampling rate of 250 Hz. Head location was fixed using a chin rest and a forehead bar.

Procedure

Upon completion of an informed consent form and a brief sociodemographic questionnaire, participants in both groups were given information on the experiment, the routes they were going to walk or see as well as the erroneous information that the eye-tracking device would be installed in order to measure their pupillary response to varying lighting conditions. When later asked to comment on their experiences and thoughts concerning the experiment, none of the participants reported being aware that our primary research focus was on (social) gaze orienting and not on changes in pupillary responses as a function of lighting conditions.

In the real-life group, the eye-tracker was calibrated using a three-point calibration, then validated, and, if validation failed, calibrated again until validation yielded a positive result. All participants wore a hat to protect the eye-tracking measurements from direct sunlight. Participants then walked the predefined paths both in a populated train station (social condition) and in a close-by parking garage where other people were largely absent (non-social condition) in a randomized order. As this study's main interest is in social attention, only analyses concerning the social condition will be described here. The experimenter unobtrusively followed the participant at a distance of at least 10 m and intervened in four cases when participants were getting off the defined path. After a walk of approximately four to five minutes, participants reached the end of the route and waited standing for approximately 5 minutes until they were picked up by the experimenter. Calibration was validated, and, if necessary, performed again after the first condition.

In the laboratory group, participants were presented with the videos obtained from their matched participants in the real-life group, with both conditions presented in the same order as for the real-life participant. Before each condition, the eye tracking system was calibrated and validated using a 9-point calibration grid. Participants were given the instruction to watch the videos as if they were watching television. Stimulus presentation and data collection were controlled using the Psychophysics Toolbox (Brainard & Vision, 1997) on MATLAB R2011b (MathWorks, Natick, MA, USA).

Upon completing the experiment, participants in both groups filled out questionnaires to allow for participant characterization regarding symptoms of depression, trait anxiety and social anxiety (see Supplemental Methods).

Data Processing

We extracted videos displaying the participants' field of view over the duration of the experiment as well as the same videos with gazed-at locations indicated as colored rings using the SMI BeGaze (Version 3.5) software. Parallel to the real-life group, we produced videos displaying the mean gazed-at location in each frame as colored rings for the laboratory group using MATLAB R2011b (MathWorks, Natick, MA, USA). Matched participants in both groups were assigned the same video ID. Using the videos displaying the gazed-at locations, gaze was manually coded at 6 Hz (i.e., every fifth frame) using in-house software written in MATLAB by raters who were not aware of the experimental group assignment. Overall, 183,753 frames were coded. For each frame, raters noted whether or not a person was present in the frame, if a person, the path, an object or nothing particular was being gazed at (category) and if the frame could be considered as valid (see Supplemental Experimental Procedures for details on coding). We furthermore noted whether gaze was located within (vicinity) or beyond (distance) the near-distant action space - a space of approximately 8 meters around a person characterized by the effective use of eye accommodation, visual convergence, and retinal disparity in distance estimation (Daum & Hecht, 2009; Grüsser, 1983). Two raters each coded videos from 15 matched pairs (interrater reliability was checked for a subset of these videos, see Supplemental Experimental Procedures). For the analyses, we removed all frames in the dataset in which no persons were visible (M = 15.23%, SD = 10.61% in the laboratory group, M = 15.30%, SD = 11.30% in the real-life group), all invalid frames (M = 3.72%, SD = 4.47% in the laboratory group, M =

11.60%, $SD = 5.40\%$ in the real-life group) and all frames in which gaze location was coded as undefined ($M = 0.14\%$, $SD = 0.29\%$ in the laboratory group, $M = 0.13\%$, $SD = 0.20\%$ in the real-life group). We then produced a dataset for each participant, listing the relative frequencies of frames for which gaze was labelled to rest on each of the six coding categories (i.e., person, path or object in the vicinity or in the distance). The sum of gaze frequencies was normalized to add up to a value of 1 within each participant.

Statistical Analysis

All statistical analyses were performed using the R software for statistical computing (version 3.2; R Development Core Team, 2015). We first compared the fixation time on different categories in both conditions using a 2 $\times$ 2 $\times$ 3 mixed ANOVA with group (real-life vs. laboratory) as between-subject factor and distance (vicinity vs. distance) as well as category (person, object or path) as within-subject factors.

We furthermore investigated the stability of gaze preferences throughout the time course of the experiment. To this end, we separated gaze data from each participant into 20 bins of equal length ($M = 23.39$s, $SD = 4.63$s), computed relative gaze frequency for each category (person, object or path) and distance (vicinity vs. distance) and tested the stability of differences between participants across the individual bins using Cronbach's α. Cronbach's α varies between 0 to 1 and is commonly employed to quantify the intra-individual consistency of responses along a set of items (Cortina, 1993).

In order to investigate the influence of social anxiety on fixation on persons, we computed linear mixed models using the lme-function of the nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2009) with group (real-life vs. laboratory), distance (near vs. far) and the SPAI score included as fixed effects and the Video ID included as random effect. We chose to rely on the SPAI score in these analyses since this questionnaire allows for a comprehensive characterization of social anxiety across a variety of situations. Please note that the current pattern of results was similar when relying on the SIAS score instead. In order to better interpret a group × distance × SPAI interaction, we computed a corresponding linear mixed model for each of the two groups separately. To ensure the social specificity of attentional effects of social anxiety, we furthermore computed a corresponding linear mixed model across both groups, but with gaze on objects as dependent variable. In all statistical analyses, α was set to .05. For ANOVAs and regression models, η_p^2 and R^2 are reported as effect size estimates, respectively. For ANOVAs, degrees of freedom were adjusted using the Greenhouse-Geisser correction to account for possible violations of the sphericity assumption, and corresponding ϵ values are reported. Post-hoc pairwise comparisons were performed using Tukey's HSD test, and Cohen's d values are reported to display effect sizes in post-hoc comparisons. Parameters in linear mixed models were estimated using the Restricted Maximum Likelihood (REML) approach and tested for significance using F-tests.

Results

Distribution of Gaze in Real-Life and in the Laboratory

The three categories (person, object, path) were gazed at for a different proportion of time ($F(2{,}106) = 161.80$, $p < .001$, $\eta_p^2 = .75$, $\epsilon = .59$, see 3.1). Gaze was directed more frequently towards the vicinity than the distance ($F(1{,}53) = 443.27$, $p < .001$, $\eta_p^2 = .89$), and there was a significant distance × category interaction ($F(2{,}106) = 82.67$, $p < .001$, $\eta_p^2 = .61$, $\epsilon = .80$) as well as a significant group × distance interaction (F(1,53) = 15.24, $p < .001$, $\eta_p^2 = .22$). The group × category interaction ($F(2{,}106) = 2.06$, $p = .154$, $\eta_p^2 = .04$, $\epsilon = .59$), as well as the group × distance × category interaction did not reach statistical significance ($F(2{,}106) = 3.02$, $p = .065$, $\eta_p^2 = .05$, $\epsilon = .80$). Across both groups, both objects ($p < .001$) and persons ($p < .001$) were gazed at more often than the path, while there was no difference between objects and persons ($p = .889$). The preference for the vicinity was present in all three categories, but more pronounced for objects ($p < .001$, d

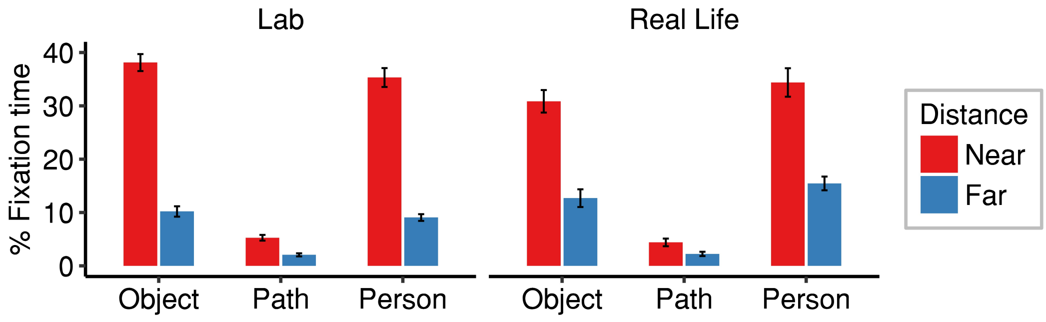

Figure 3.1: Percentage of time gaze was directed at objects, the path or persons, in the vicinity or distance, for both the laboratory and the real-life group. Error bars indicate SEM.

= 2.65) and persons ($p < .001$, $d = 2.48$) than for the path ($p = .043$, $d = 1.02$). The preference for the vicinity was furthermore present in both groups, but more emphasized in the laboratory ($p < .001$, $d = 1.55$) than in the real-life group ($p < .001$, $d = 0.98$).

Consistency of Gaze Behavior along the Experiment's Time Course

Table 3.2 displays internal consistency (represented as Cronbach's α) in gaze towards each category (person, object or path) and distance (vicinity vs. distance), separately for participants in the real-life group and laboratory group as well as for all participants combined. The interpretation of Cronbach's α was argued to depend on the research question, but one influential report recommended .70 as a goal for early stages of research and noted that values above .90 may be hinting more towards unnecessary redundancy in the measurement rather than consistency (Streiner, 2003). In the present study, consistency in gaze behavior was moderate or high for all categories and distances (values between .71 and .91). For all participants, α for gaze at near people was .88 and α for gaze at distant people was .86, highlighting high intra-individual stability in gaze towards other people throughout the experiment.

Table 3.2: Consistency (measured as Cronbach's α) of participants' viewing preferences along the experiment's time course

	Real-Life		Laboratory		Combined	
	near	far	near	far	near	far
Persons	.91	.84	.84	.71	.88	.86
Objects	.86	.91	.84	.89	.87	.91
Path	.84	.79	.84	.83	.84	.81

Effects of Social Anxiety on Gaze at Persons

When incorporating social anxiety (SPAI) scores into a linear mixed model predicting fixations on persons for the real-life and the laboratory group 3.2), we again found a main effect of distance (F(1,74) = 207.93, p < .001), a group × distance interaction (F(1,74) = 7.22, p = .009) and a group × distance × SPAI interaction (F(1,74) = 4.43, p = 0.039). None of the other main or interaction effects reached the level of significance (SPAI: F(1,74) = 0.12, p = .729, group: F(1,74) = 2.89, p = .093, distance × SPAI: F(1,74) = 1.79, p = .185; group × SPAI: F(1,74) = 0.05, p = .820).

To follow up on the significant three-way interaction, we calculated separate mixed models within each group, again selectively focusing on fixations on persons. In the real-life group alone, we found a significant main effect for distance (F(1,24) = 50.19, p < .001) and, importantly, a distance × SPAI interaction (F(1,24) = 5.97, p = .022), but no significant main effect for SPAI (F(1,24) = 0.10, p = .753). In the laboratory group alone, we found a significant main effect for distance (F(1,27) = 188.73, p < .001), but no main effect for SPAI (F(1,27) = 0.14, p = .715) and no distance × SPAI interaction (F(1,27) = 0.02, p = .892).

Specifically, participants in the real-life group, but not the laboratory group, showed a reduced bias to gaze at near compared to distant people when high on social anxiety. This pattern of results was similar when relying on the SIAS score

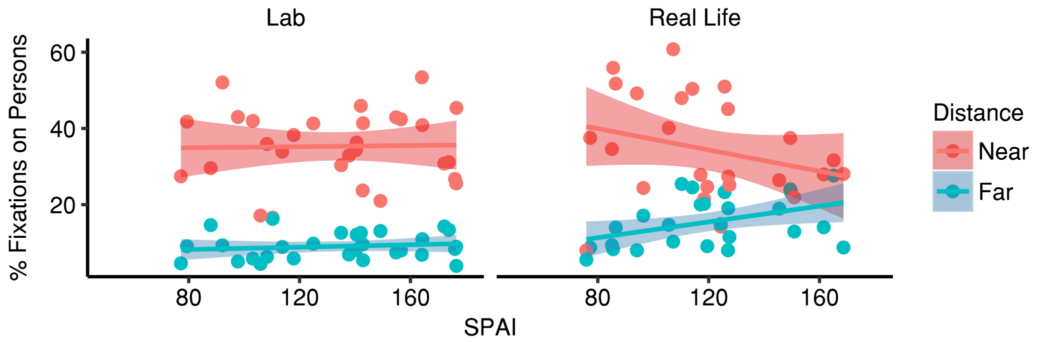

Figure 3.2: Frequency of gaze at persons as a function of group, distance and SPAI. Shaded areas represent 95% confidence intervals of the linear regression.

instead, while general anxiety as measured by the STAI did not modulate gaze behavior (see Supplementary Results). In a parallel model across both groups with fixations on objects (instead of persons) as dependent variable, SPAI scores were in no way related to the frequency of fixations on objects in real-life or in the laboratory (see Supplemental Results and Figure 6.3).

Discussion

The present study found several general similarities and differences in viewing behavior between matched participants in a real-life and a laboratory group. Specifically, participants looked at objects and persons more frequently than at the path in both groups and exhibited a general preference for looking into the vicinity (vs. distance), which was, however, enhanced in the laboratory group. Unlike other studies that utilized gaze monitoring in naturalistic conditions (Foulsham et al., 2011), we did not find a general avoidance of gaze at other persons in real life, but rather more fixations on near as compared to distant persons in laboratory as well as in field conditions.

Interestingly, we found moderate to high stability in participants' viewing preferences for all categories and distances, both in the laboratory and the real-life group. Given the high variability in viewing conditions between participants, this consistency seems surprising but similar results have been reported for participants watching videos on a computer screen (Rubo & Gamer, 2019). To the best of our knowledge, such consistency in viewing patterns has not yet been documented in a naturalistic situation outside the laboratory but it confirms that the current findings were not driven by a small number of exceptional samples in individual recordings. Crucially, extensive measurements of stable response patterns were argued to partially provide self-replication within a single experiment and can explain why certain fields relying on experiments with small sample sizes never faced a replication crisis (Smith & Little, 2018). Nonetheless, while high intra-personal stability does highlight the robustness of observations made within the present sample, future research will need to test a different group of individuals to better estimate the generalizability of between-subject effects.

Regarding the influence of social anxiety on viewing patterns, we found a bias to look more frequently at near compared to distant people, which was not modulated by social anxiety in the laboratory group. In the real-life group, by contrast, the bias to look at near people was reduced with increasing levels of social anxiety; that is, while individuals high on social anxiety did not generally avoid looking at other people in real-life, they had a reduced preference to do so when conspecifics were near compared to less socially anxious individuals. This compensatory pattern was furthermore specific to social attention and was not present in attention towards objects, showing that social anxiety specifically affected attention to persons, not attention in general. While clinical observations and self-reports (Schneier et al., 2011) canonically report an avoidance of gaze at other persons in socially anxious individuals, the present study extends these observations by showing that socially anxious observers may instead prefer to gaze at others at a greater distance compared to less socially anxious individuals. In our opinion, this observation seems plausible from a theoretical point of view considering that conspecifics' scrutiny - the target of fear in social anxiety (Stein & Stein, 2008) - might be preferentially elicited when the conspecific is near enough to detect and reciprocate one's gaze or to even initiate a conversation following eye contact. Since socially anxious individuals are not thought to be less interested in the social world per se, they may, compared to less socially anxious individuals, preferentially allocate overt attention towards people when they are located in the distance, thereby satisfying their need for social information while anticipating (and thereby coping with) their fear of scrutiny. However, we would like to stress that future research will need to address this proposed mechanism more directly by investigating gaze behavior in social situations in which the distance of other people varies systematically.

By providing the first direct comparison of gaze behavior in two participant groups matched on social anxiety viewing the same stimulus material in real-life and in a laboratory situation, the present study resolves a long-standing conflict between clinical practitioners' observations and laboratory-based research. Critically, the absence of gaze modulation by social anxiety in the laboratory group substantiates previously expressed concerns (Risko et al., 2016) that laboratory-based passive viewing tasks may not provide an appropriate proxy for real-world social attentional phenomena in humans, likely because participants are aware that persons in a video will not be able to evaluate them. Interestingly, absence of atypical viewing behavior towards images of persons was also reported in autism (Rutherford & Krysko, 2008), although a substantial amount of clinical reports document deviations in social attention in real-world social situations in these patients as well (Senju & Johnson, 2009).

Further support for the hypothesis that the presence of real persons may stimulate distinctly different attentional processes compared to the presence of images comes from related fields (Risko & Kingstone, 2010). For instance, it was shown that compared to averted gaze, direct gaze elicited larger visual brain responses (Pönkänen, Alhoniemi, Leppänen, & Hietanen, 2010), stronger left-sided frontal EEG alpha-asymmetry (Pönkänen, Peltola, & Hietanen, 2011) and a larger skin conductance response (Pönkänen et al., 2011), but only in a live condition and not when participants viewed images of the same faces gazing at them.

By recruiting a stratified sample of participants on the basis of a pre-screening procedure, we were able to cover a broad range of social anxiety traits ranging from low to medium levels to (sub-)clinical symptoms. This approach complies with the conceptualization of social anxiety as a continuum (Rapee & Spence, 2004) with subjects at the upper end representing high degrees of social fear (commonly diagnosed as SAD). Nevertheless, future studies should examine the stability and generalizability of the current findings to individuals who received the formal diagnosis of a social anxiety disorder or to compare such persons to other patient groups to elucidate the specificity of the current findings. Another aspect calling for a more detailed inquiry is the precise distance at which gaze at other persons is reduced or enforced in social anxiety. Grounded in basic models of visual perception (Daum & Hecht, 2009), we classified gaze as falling either in the vicinity (8 m or nearer) or distance (beyond 8 m) of the participants. This data aggregation strategy was also motivated by technical boundaries, as present eye-tracking devices do not measure the distance to the looked-at object. Future studies may achieve more fine-grained data acquisition either by employing virtual reality technology, where precise distance measures are easily available (Ben-Moussa, Rubo, Debracque, & Lange, 2017), or by installing identification marks in the test environment to allow for a more precise estimation of distances.

Future research should furthermore aim at examining the generalizability of the current findings by comparing viewing behavior across different situations within the same participants. In the present study, viewing behavior was investigated in only one type of social situation, a populated train station. To avoid an influence of memory effects, the laboratory condition was not presented to the same persons as the real-life condition, but to other participants matched on gender and social anxiety. In order to better estimate the generalizability of the present findings, participants in future research should be confronted with a larger variety of everyday situations (e.g., pedestrian zones, supermarkets, waiting rooms, concerts, sports events etc.). This will allow to better highlight the specificity or generalizability of gaze patterns across environments, and, by randomly assigning participants to a real-life or laboratory version of the same scenes, will help to better understand influences of social presence (as manipulated via different scene types) on viewing behavior within the same individuals.

Summing up, the present study is the first to directly compare gazing behavior of persons with a high range of social anxiety symptoms both in a real-life and a closely matched laboratory group. Extending on clinical observations, we found high social anxiety to be associated with a relative avoidance of gaze at near compared to distant people in real-life, but no such modulation of gaze behavior by social anxiety in the laboratory group. Our findings furthermore provide a basis for recently expressed assertions that the field of social attention needs to move beyond laboratory research and into real-world situations to do justice to all basic social mechanisms that are at the heart of social attention (Schilbach et al., 2013). This applies all the more to research on impairments of social functions in psychiatric conditions.

Data availability The data that support the findings of this study are available at https://osf.io/h7pf2/.

Acknowledgements This work was supported by the European Research Council (ERC-2013-StG $#$336305). The authors thank Jasin Sahraoui for his help during data collection.

Author Contributions M.R., M.G. and L.H. designed the experiment. M.R. collected and analysed the data. M.G. supervised data analysis. M.R., M.G. and L.H. wrote and reviewed the manuscript.

CHAPTER 4

Social Attention in Virtual Reality

Natural Gaze Behavior and Stable Intra-individual Gaze Preferences in a Social Virtual Scenario with a Reactive Virtual Agent

Marius Rubo, Matthias Gamer

Julius-Maximilians-Universität Würzburg

Corresponding Author: Marius Rubo

Email adress: marius.rubo@uni-wuerzburg.de

Manuscript submitted for publication.

Abstract

People show a robust tendency to gaze at other human beings when viewing images or videos, but were also found to relatively avoid gaze at others in several real-world situations. This discrepancy, along with theoretical considerations, spawned doubts about the appropriateness of classical laboratory-based experimental paradigms in social attention research. Several researchers instead suggested the use of immersive virtual scenarios in eliciting and measuring naturalistic attentional patterns, but the field, struggling with methodological challenges, still needs to establish the advantages of this approach. Here we show that in a complex social scenario displayed in virtual reality, participants distribute relatively little overt attention towards a virtual agent's face compared to virtual objects, but nonetheless reacted dynamically to the agent's social behavior. Moreover, participants exhibited stable intra-individual gaze preferences towards the face and objects. The present study suggests that reactive virtual agents observed in immersive virtual reality can elicit natural modes of information processing and can help to design ecologically valid experiments while maintaining high experimental control.

Introduction

Humans pay attention to other human beings in many situations and in various ways: they preferentially gaze at human heads and eyes (Birmingham et al., 2008) as well as objects gazed at by conspecifics (Borji et al., 2014) when viewing images or videos, automatically form representations of conspecifics' tasks even when there is no incentive to do so (Sebanz, Bekkering, & Knoblich, 2006) and show a similar inhibition when reaching towards objects which were previously touched by themselves or by another person (*social inhibition of return*, Welsh et al., 2007).

However, while several phenomena in the field of social cognition are typically investigated in real social situations (i.e., where a conspecific is physically present), gaze behavior is often investigated in participants viewing images or videos of social situations. Interestingly, several studies measuring gaze allocation while other people were physically present found an avoidance of, not a preference for looking at others (Gallup, Chong, & Couzin, 2012; Laidlaw et al., 2011), contrasting findings from laboratory-based studies. Furthermore, a recent study reported a modulation of social gaze in socially anxious persons, but only in a real-life situation and not in a matched laboratory condition (Rubo, Huestegge, & Gamer, n.d.).

This discrepancy spawned claims that passive viewing of images or videos may represent inept proxies for real-world social attentional phenomena (Risko et al., 2016). Specifically, it was argued that in natural social situations, people are aware of their conspecifics' responsiveness and therefore use their gaze not only to collect visual information, but also to signal their mental states to others. It was argued that in several social situations, there exist social norms which interdict "staring" at others, leading individuals to avoid rather than seek direct social gaze (Zuckerman, Miserandino, & Bernieri, 1983). Since such social norms are believed to be activated in the presence of other persons, but not by depictions of persons, it is argued that social attention should be studied within natural social contexts where attentional processes operate in unadulterated ways.

When investigating attention in real-life situations, however, precise experimental control is virtually impossible, and even confederates are not capable of reacting in identical ways towards each participant. As one methodological approach to simultaneously achieve high ecological validity as well as experimental control, researchers have suggested the use of virtual reality (VR) technology (Pan & Hamilton, 2018). Here, the idea is to model experiences in real-world social situations more closely compared to traditional laboratory-based setups, with the participants being able to freely look around in the scenario and virtual agents naturally acting and responding the the participant. Critically, observing scenes in VR typically creates a sense of *presence* or *being there* in the scenario (Skarbez et al., 2017), which is not the case when looking at depictions on a computer screen. Nonetheless, there is still only relatively scarce empirical evidence that social scenarios observed in virtual reality really do elicit natural modes of information processing (Kulik, 2018). Zimmer, Buttlar, Halbeisen, Walther, & Domes (2019) recently found comparable self-reported, autonomic and endocrine stress markers in response to a standardized social stress test carried out by real or virtual conspecifics. Wienrich, Gross, Kretschmer, & Muller-Plath (2018) documented an inhibition of return effect towards virtual avatars — an effect which was previously believed to only be triggered by other human beings. In a study by Gallup, Vasilyev, Anderson, & Kingstone (2019), by contrast, contagious yawning was inhibited in the presence of a real conspecific, but not in the presence of a virtual agent. In the latter study, however, the virtual agent showed no meaningful social behavior — a possible prerequisite for perceiving an agent as believable and activating natural modes of social information processing.

In the present study, participants were located in a complex social scenario in virtual reality and had the opportunity to observe a reactive virtual agent with a naturalistic behavioral repertoire. Assuming that such a virtual agent elicits naturalistic modes of information processing, we expected a relative avoidance of participants' gaze at the agent's face, but a reactivity to the agent's social behavior (i.e. looking and smiling at the participant). We additionally tested whether interindividual differences in attentional preferences were stable throughout the experiment — a largely unexplained phenomenon which was recently discovered in participants viewing images (de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2018), videos (Rubo & Gamer, 2018) and a real-life situation (Rubo et al., n.d.). In sum, the present study aims to reveal more detailed insights into gaze behavior in a reactive virtual environment. We tackled prevailing methodological challenges, raising the bar in terms of the virtual scene's semantic credibility and hoping to stimulate the promising but largely unexplored new field of social attention research using virtual reality.

Methods

Participants, Apparaturs and Software

Forty participants (32 female, mean age = 24.62 years, *SD* = 6.83 years) watched a virtual reality scene using an HTC Vive system (2160 × 1200 pixel resolution, 110° field of view) while a built-in eye-tracker (SMI Eye Tracking Glasses) recorded their gaze direction from both eyes at a sampling rate of 60 Hz. Experimental stimuli were displayed using the Unity 3D Game Engine (https://unity3d.com) and the virtual agent's behavior was controlled using in-house software available at github.com/mariusrubo/Unity-Humanoid-TransportObjects which was built using an inverse kinematics algorithm package (www.root-motion.com). The sample size and experimental procedure were determined a priori and were pre-registered (AsPredicted #14290, see https://aspredicted.org/5wj8u.pdf). All participants had normal or corrected-to-normal vision. The study was approved by the ethics committee of the Department of Psychology, University of Würzburg and conducted in accordance with the Declaration of Helsinki.

Procedure

Participants were invited to the laboratory individually, informed about the purpose of the study and completed an informed consent form. They were instructed to merely observe the scenario (see Figure 4.1) for about six minutes as if they were really there, but not walk away from the location where they were positioned. The virtual scenario consisted of a suburban neighborhood in which a garage sale was being prepared on a driveway in front of a house. Participants found themselves located at the driveway while a virtual character — a woman in the age of roughly 60 years — was engaged in carrying household objects (e.g. a radio, a lamp, a toy car) from her house onto two tables positioned approximately 2 meters in front of the participant. Participants were specifically asked to imagine that they had come across this scene by chance while they were going for a walk in the neighborhood and had decided to observe the scene for a while.

Altogether, the virtual character carried ten items onto the tables during the experiment, each time approaching the participant in a similar manner: After leaving the veranda with an item in her hands, the character walked a straight path with a distance of approximately 12 meters through the front yard before placing the item in front of the character. The character then turned around and walked the same itinerary back and into her house, picking up the next item. Participants were therefore confronted with ten similarly structured opportunities to gaze at the character or the object carried in her hands as she approached (ten *trials*). The character furthermore glanced at the participant and smiled in random trials (but never showing the same behavior in more than two consecutive trials), starting at a distance of approximately four meters.

Prior to the start of the scenario, the eye-tracker was calibrated using a 5-point-calibration technique. Gaze measurement validation was performed using in-house software to allow for a drift correction procedure and to achieve the most detailed insight into potential problems with gaze measurement. The general implementation of the procedure is described at github.com/mariusrubo/Unity-EyeTracking-RegionsOfInterest. Specifically, we presented a red sphere in the virtual scenario with a diameter of 10 cm and a distance to the participants of about 4 meters. Participants were instructed to gaze at the sphere until it vanished about three seconds later. This allowed us to compare the participant's gaze rays from both eyes as they were estimated by the eye-tracker with hypothetical gaze rays perfectly hitting the sphere's center. We geometrically transposed this arrangement to represent the measured and the ideal rays' relative slope as a 2-dimensional deviation, comparable to the shot on a target and parallel to monitor-based eye-tracking where gaze deviations are described along the screen's x- and y-coordinates. This validation process was performed directly after the initial calibration and repeated several times throughout the experiment. Specifically, validation was performed in moments when the virtual avatar was walking back into her house, since the participants' natural gaze behavior in these moments were not of interest for the present study. If validation was unsuccessful, calibration was performed again. Data from all validation procedures was used to correct for drifts in gaze measurement (see *Data processing*).

Upon completing the experiment, participants were asked to report on their sense of *presence* in the virtual environment (Skarbez et al., 2017) using the German version of the *Igroup Presence Questionnaire* (IPQ; Schubert et al. (2001)) as well as on sensations of simulator sickness using the *Simulator Sickness Questionnaire* (SSQ, Kennedy et al. (1993)). Participants then filled out a sociodemographic questionnaire. Participants gave moderate ratings for presence (M = 57.25, SD = 11.06, on a scale from 14 to 98) and low ratings for simulator sickness (M = 3.35, SD = 3.29, on a scale from 0 to 48).

Data Processing and Statistical Analyses

Data were analyzed using R for statistical computing (version 3.2; R Development Core Team, 2015). We were only interested in participants' gaze behavior while the virtual character approached the participants, i.e. from the moment the character stepped onto the garden path (at a distance of 12 meters from the participants) to the moment the character placed the object on the table in front of the participants and turned around.

Gaze data collected during the validation procedures was used to correct for drifts in gaze measurements resulting from shifts in the HMD's position on the participants' head. To exclude moments when participants did not look at the validation sphere during the validation procedures (e.g. when the sphere had only just appeared and participants were not yet looking at it), drift-correction was performed using only data points when participants' gaze from both eyes missed the validation sphere by less than 15° (this was the case in 93.25%, SD = 9.59%, of data points in all validation procedures). In one participant, the eye-tracker

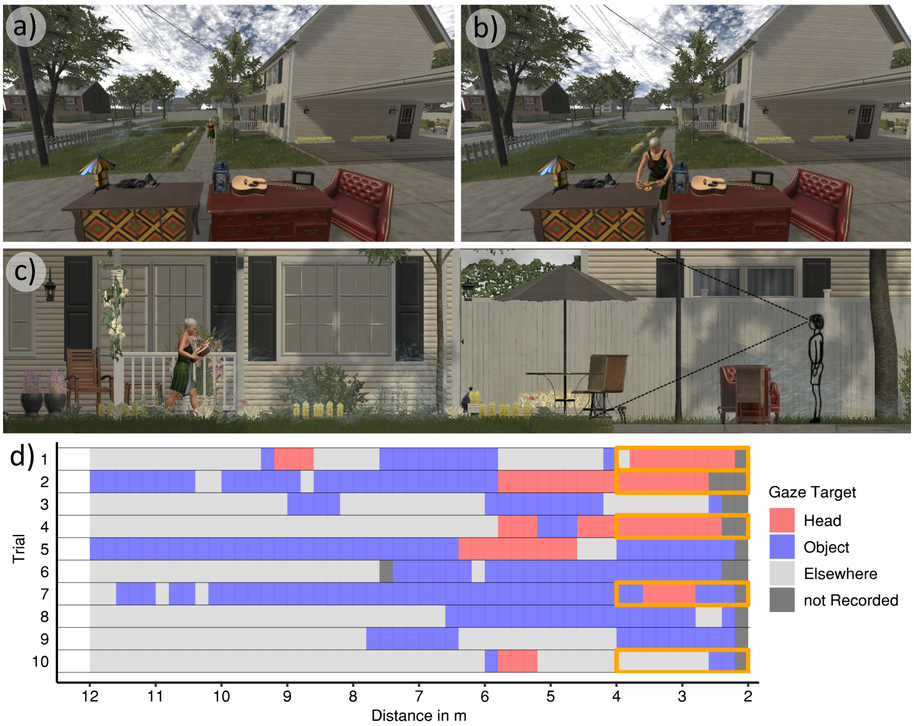

Figure 4.1: The social situation taking place in the virtual scenario. Images (a) and (b) are from the perspective of the participants in moments when the virtual character has just left her house carrying an object and when she places the object on the table in front of the participant, respectively. c) shows the scene from a lateral perspective (as an orthographic projection), with the participant and her typical viewing angle (black dotted lines) pictured on the right of the image. d) illustrates one exemplary participant's gaze while the virtual character is approaching the participant. The distance scale corresponds to c). For visualization purposes, gaze is categorized as being directed towards the character's head or the object if either is being gazed within a distance bin of 20 cm. Orange frames indicate moments when the virtual character looked at the participant and smiled.

only correctly tracked the left eye (during all validation procedures, observed gaze from the left eye missed the validation sphere by 2.45°, while observed gaze from the right eye missed the sphere by 64.69°). In this case, only data from the left eye was used for further analyses.

We used a recursive outlier removal algorithm to estimate drift in each validation process. Separately for the deviation measurements in x- and y-direction of each eye, the lowest and highest values were both removed from the distribution, individually compared to the distribution of the remaining data and entered again if they were located within 3 standard deviations from the mean. This process was recursively applied to the remaining data until both the highest and the lowest data points met the criterion to be re-entered to the distribution (for a similar procedure see End & Gamer, 2017b, and @Rubo_2018). The values in the remaining distributions were averaged to arrive at *baseline deviation* values.

Gaze data was represented as deviations from two regions of interest (ROIs) — the character's face and the object she was carrying — in an x and y-direction and separately for both eyes. We subtracted baseline deviation values from both eyes' measurements individually and then averaged deviations from both eyes to arrive at a deviation representing both eyes' gaze. Gaze data in which either of the two eyes could not be tracked during the relevant phases of the experiment either because of the participants' eye blinking or technical issues (M = 7.83%, SD = 7.15% of the time) were removed from the dataset. Based on these data, we classified for each time point if gaze was directed towards the virtual agent's head, the object in her hands, or elsewhere. To achieve identical a priori probabilities for gaze at both ROIs, gaze was classified to hit either of the two if it collided with a sphere of a diameter of 30 cm positioned at the same place as each region of interest. This procedure implies that the angular range in which gaze is classified as being directed towards one of the two ROIs increases as the virtual agent approaches (illustrated in Figure 4.2). Note that unlike typical experimental setups measuring gaze in real-life (e.g. Rubo et al., n.d.), the approach used here does not require manual classification of fixated locations and allows to more precisely identify the distances at which a ROI is gazed at (illustrated in Figure 4.1d). For each participant and trial, we aggregated relative gaze dwell times towards both ROIs within distance bins with a width of one meter, and, for analyses across trials, further aggregated gaze dwell times across trials. This approach allows to analyze the relative gaze dwell times at the two ROIs as a function of the distance towards the ROIs.

We tested the influence of individual parameters on gaze allocation by means of linear mixed models where participant ID was inserted as a random effect. All parameters were tested for significance using an F-test, with α set to 0.05. Reported correlations are Neyman-Pearson correlations.

Results

Gaze Allocation as a Function of the Agent's Distance and Social Behavior

We first aimed to describe the gaze pattern as a function of the character's distance (see Figure 4.2), therefore including the ROI (either the character's head or the object in her hands), the distance as well as a ROI × distance interaction as fixed effects into a linear mixed model and the percentage of gaze dwell time as dependent variable. We found a significant effect of ROI ($F(1{,}1557) = 575.25$, $p < .0001$), distance ($F(1{,}1557) = 352.08$, $p < .0001$) and the ROI × distance interaction ($F(1{,}1557) = 83.44$, $p < .0001$). Overall, the head was gazed at 6.05% (SD = 10.60%) of the time, while the object was gazed at 23.29% (SD = 24.38 %) of the time. There was a negative correlation between distance and gaze dwell time for both regions of interest which was somewhat stronger for objects (r = -.41, 95%-CI = [-.48, -.35], $p < .0001$) than for the character's head (r = -.33, 95%-CI = [-.39, -.26], $p < .0001$). In sum, participants spent more time looking at the object compared to the character's face, and furthermore increased gaze dwell time towards the object more strongly as the character approached compared to gaze dwell time towards the character's face.

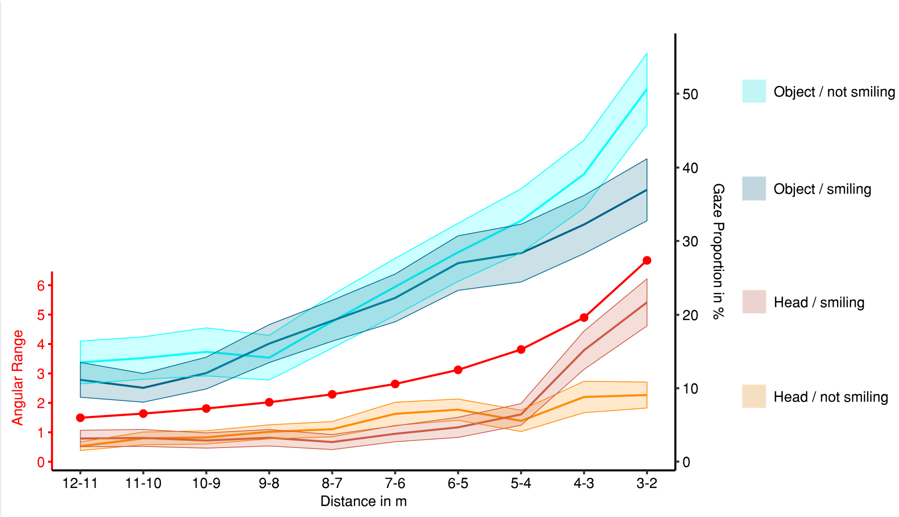

Figure 4.2: Gaze allocation as a function of the virtual character's distance and its social behavior (smiling at the participant while approaching or merely looking at her hands and not smiling). Gaze data is allocated within distance bins with a width of one meter. Ribbons represent SEM. The red points (and the corresponding y-axis on the left of the figure, also printed in red) highlight the angular range in degrees within which the participants' gaze is classified as being directed towards each of the two regions of interest. Note that as the virtual character approaches, this angular range naturally enlarges, thereby possibly increasing the measurement's sensitivity.

We then focused on time points when the virtual avatar was closer than 4 meters to the participant — the distance range in which she looked and smiled at the participant in selected trials. Here we describe the relative gaze dwell time as a function of the region of interest, the character's social behavior (smiling or not smiling) and an interaction of these two factors. We again found a significant effect of the ROI ($F(1{,}277) = 139.39$, $p < .0001$), furthermore a significant ROI $\times$ smiling interaction ($F(1{,}277) = 20.01$, $p < .0001$), but no statistically significant main effect of smiling ($F(1{,}277) = 0.03$, $p < .86$). The head was gazed at 18.41% ($SD = 18.64\%$) of the time when the character smiled and 8.93% ($SD = 12.31\%$) when she did not smile. The object was gazed at 34.61% ($SD = 25.78\%$) of the time when the character smiled and 44.88% ($SD = 30.34\%$) when it did not smile. In sum, participants clearly reacted to being smiled at by the virtual agent, shifting their attention away from the object in her hands and towards the agent's face.

Variations in Gaze Behavior throughout the Experiment

We subsequently analyzed the stability of gaze patterns throughout the course of the experiment (see Figure 4.3). Again focusing on time points when the virtual avatar was closer than 4 meters, we describe gaze dwell time as a function of the trial, the region of interest (ROI), the avatar's smiling as well as a trial × ROI interaction, a trial × smiling interaction, a ROI × smiling interaction and a trial × ROI × smiling interaction. Like above, we found a general effect of the ROI ($F(1{,}753) = 176.69$, $p < .0001$) and a ROI × smiling interaction ($F(1{,}753) = 28.19$, $p < .0001$). This analysis furthermore revealed a trial × ROI interaction ($F(1{,}753) = 32.41$, $p < .0001$) as well as a trial × smiling × ROI interaction ($F(1{,}753) = 1.55$, $p = .03$). Neither a main effect of the trial ($F(1{,}753) = 0.32$, $p = .57$) or the character's smiling ($F(1{,}753) = 0.22$, $p = .64$) nor the trial × smiling interaction ($F(1{,}753) = 1.55$, $p = .21$) were statistically significant.

There was a negative correlation between gaze dwell time on the character's head and the experiment's trial ($r = -.25$, 95%-CI = [-.34, -.15], $p < .0001$), which was somewhat stronger for trials in which the character smiled ($r = -.36$, 95%-CI = [-.47, -.23], $p < .0001$) compared to trials in which the character did not smile ($r = -.14$, 95%-CI = [-.27, .00], $p = .05$). By contrast, gaze dwell time on the object increased along the experiment's trials ($r = .14$, 95%-CI = [.04, .23], $p < .01$), and the correlations here were more similar in trials in which the character smiled ($r = .17$, 95%-CI = [.03, .30], $p = .02$) and trials in which the character did not smile ($r = .13$, 95%-CI = [-.01, .26], $p = .07$).

Consistency of Gaze Behavior within Participants

Finally, we estimated the consistency of gaze preferences within individual observers. To this end, we randomly split the ten trials into two halves for each participant, aggregated relative gaze dwell times on both regions of interest across both trial groups (both along the entire walking distance and for the phase when the agent was nearer than 4 meters) and computed split-half correlations of gaze

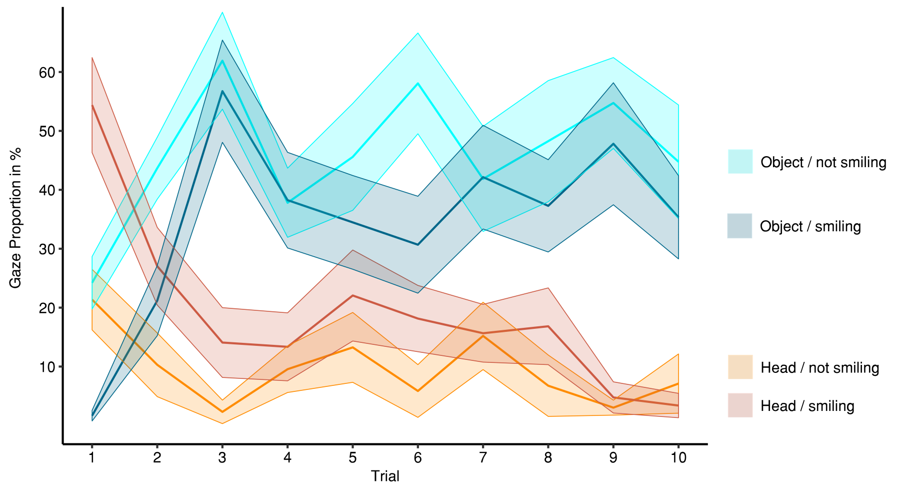

Figure 4.3: Gaze allocation along the experiment's ten trials while the virtual agent was nearer than 4 meters. Note that in each trial and for each participant, the virtual agent either did or did not smile at the participant while approaching, and data is therefore aggregated along different subsets of participant within each trial and condition (smiling vs. not smiling). Error ribbons represent SEM.

dwell times between the two halves. We repeated this procedure 1000 times, and determined the 2.5^{th} and 97.5^{th} percentile rank among correlation estimates in all iterations as an unbiased estimate of the 95% confidence interval. For the entire distance, split-half correlation in gaze allocation were r = .55 (95%-CI = [.37, .72]) for the face and r = .82 (95%-CI = [.73, .89]) for the object. When only considering phases when the virtual agent was near, split-half correlations were r = .42 (95%-CI = [.19, .63]) for the face and r = .73 (95%-CI = [.63, .82]) for the objects.

We furthermore performed the above procedure within a sliding window of four trials (i.e. trials 1-4, 2-5 etc.) in order to more closely investigate the temporal emergence of individual gaze preferences along the course of the experiment (see Figure 4.4). Inter-individual differences in gaze preferences were present both when considering the entire time of each trial and when considering only phases when the virtual agent was near. However, along the entire trials, gaze preferences were present right from the experiment's beginning, but only emerged after several trials when considering only moments when the agent was near. Moreover, gaze preferences were generally more distinct with regards to gaze at objects compared

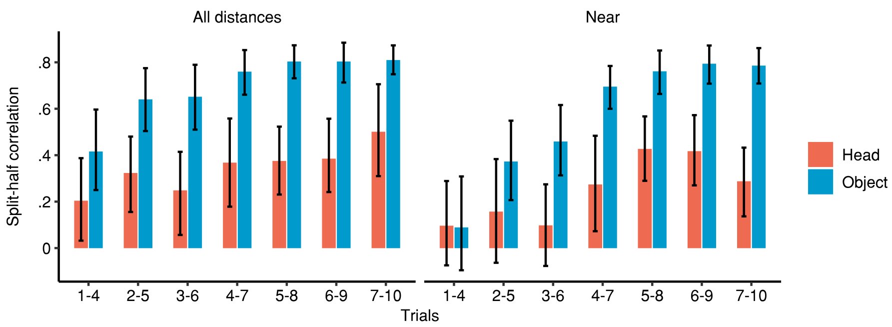

Figure 4.4: Split-half correlations of gaze dwell time for gaze towards the virtual agent's face and the carried object along a sliding window with a width of four trials, separately for the entire trial and the phase when the agent was nearer than four meters. Error bars represent 95% confidence intervals.

to gaze at the agent's face.

Discussion

The present study confronted participants with a naturalistic social situation and a reactive social agent presented in a virtual environment. Participants spent more time fixating objects in the agent's hands compared to the agent's face. Participants increased gaze dwell times towards both of these regions of interest while the agent approached — an effect which may partly result from an increased measurement sensitivity —, but did so more strongly for the object compared to the face. At the same time, fixation time towards the agent's face was increased in moments when the agent looked and smiled at the participants, highlighting a reactivity to the virtual agent's social behavior. The participants' gaze behavior is in line with the idea of a *civil inattention*, a social norm which prohibits starring at others when one is not engaged in a conversation (Zuckerman et al., 1983). Such relative gaze avoidance is typically not found in the context of laboratory experiments, but was found in several (Gallup et al., 2012; Laidlaw et al., 2011), though not all (Rubo et al., n.d.) investigations of gaze behavior in real-life situations.

Participants furthermore modulated their gaze behavior along the course of the experiment: While the first trials were characterized by a stronger tendency to gaze at the agent's face — especially when she smiled at the participants —

participants more strongly prioritized gazing at the objects during the second half of the experiment. This behavioral pattern may be explained by a general bias towards novelty in visual attention (Desimone & Duncan, 1995), but may also in part result from complying to the civil inattention norm: Looking at a conspecific's face at the first encounter may be a sensible social behavior, while repeated looking at the conspecific in the absence of a conversation could be considered inadequate starring.

We additionally found robust inter-individual differences in gaze allocation towards the virtual agent's face as well as towards the object in her hands. Interestingly, gaze preferences for the objects were more tangible compared to gaze preferences for the face, highlighting a comparably higher variability in individuals' social attention compared to attention towards objects. When investigating gaze allocation along the experiment's time course, gaze preferences in the beginning of the experiment were weaker when considering the whole trials and even absent when considering only phases when the agent was near. Conversely, gaze preferences reached their highest tangibility in the second half of the experiment. In our view, such an incremental emergence of intra-individual gaze preferences hints towards the presence of modulating variables at the beginning of the experiment — e.g. novelty effects may exert a differential influence on different participants — while in later trials, habitual attentional processes may surface more strongly. This finding is in line with previous research which showed gaze preferences when viewing videos (Rubo & Gamer, 2018) and in real-life (Rubo et al., n.d.), but partly contrasts documented gaze preferences from the first saccade after stimulus onset when viewing images (de Haas et al., 2018). Several studies have furthermore hinted towards a genetic component in social gaze, as evident in a reduced tendency to attend to semantically salient regions in persons with autism spectrum disorder (Wang et al., 2015) and in higher concordance between identical (monozygotic) compared to fraternal twins (Kennedy et al., 2017). However, little is known about the causes and consequences of such gaze preferences in the general population.

The present study plunged into exploring social attention in a lifelike virtual scenario where a reactive virtual agent staged a complex working process taken from real life. Participants exhibited signs of naturalistic attentional processes, suggesting that such a complex virtual scenario — though still laborious to establish with current software development techniques — may be worthwhile in gaining more detailed insights into the precise sequence of social attentive phenomena in naturalistic environments. As it is naturally the case when investigating behavior in more lifelike scenarios, many variables were not individually controlled, diluting attributions of statistical effects to clear causal mechanisms. For instance, we did not systematically vary the agent's facial expression when looking at the participant (she always began to smile in these occasions), leaving us unable to decide if participants reacted to the agents *looking* at them or *smiling* at them. We furthermore did not systematically vary the agent's general appearance, her walking speed, or even the scenario's arrangement as a whole.

However, in our view, future research should not prioritize the systematic variation of individual variables since this strategy would bear the risk of creating stilted and unnatural scenarios which cannot harvest the methodology's potential to stimulate naturalistic attentional processes. Instead, we believe that social attention research using virtual reality should employ even less tabulated scenarios than the one used in the present study, possibly even abandoning the repetition of homogeneous trials and instead immersing participants in social situations as semantically rich as many situations we encounter in real life. Note that it is the repetition of similarly structured trials used in the present study which make it more difficult to differentiate if the decline in attention towards the virtual agent's face reflects civil inattention, or more simply the fading of a novelty effect. Confronting participants with a variety of semantically coherent (and not necessarily systematically varied) situations may nonetheless allow to carve out contributions of individual factors to attention allocation, but will also help to identify broad and robust patterns which can remain hidden in overly standardized experiments (Würbel, 2000). The present study suggests that this strategy may enable researchers to investigate truly lifelike behavior while at the same time being able to confront participants with identical stimuli.

Data availability The data that support the findings of this study are available at https://osf.io/nqv8x/.

Acknowledgements This work was supported by the European Research Council (ERC-2013-StG #336305). We thank Jasin Sahraoui for his help in data collection.

Author Contributions M.R. and M.G. designed the experiment. M.R. collected and analysed the data. M.G. supervised data analysis. M.R. and M.G. wrote and reviewed the manuscript.

CHAPTER 5
DISCUSSION

The present work illustrates and contextualises three studies investigating gaze behavior in diverse social situations, each aiming to illuminate a different aspect in the vast field of social attention. While each study does present novel data and, arguably, interesting relationships in a specific subfield, they all leave nagging research questions unanswered.

Study 1 found participants' gaze to be similarily attracted by physical saliency and human faces when viewing videos, providing the first direct side-by-side comparison of these two classes of predictors when viewing dynamic stimuli and perhaps contributing to a harmonization of the rivaling scientific fields behind them. Although this study documented the presence of both social and saliency-driven attentional mechanisms while watching videos, it did not differentiate if both factors exert a similar effect throughout each video's timecourse, or if specific subsystems take over the attentional steering wheel in predictable key moments. While it is possible to identify moments when gaze from several viewers converges (e.g. Kennedy et al., 2017) and there exist more detailed accounts of what types of objects especially attract attention (e.g. objects which are known to produce sounds or smell, Xu et al., 2014), it remains largely unclear if more specific predictions of gaze are generally possible: An ideal model, one may argue, would predict the likelihood of gaze at a specific area as a function of the specific image as well as background knowledge about the depicted situation (Henderson & Hayes, 2017), but also the viewer's characteristics. However, such specific predictions about gaze allocation may be a distant vision considering that even seamingly basic questions like the general frequency of spontaneous fixations towards faces are still debated (Pereira, Birmingham, & Ristic, 2019).

Study 2 investigated gaze behavior in a real-life situation and a closely matched laboratory situation, and found that social anxiety affected gaze in the real-life condition where conspecifics were physically present, but not in the laboratory condition. This finding is theoretically plausible considering that socially anxious individuals do not fear other people per se, but their scrutiny (Stein & Stein, 2008) and that scrutiny need not be feared when observing people on a computer screen. The finding is furthermore corrobated by clinical descriptions of social anxiety, but nonetheless novel in experimental research. Despite its interesting core finding, this study, too, may be seen as a methodologically crude approach to social attention as it mostly reflects on aggregated gaze dwell times across a timecourse of several minutes. Here again, future research may aim at developing a more fine-grained understanding of gaze allocation. Specifically, the functional roles of gaze and its interaction with verbal and other non-verbal aspects of language during or in the initiation of conversations should be described. Indeed, seemingly subtle details in a social situation may delicately, but predictably influence social gaze. In a study by Macdonald & Tatler (2018), participants looked at a partner's eyes longer if the dyad had been assigned specific roles (*chef* and *gatherer* in a kitchen) than if they had not been assigned social roles. Fixating an interacting partner's eyes furthermore often occurs in specific moments like the end of a speaking turn (Hessels, Holleman, Kingstone, Hooge, & Kemner, 2019), and this specific action may be modulated by the (perceived) social role or personal characteristics such as social anxiety.

Study 3 employed virtual reality technology to investigate social gaze in a naturalistic, but still completely controlled situation. While virtual reality scenes can be undoubtly controlled to a high degree, it remains an open questions whether they really do excel at eliciting less adulterated information processing compared to traditional laboratory-based methodology (Kulik, 2018). Participants in this study showed signs of naturalistic social behavior during the experiment: they frequently gazed at an avatar's face — especially when the avatar gazed at the participant and smiled — but also at an object which the avatar was carrying. More intriguingly, however, is a finding which marks a common denominator among all three studies in this work: Participants showed stable gaze preferences for the avatar's face or objects in its hands across several trials in the experiment. A similar effect was found in study 1 in this work where inter-individual differences in predictor strength for physical saliency and human faces were robust across videos, and also in study 2, where inter-individual differences in gaze allocation to humans as well as objects were robust across different time bins throughout the experiment.

The moderate to high stability in gaze preferences in the three studies presented here came as a surprise, as very few in the numerous eye-tracking studies published so far report similar findings. One explanation for this research gap may be a wide-spread omission to inquire psychometric properties in gaze data across different trials of an experiment. Unlike explicit measures such as personality questionnaires, which are routinely and extensively tested for psychometric properties, gaze behavior is mostly analyzed at a group level, typically after averaging data from individual trials. Smith & Little (2018) criticized this emphasis and argued that any quantitative measurements from humans, not only questionnaire data, should primarily be analyzed with regards to functional relationships at the individual level, and only secondarily at the group level by averaging data.

What previous research did show is alterations in gaze behavior in several neuropsychiatric conditions such as attention deficit hyperactivity disorder (ADHD), fetal alcohol spectrum disorder (FASD), and Parkinson's disease (PD, Tseng et al., 2012) as well as a lack of gaze towards other humans in children with autism spectrum disorder (Murias et al., 2017). Furthermore, recent studies found higher correlations of social-information seeking viewing behavior in monozygotic twins than in dizygotic twins (Constantino et al., 2017; Kennedy et al., 2017), substantiating the idea that differences in gaze preferences between individuals are stable. Nonetheless, stability in gaze preferences in the healthy population remains poorly understood. It is unclear how well gaze preferences generalize from one modality to another (e.g. from looking at images or videos to situations in which one is actively engaged in a conversation with others), how stable gaze preferences are over longer timescales and how closely gaze preferences are associated with other characteristics of an individual.

Explaining Stability in Gaze Preferences

Participants in the three studies incorporated in this work exhibited moderate to high levels of intrapersonal stability with regards to their viewing preferences, but no evidence was presented that these viewing preferences are linked to other psychological variables. Social anxiety played a role in gaze behavior in the real-life condition of study 2, but most of the variance in gaze preferences remained unexplained in this condition as well as in the lboratory condition as well as in studies 1 and 3. From a theoretical perspective, it is possible that gaze preferences may be tightly related to other psychological properties which were not covered. On the other hand, it is also conceivable that gaze preferences represent a mostly epiphenomenal phenomenon — comparable to fingerprints or the structure of the retina — with little significance for other psychological properties. In order to outline how gaze preferences may be characterized more thoroughly in future research, one may first turn to common psychological concepts used to characterize behavioral, cognitive or affective stability in other areas.

Psychological Traits

Humans differ on an abundance of psychological variables and no framework exists which can structure all of these characteristics into one comprehensive list (McAdams & Pals, 2006). Nonetheless, attempts were made to identify *traits* in individuals which describe interindividual differences in behavior, thoughts and feelings across a range of situations and points in time. The most widespread approach to characterizing personality traits to date is the *Big Five* personality model (Goldberg, 1993), which resulted from factor-analytically concentrating descriptive adjectives into five categories now labelled as *extraversion*, *neuroticism*, *conscientiousness*, *agreeableness*, and *openness*. However, the usefulness and meaning of traits and the approach in general were repeatedly questioned and remain controversial.

In an influential critique, Mischel (1968) objected that what researchers had identified as traits describing people could be better attributed to a consistency in the situations with which they were confronted during psychological assessments. He highlighted that the proposed traits were poor predictors for behavior in other situations — an accusation which is still sometimes held against personality psychology today. The criticism can now be partly refuted after meta-analytic analyses found personality traits to be similarily predictive for important life outcomes (e.g. mortality, divorce, and occupational attainment) as socioeconomic status and cognitive ability (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Nonetheless, personality trait models are still sometimes accused for downplaying the extent to which behavior depends on an interaction between a person and the situation he or she is confronted with (Mischel & Shoda, 1995). At the same time, it must be noted that proponents of trait models do not necessarily claim a general *cross-situational* consistency, but rather a *temporal* consistency in the reaction to similar situations (Roberts, 2009). And indeed, a variety of personality traits was shown to largely stabilize around the age of 30 (McCrae & Costa, 1994), although newer accounts found an increase in agreeableness and conscientiousness along the course of adulthood (Roberts & Mroczek, 2008).

In studies 1 and 3, participants were characterized along the *Big Five* personality traits as well as other trait concepts using self-report measures, allowing us to exploratively search for any cohesion between traits and gaze preferences. Participants in study 1 filled out the NEO-FFI (Costa & McCrae, 1989; Körner et al., 2002) questionnaire to characterize participants along the dimensions *extraversion* ($M = 29.47$, $SD = 5.81$), *neuroticism* ($M = 19.03$, $SD = 6.80$), *conscientiousness* ($M = 31.91$, $SD = 7.86$), *agreeableness* ($M = 32.28$, $SD = 6.29$), and *openness* (M

= 32.41, *SD* = 6.79; all values are raw values and all subscales range from 0 to 48) and were furthermore tested for their general anxiety [STAI; Spielberger (1983); *M* = 34.63, *SD* = 6.91; scale from 20 to 80], anxiety when interacting in social situations [SIAS; Mattick & Clarke (1998); *M* = 15.47, *SD* = 8.08; scale from 0 to 80] and autistic traits [AQ-K; Freitag et al. (2007); *M* = 7.25, *SD* = 3.30, scale from 0 to 33].

We used the same approach described in the main analysis (section *Directly predicting gaze using GLMMs*) to estimate the gaze preference for faces, physical saliency or the center of the screen in each individual participant. We computed generalized linear models for the entirety of all social videos, but separately for each participant, again each time describing the influence of centrality, low-level saliency and social information on gaze allocation, and z-standardized resulting coefficients along the participant group. We then computed Neyman-Pearson correlations of each of these three predictor weights with each of the individual self-report measures in an exploratory manner with no correction for multiple testing. The results are depicted in Figure 5.1. There was a significant positive association between the tendency to allocate attention to the center of the screen and the personality trait *agreeableness* ($r(30) = .35$, $p = .047$). However, note that with 24 individual tests, the expectancy value for statistically significant results assuming the null hypothesis (i.e., no true correlations between any of the measures) is 1.2, hence the individual significant result is easily explained by chance alone. To sum up, it seems fair to say that the exploratory search for correlates of gaze behavior at the level of self-report measures was not successful within the current set of individuals, stimuli and selected measures of self-report.

Participants in study 3 filled out a short version of the Big Five Inventory, BFI-K (Rammstedt & John, 2005)), to characterize participants along the dimensions *extraversion* (*M* = 14.55, *SD* = 3.42; scale from 4 to 20), *neuroticism* (*M* = 11.45, *SD* = 4.28; scale from 4 to 20), *conscientiousness* (*M* = 15.15, *SD* = 2.43; scale from 4 to 20), *agreeableness* (*M* = 12.26, *SD* = 3.86; scale from 4 to 20), and

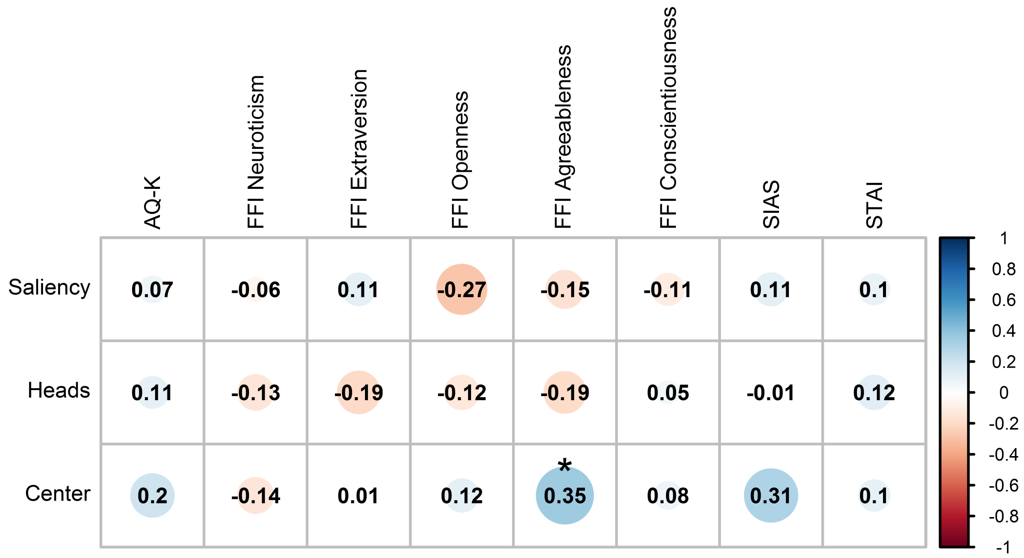

Figure 5.1: Correlations between participant's gaze preferences towards areas of high (physical) saliency, (human) heads and the screen's center with autistic traits (AQ-K) the Big Five personality traits measured via the NEO-FFI, general anxiety (STAI) and social anxiety (SIAS) in Study 1. The asterisk indicates a statistically significant correlation.

openness ($M = 19.26$, $SD = 4.65$; scale from 5 to 25) and were furthermore tested for their anxiety when interacting in social situations [SIAS; Mattick & Clarke (1998); $M = 22.08$, $SD = 12.28$; scale ranges from 0 to 80] and autistic traits [AQ-K; Freitag et al. (2007); $M = 3.11$, $SD = 0.33$; scale from 0 to 33]. Here, trait measures were correlated with relative gaze dwell times towards the head and objects throughout the entire trials, and furthermore with relative gaze dwell times while the virtual agent was near and either smiling at the participant or not (see Figure 5.2). None of the correlations reached the level of significance.

In conclusion, while the Big Five personality traits do explain some of the variance in individual life trajectories as documented in previous studies, they did not explain the stable gaze preferences found in Study 1 and Study 3 presented here. In Study 2, social anxiety was associated with a relative avoidance of gaze at near persons only in the real-life group. Although a lack of power may play a role in this absence of explained variance, it also seems plausible that gaze preferences exist at a different level of abstraction compared to personality traits. In this case, the psychological mechanisms driving gaze preferences could still be

Figure 5.2: Correlations between participant's gaze preferences towards the virtual agent's head and object in her hands (separately for gaze data from the entire trials as well as phases when the participant was near and smiling or not smiling at the participant) with autistic traits (AQ-K) the Big Five personality traits measured via the BFI-K and social anxiety (SIAS) in Study 3.

meaningful for other psychological variables, but would operate below the radar of most (self-report) personality assessment.

Other Levels of Psychological Stability

Gaze preferences may exhibit little overlap with psychological traits, but could possibly be driven by a person's *characteristic adapations* to current life circumstances. *Characteristic adapation* is an umbrella term for a variety of constructs such as values, goals, the self-image, world views or developmental tasks — all of which are usually understood in the context of a certain social environment and in reference to an individual's life circumstances. Characteristic adaptations are situated on a level of abstraction between general traits and specific behavior, and, unlike traits, are not organized into a single comprehensive structure (McAdams & Pals, 2006). Although such a lack of internal coherence can be seen as unsatisfactory, clinical practitioners predominantly choose this level of analysis when characterizing and treating their patients, highlighting the power of this approach in examining more contextualized fields of inquiry. At the same time, with few hypotheses guiding research, linking gaze preferences to one of the many aspects of characteristic adapations might require a more exploratory, or hit or miss method.

A rather crude and manageable class of constructs known to guide human behavior are those discussed in evolutionary psychology, including physical attractiveness, body size or social dominance. For instance, Maner, DeWall, & Gailliot (2008) found that when viewing images of a set of persons, participants from both sexes preferentially gazed at males who were displaying signs of social dominance, but not at dominant females. At the same time, participants from both sexes preferentially gazed at physically attractive females, while physical attractiveness did not modulate gaze at men. Another study found men to preferentially gaze at reproductively relevant regions of the female body (Garza, Heredia, & Cieslicka, 2016). These results were interpreted to reflect differences in mating strategies between the sexes resulting from divergent biological affordances for successful reproduction (Feingold, 1992; Trivers & others, 1972), although cultural influences need to be considered as well (Eagly & Wood, 1999). Again, these studies investigated gaze behavior as a function of the stimuli, but did not validate its conclusions by investigating systematic differences in gaze behavior as a function of a viewer's characteristics. For instance, following the authors' conclusions that mating strategies can play important role in gaze allocation towards other humans, one should expect female participants' own physical attractiveness, which is known to modulate mating strategies (Perilloux, Cloud, & Buss, 2013), to influence gaze behavior.

The relative lack of such crude biological markers as moderating variables in the analysis of gaze behavior stands at contrast with their well-documented impact on individuals' life trajectories. As one example, body height was found to predict important life outcomes such as economic success both in men and women in such diverse cultures as Germany (Hübler, 2009) and Indonesia (Sohn, 2015). Stulp et al. (2013b) analyzed body height in presidents of the United States of America (all were men) up to the year 2008 and found them to be about one standard deviation taller then the average of each president's birth cohort. Moreover, elected presidents were slightly but significantly taller compared to the losing presidential candidates. Stulp et al. (2013a) furthermore found women to more strongly emphasize body height in their partner selection compared to men. Such findings hint towards the existence of social mechanisms which may be more blunt or even "neanderthal" compared to the more sophisticated processes sometimes suggested by personality psychologists working with trait concepts. Although often relatively robust, concepts from evolutionary psychology may be somewhat sidelined in the study of personality development. Some authors have even (controversially) asserted that evolutionary ideas are downplayed in several areas of research due to a seeming misfit with researchers' political ideology (Von Hippel & Buss, 2017), which is quite dominantly the political left among psychologists (Inbar & Lammers, 2012; Redding, 2001).

Overall, the present work can admittedly help little to identify what variables are linked to a stronger tendency to seek social visual information, but argues for a broad search both at a high level of abstraction represented in psychological traits, as well as more concrete forms of characteristic adaptations and more biologically driven types of behavior including mating strategies.

Comparing Methodologies in Social Attention Research

Study 2 in the present work found an influence of social anxiety on gaze behavior in a real-life situation, but not in a matched laboratory situation. Together with other studies, this finding calls for a more thorough investigation of how the testing modality itself may influence research outcomes. Throughout the past decades, social attention was mostly studied by showing participants images or videos on a computer screen. While this line of research will likely continue to help identifying mechanisms of social attention, several novel findings may be expected from studies investigating social gaze in real situations or in virtual reality.

Methodological problems with Laboratory-based Research

The methodological problems with studying social gaze using images or videos have been described several times (Gobel et al., 2015; Risko et al., 2016; Schilbach et al., 2013). Criticism mostly focusses on the missing reactivity of the observed social stimuli (that is, humans and their faces), thereby ignoring the basic fact that human gaze serves two different purposes in social situations — collecting visual information and signalling mental states to others.

Another consideration, which is less frequently expressed but may additionaly dissuade participants from following their natural viewing behavior even in social scenes is the lack of three-dimensional, truly binocular visual input when viewing a computer screen. As described above (in section “The ventral stream and the dorsal stream of the visual system”), humans possess two largely separate visual systems, one for perceiving objects (emphasizing *what* can be seen) and one for acting upon them (emphasizing *where* something is located). The independence of the two systems is most impressively demonstrated in some neurological conditions (Himmelbach et al., 2012; Perenin & Vighetto, 1988), but was also found to modulate attentional processes in the general population (Dyde & Milner, 2002). While it seems reasonable to assume that perceptual information processing is rather unadulterated when viewing a computer screen compared to physical objects, this can hardly be claimed for information processing in action-oriented neural pathways which more strongly rely on binocular input (Milner, 2017). An interesting exemplification for the role of three-dimensional input in processing the environment again comes from a neurological patient with a damaged ventral stream who shows an impaired ability to recognize and describe objects but performs well when asked to judge an object’s distance (Read, Phillipson, Serrano-Pedraza, Milner, & Parker, 2010). Objects displayed on pictures or computer screens cannot be located in space in a naturalistic way, can usually not be processed as three dimensional entities using binocular vision, and participants have no expectation of being able to touch and manipulate them. The existence of the visual system's action-oriented dorsal stream therefore provides an additional challenge for the view that classic laboratory-based tasks carried out using compuer screens will adequately elicit real-life attentional phenomena. Critically, direct comparisons of viewing behavior in three-dimensional scenarios with viewing behavior towards content displayed on computer screens are still scarce.

Studying social attention in participants viewing computer screens is sometimes described as a *reductionist* approach (e.g. Schilbach et al., 2013). However, concerns may be raised if this characterization is accurate. Reductionism refers to describing relations along underlying, more basic structures, but is nowadays sometimes used pejoratively in order to denounce a position as naive (Andersen, 2001). This development may appear somewhat inept when one considers the multitude of scientific fields where a reductionist approach provoked groundbreaking advances, especially in the natural sciences and medicine. Take, as just one example, the periodic table in chemistry. This concept does not explain all observations in the field of chemistry in their entirety, but it does structure the plethora of substances into a few dozens of elements, indicates how these can form compounds and even predicted the existence of substances and elements before they were first observed. By contrast, laboratory-based research in social attention is not only characterized by a search for basic building blocks of the phenomena in question, but also by a restraint in where researchers look for observations of these phenomena. One may argue that a metaphor for this strategy is not so much the search for the periodic table, but a (hypothetical) version of chemistry where all experiments are performed at room temperature and under normal atmosphere pressure, rather than attempting to explore the behavior of substances in a wide range of circumstances. In my view, a thorough justification of why a laboratory-based research strategy in the field of social attention should be labelled *reductionist*, and not, more simply, *curtailed*, is still pending.

Experimentally manipulating Social Gaze using Virtual Body Illusions

Researchers can evade problems with laboratory-based research by collecting data in real-world situations — investigating not a surrogate, but a true example of the phenomenon of interest —, but may then see experimental standardization dwindle away amidst the stream of unscripted events that naturally occur in social situations. It was argued that a balance between ecological validity and experimental standardization can be achieved by using virtual reality (Pan & Hamilton, 2018), which was the chosen methodology in study 3 in this work. Observing scenes in virtual reality entails several alterations compared to watching similar scenes on a computer screen. In virtual reality, the scene is displayed at a much larger field of view with nothing but the virtual scene visible to the participant, and the visual perspective can be changed by rotating or laterally moving one's head. As a result, viewers typically locate themselves *in* the virtual scenario (i.e. feel *present* in the scene), which is often described as a markedly different experience compared to watching something *on* a computer screen.

An important aspect of feeling present in a social situation is the experience of being a person with a body (*corporeal awareness*), constituting a part of the social scene itself (Monti & Aglioti, 2018). Interestingly, virtual reality technology can be used to create the illusiory perception of owning a virtual body by tracking the physical body's movement and mapping it into a computer-generated mannequin which is seen from a first-person perspective (Blanke & Metzinger, 2009). Allowing participants to take ownership over a virtual body which differs from their physical body (e.g. in that it is taller or shorter, more or less phyiscally attractive) might therefore allow to experimentally manipulate social schema and social (gaze) behavior. The idea to actively manipulate variables that modulate social behavior is not new in psychological research. A prominent line of research following this strategy is the field of *social priming*, where subtle or even subliminally presented cues are used to activate social representations and to subsequently affect thoughts and behaviors. The literature documents measurable effects from quite fleeting social stimuli (Bargh, Schwader, Hailey, Dyer, & Boothby, 2012). However, several of these findings failed to replicate in a number of attempts (like a reported increase on the performance on cognitive tasks following the exposure to achievement-related words as in Harris, Coburn, Rohrer, & Pashler, 2013) or could only be replicated under unexpected boundary conditions (i.e., participants only walked slower after reading words associated with "being old" if the experimenter believed in such an effect; Doyen, Klein, Pichon, & Cleeremans, 2012), spawning criticism in the field's credibility as a whole (Molden, 2014).

Providing ownership over varying virtual bodies which embody varying social roles may be a powerful technique to activate specific aspects of the self (Maister, Slater, Sanchez-Vives, & Tsakiris, 2015). In a seminal study, Yee & Bailenson (2007) found participants to reveal more intimate details about themselves when identifying with a more attractive virtual avatar, and showed more confidence in a negotiation task when identifying with a taller compared to a shorter avatar. In studies by Banakou, Groten, & Slater (2013) as well as Tajadura-Jiménez, Banakou, Bianchi-Berthouze, & Slater (2017), participants took ownership over a 4-year-old child, experiencing a scene with a smaller body and from a lower perspective. In both studies, embodiment resulted in a shift towards associating the self with child-like compared to adult-like categorizations in an *Implicit Association Test* (IAT), but only if the avatar's proportion resembled a child and was not merely an adult body shrinked in size. Peck, Seinfeld, Aglioti, & Slater (2013) found embodiment of a dark-skinned avatar to reduce racial bias as measured in an IAT in light-skinned participants. However, to the best of my knowledge, no study has yet investigated the effects of embodying other persons in more complex social situations and included the measurement of gaze behavior.

Compared to the incidental or even subliminal presentation of stimili, a strong and consciously experienced manipulation like a virtual body illusion bears an increased risk of eliciting expectancy effects. It therefore seems critical for researchers in the field to include unobtrusive and difficult-to-fake measurements in order to reach well-founded conclusions. While explicit questions can usually not be seen as impervious to deliberate manipulation or expectancy effects, implicit measures like the IAT are only prone to deliberate alterations when participants' attention is focussed to the underlying parameters of the task (Hughes et al., 2016). Likewise, behavioral measures such as movement (Rubo & Gamer, 2019) or gaze behavior may possibly reveal deeper insights into the suggestibility of social attentive processes in the future.

Conclusion and Outlook

The present work presented three studies investigating phenomena within the field of social attention. Collectively, they contributed to weighing up the relevance of two contending classes of gaze predictors (physical saliency and human heads) when viewing videos (Study 1), highlighted alterations of gaze behavior in social anxiety and fueled the dispute over suitable research methods in the field (Study 2), provided further basis that virtual reality is a useful tool when studying social attention (Study 3) and introduced compelling evidence that preferences in gaze behavior are relatively stable along an experiment's timecourse (Studies 1-3).

Nonetheless, an overwhelming amount of complex research questions continues to haunt researchers in the field and was, if at all, only briefly touched upon in the presented studies. The latter part of this work attempts to sketch how future research might tackle the sometimes cumbersome complexity of social attentive phenomena. It is argued that future research should include variables of the viewer into models describing gaze allocation (possibly explaining the intrapersonal stability in gaze preferences observed here), more precisely classify different types of social situations, and use novel techniques such as virtual body illusions to actively manipulate implicit social concepts and behavior.

A first series of studies following up to the present work should investigate the stability of gaze preferences along a longer timecourse and across more diverse situations. Note that while the three studies presented here found gaze preferences to be stable within each experiment, the testing sessions were relatively short (less than one hour) and the testing condition was not varied within participants. It therefore remains unclear to what extent gaze preferences vary within an individual during the course of a day or across days, and if preferences for gaze at other human beings when watching videos is associated with a corresponding preference in real-life situations.

Secondly, future research should more broadly search for a cohesion between gaze preferences and other psychologically relevant variables. With few tangible hypotheses guiding the search, it seems appropriate to (exploratively) integrate a variety of measures which were found to be reliable and predictive in other psychological areas and to successively move to a more confirmatory research strategy incorprating fewer variables after promising candidate variables were identified. Such measures may include more detailed assessments of psychological traits, a thorough analysis of individuals' life circumstances as well as their corresponding characteristic adaptations, but also more crude biological variables such as body height or physical attractiveness.

Thirdly, if gaze preferences were to be influenced by tangible psychological variables of any kind, research should, where feasible, attempt to experimentally manipulate these variables in order to establish clear causal links. For instance, if body height or physical attractiveness were associated with a more brisk social behavior and more uninhibited gaze at others in real-life situations, one might further test this relationship by allowing participants to embody virtual characters varying on these variables and to communicate with each other within a virtual world.

Finally, if such experimental manipulations were to reliably modulate social attention and behavior, additional research should inquire the stability of such effects beyond the experimental situation and test if experiences in virtual worlds might even be beneficial for therapeutic purposes. For instance, experiencing oneself as more socially confident in a modified (virtual) environment might possibly pave a way for establishing novel behavioral or modulating social schemas in socially anxious patients. To sum up, while all attempts to sketch a pathway for future research must be seen as speculative in the current state of research, the present work hopes to contribute to the still emerging field of social attention research and foster fruitful continuative investigations.

REFERENCES

Adamovich, S. V., Fluet, G. G., Tunik, E., & Merians, A. S. (2009). Sensorimotor training in virtual reality: A review. *NeuroRehabilitation*, *25*(1), 29–44.

Adolphs, R. (2010a). Conceptual challenges and directions for social neuroscience. *Neuron*, *65*(6), 752–767.

Adolphs, R. (2010b). 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

Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, *5*(6), 679–685.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (dsm-5)*. American Psychiatric Publishing.

Amir, N., Foa, E. B., & Coles, M. E. (1998). Automatic activation and strategic avoidance of threat-relevant information in social phobia. *Journal of Abnormal Psychology*, *107*(2), 285–290. https://doi.org/10.1037/0021-843x.107.2.285

Andersen, H. (2001). The history of reductionism versus holistic approaches to scientific research. *Endeavour*, *25*(4), 153–156. https://doi.org/10.1016/s0160-9327(00)01387-9

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

Argyle, M., & Cook, M. (Eds.). (1976). *Gaze and mutual gaze*. Cambridge U Press.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Baker, S. R., & Edelmann, R. J. (2002). Is social phobia related to lack of social skills? Duration of skill-related behaviours and ratings of behavioural adequacy. *British Journal of Clinical Psychology*, *41*(3), 243–257. https://doi.org/10.1348/014466502760379118

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences*, *34*(4), 210–224. https://doi.org/10.1016/j.tins.2011.02.003

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, *110*(31), 12846–12851. https://doi.org/10.1073/pnas.1306779110

Bandura, A. (1978). On paradigms and recycled ideologies. *Cognitive Therapy and Research*, *2*(1), 79–103. https://doi.org/10.1007/bf01172518

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences*, *16*(12), 593–605. https://doi.org/10.1016/j.tics.2012.10.002

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & IJzendoorn, M. H. van. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The autism-spectrum quotient (AQ)Adolescent version. *Journal of Autism and Developmental Disorders*, *36*(3), 343–350. https://doi.org/10.1007/s10803-006-0073-6

Barton, J. J., Press, D. Z., Keenan, J. P., & Connor, M. O. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, *58*(1), 71–78. https://doi.org/10.1212/wnl.58.1.71

Bates, D., Maechler, M., Bolker, B., Walker, S., & others. (2014). Lme4: Linear mixed-effects models using eigen and s4. *R Package Version*, *1*(7), 1–23.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, *23*(2), 223–229. https://doi.org/10.1016/s0163-6383(01)00037-6

Baumgartner, T. (2008). Feeling present in arousing virtual reality worlds: Prefrontal brain regions differentially orchestrate presence experience in adults and children. *Frontiers in Human Neuroscience*, *2*. https://doi.org/10.3389/neuro.09.008.2008

Beck, A. T. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561. https://doi.org/10.1001/archpsyc.1961.01710120031004

Beller, S., Bender, A., & Medin, D. L. (2012). Should anthropology be part of cognitive science? *Topics in Cognitive Science*, *4*(3), 342–353. https://doi.org/10.1111/j.1756-8765.2012.01196.x

Ben-Moussa, M., Rubo, M., Debracque, C., & Lange, W.-G. (2017). DJINNI: A novel technology supported exposure therapy paradigm for SAD combining virtual reality and augmented reality. *Frontiers in Psychiatry*, *8*. https://doi.org/10.3389/fpsyt.2017.00026

Benuzzi, F., Pugnaghi, M., Meletti, S., Lui, F., Serafini, M., Baraldi, P., & Nichelli, P. (2007). Processing the socially relevant parts of faces. *Brain Research Bulletin*, *74*(5), 344–356. https://doi.org/10.1016/j.brainresbull.2007.07.010

Bernat, E., Patrick, C. J., Benning, S. D., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, *43*(1), 93–103.

Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. *Communication in the Age of Virtual Reality*, *15*, 32.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). 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. (2009a). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, *17*(6-7), 904–924.

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.

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

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*(1), 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, *13*(8), 556.

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13. https://doi.org/10.1016/j.tics.2008.10.003

Block, N. (1981). Psychologism and behaviorism. *The Philosophical Review*, *90*(1), 5–43.

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

Borji, A. (2012). Boosting bottom-up and top-down visual features for saliency estimation. In *Computer vision and pattern recognition (cvpr), 2012 ieee conference on* (pp. 438–445). IEEE.

Borji, A., Parks, D., & Itti, L. (2014). Complementary effects of gaze direction and early saliency in guiding fixations during free viewing. *Journal of Vision*, *14*(13), 3–3. https://doi.org/10.1167/14.13.3

Bouchard, S., Dumoulin, S., Talbot, J., Ledoux, A.-A., Phillips, J., Monthuy-Blanc, J., … Renaud, P. (2012). Manipulating subjective realism and its impact on presence: Preliminary results on feasibility and neuroanatomical correlates. *Interacting with Computers*, *24*(4), 227–236. https://doi.org/10.1016/j.intcom.2012.04.011

Bögels, S. M., & Mansell, W. (2004). Attention processes in the maintenance and treatment of social phobia: Hypervigilance, avoidance and self-focused attention. *Clinical Psychology Review*, *24*(7), 827–856. https://doi.org/10.1016/j.cpr.2004.06.005

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation i: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*(7), 513–531. https://doi.org/10.1037/0003-066x.32.7.513

Brunet, P. M., Heisz, J. J., Mondloch, C. J., Shore, D. I., & Schmidt, L. A. (2009). Shyness and face scanning in children. *Journal of Anxiety Disorders*, *23*(7), 909–914. https://doi.org/10.1016/j.janxdis.2009.05.009

Bruno, N., Bernardis, P., & Gentilucci, M. (2008). Visually guided pointing, the müller-lyer illusion, and the functional interpretation of the dorsal-ventral split: Conclusions from 33 independent studies. *Neuroscience & Biobehavioral Reviews*, *32*(3), 423–437. https://doi.org/10.1016/j.neubiorev.2007.08.006

Bruno, N., & Franz, V. H. (2009). When is grasping affected by the müller-lyer illusion? *Neuropsychologia*, *47*(6), 1421–1433. https://doi.org/10.1016/j.neuropsychologia.2008.10.031

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, *50*(3), 255–272. https://doi.org/10.1037/h0060889

Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer-Verlag.

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development: An International Journal of Research and Practice*, *10*(1-2), 67–74.

Butterworth, G., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development*, *3*(3), 253–272. https://doi.org/10.1177/016502548000300303

Byrnes, J. E., & Stachowicz, J. J. (2009). The consequences of consumer diversity loss: Different answers from different experimental designs. *Ecology*, *90*(10), 2879–2888.

Cameron, A. C., & Windmeijer, F. A. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics*, *14*(2), 209–220.

Cardoso-Leite, P., & Gorea, A. (2010). On the perceptual/motor dissociation: A review of concepts, theory, experimental paradigms and data interpretations. *Seeing and Perceiving*, *23*(2), 89–151.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, *46*(26), 4333–4345.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i. https://doi.org/10.2307/1166214

Caspi, A., Elder, G. H., & Bem, D. J. (1988). Moving away from the world: Life-course patterns of shy children. *Developmental Psychology*, *24*(6), 824.

Castelhano, M. S., Wieth, M., & Henderson, J. M. (2007). I see what you see: Eye movements in real-world scenes are affected by perceived direction of gaze. In *International workshop on attention in cognitive systems* (pp. 251–262). Springer.

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.

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).

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284.

Cobb, S. V. G., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). *Presence: Teleoperators and Virtual Environments*, *8*(2), 169–186. https://doi.org/10.1162/105474699566152

Cohen, A. S., Sasaki, J. Y., German, T. C., & Kim, H. S. (2015). Automatic mechanisms for social attention are culturally penetrable. *Cognitive Science*, *41*(1), 242–258. https://doi.org/10.1111/cogs.12329

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

Constantino, J. N., Todorov, A., Hilton, C., Law, P., Zhang, Y., Molloy, E., … Geschwind, D. (2012). Autism recurrence in half siblings: Strong support for genetic mechanisms of transmission in ASD. *Molecular Psychiatry*, *18*(2), 137–138. https://doi.org/10.1038/mp.2012.9

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

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, *34*(1), 28–38. https://doi.org/10.1037/0012-1649.34.1.28

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: Eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98.

Costa, P. T., & McCrae, R. R. (1989). *NEO pi/ffi manual supplement for use with the neo personality inventory and the neo five-factor inventory*. Psychological Assessment Resources.

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 5–5.

Coutrot, A., & Guyader, N. (2017). Learning a time-dependent master saliency map from eye-tracking data in videos. *arXiv Preprint arXiv:1702.00714*.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116.

Damasio, A. R., & Benton, A. L. (1979). Impairment of hand movements under visual guidance. *Neurology*, *29*(2), 170–170.

Daum, S. O., & Hecht, H. (2009). Distance estimation in vista space. *Attention, Perception, & Psychophysics*, *71*(5), 1127–1137. https://doi.org/10.3758/app.71.5.1127

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

DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, *22*(5), 1915–1953.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.neuro.18.1.193

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*(6), 959–988. https://doi.org/10.1037/0033-2909.130.6.959

Dillon, C., Keogh, E., & Freeman, J. (2002). It's been emotional': Affect, physiology, and presence. *Proceedings of Presence*, 223–232.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28–28.

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming:

It's all in the mind, but whose mind? *PloS One*, *7*(1), e29081.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, *1*(1), 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Dyde, R. T., & Milner, A. D. (2002). Two illusions of perceived orientation: One fools all of the people some of the time; the other fools all of the people all of the time. *Experimental Brain Research*, *144*(4), 518–527.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*(6), 408–423. https://doi.org/10.1037/0003-066x.54.6.408

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. *Psychological Methods*, *12*(1), 1–22. https://doi.org/10.1037/1082-989X.12.1.1

Eimer, M. (1997). Uninformative symbolic cues may bias visual-spatial attention: Behavioral and electrophysiological evidence. *Biological Psychology*, *46*(1), 67–71. https://doi.org/10.1016/s0301-0511(97)05254-x

Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist*, *30*(10), 967–976. https://doi.org/10.1037/0003-066x.30.10.967

Emery, N. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604. https://doi.org/10.1016/s0149-7634(00)00025-7

Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., & Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (macaca mulatta). *Journal of Comparative Psychology*, *111*(3), 286–293. https://doi.org/10.1037/0735-7036.111.3.286

End, A., & Gamer, M. (2017a). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, *8*, 418.

End, A., & Gamer, M. (2017b). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00418

Essau, C. A., Conradt, J., & Petermann, F. (1999). Frequency and comorbidity of social phobia and social fears in adolescents1. *Behaviour Research and Therapy*, *37*(9), 831–843.

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., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy*, *5*(1), 39–60. https://doi.org/10.1207/s15327078in0501_2

Fecteau, J., & Munoz, D. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Fehm, L., Pelissolo, A., Furmark, T., & Wittchen, H.-U. (2005). Size and burden of social phobia in europe. *European Neuropsychopharmacology*, *15*(4), 453–462.

Feingold, A. (1992). Gender differences in mate selection preferences: A test of the parental investment model. *Psychological Bulletin*, *112*(1), 125–139. https://doi.org/10.1037/0033-2909.112.1.125

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.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, *117*(3), 319–331.

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

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170.

Freeth, M., & Bugembe, P. (2018). Social partner gaze direction and conversational phase - factors affecting social attention during face-to-face conversations in autistic adults? *Autism*, 136236131875678. https://doi.org/10.1177/1362361318756786

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., … Gontard, A. von. (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

French, L. R., Bertone, A., Hyde, K. L., & Fombonne, E. (2013). Epidemiology of autism spectrum disorders. In *The neuroscience of autism spectrum disorders* (pp. 3–24). Elsevier. https://doi.org/10.1016/b978-0-12-391924-3.00001-6

French, R. M. (2000). The turing test: The first 50 years. *Trends in Cognitive Sciences*, *4*(3), 115–122.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. https://doi.org/10.3758/bf03208827

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319–329. https://doi.org/10.1037/0096-1523.30.2.319

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

Fydrich, T. (2002). SPAI-soziale phobie und angst inventar. In *Diagnostische verfahren in der psychotherapie. Göttingen: Hogrefe* (pp. 335–338).

Gallup, A. C., Chong, A., & Couzin, I. D. (2012). The directional flow of visual information transfer between pedestrians. *Biology Letters*, *8*(4), 520–522. https://doi.org/10.1098/rsbl.2012.0160

Gallup, A. C., 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. (2010). The time-course of attention to emotional faces in social phobia. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(1), 39–44. https://doi.org/10.1016/j.jbtep.2009.08.008

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

Gamer, M., Hecht, H., Seipp, N., & Hiller, W. (2011). Who is looking at me? The cone of gaze widens in social phobia. *Cognition & Emotion*, *25*(4), 756–764. https://doi.org/10.1080/02699931.2010.503117

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). Irr: Various coefficients of interrater reliability and agreement. R package v. 0.84.

Gamer, M., Schmitz, A. K., Tittgemeyer, M., & Schilbach, L. (2013). The human amygdala drives reflexive orienting towards facial features. *Current Biology*, *23*(20), R917–R918. https://doi.org/10.1016/j.cub.2013.09.008

Gardner, H. (1987). *The mind's new science: A history of the cognitive revolution*. Basic books.

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

Garza, R., Heredia, R. R., & Cieslicka, A. B. (2016). Male and female perception of physical attractiveness. *Evolutionary Psychology*, *14*(1), 147470491663161. https://doi.org/10.1177/1474704916631614

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. https://doi.org/10.1038/72140

Geschwind, D. H. (2009). Advances in autism. *Annual Review of Medicine*, *60*(1), 367–380. https://doi.org/10.1146/annurev.med.60.053107.121225

Gibson, B. S., & Kingstone, A. (2006). Visual attention and the semantics of space. *Psychological Science*, *17*(7), 622–627. https://doi.org/10.1111/j.1467-9280.2006.01754.x

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "estimating the reproducibility of psychological science". *Science*, *351*(6277), 1037–1037. https://doi.org/10.1126/science.aad7243

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

Goffman, E. (Ed.). (1963). *Behavior in public places.* Simon; Schuster.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037/0003-066x.48.1.26

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Goodale, M. A., Milner, A. D., Jakobson, L., & Carey, D. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*(6305), 154.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*(6666), 481–484. https://doi.org/10.1038/35135

Grüsser, O.-J. (1983). Multimodal structure of the extrapersonal space. In *Spatially oriented behavior* (pp. 327–352). Springer New York. https://doi.org/10.1007/978-1-4612-5488-1_18

Guerin, B. (2010). Social facilitation. *The Corsini Encyclopedia of Psychology*, 1–2.

Gurman, A. S., & Fraenkel, P. (2002). The history of couple therapy: A millennial review. *Family Process*, *41*(2), 199–260. https://doi.org/10.1111/j.1545-5300.2002.41204.x

Haaf, R. A., & Bell, R. Q. (1967). A facial dimension in visual discrimination by human infants. *Child Development*, *38*(3), 893. https://doi.org/10.2307/1127266

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development*, *67*(5), 1940. https://doi.org/10.2307/1131602

Haith, M., Bergman, T., & Moore, M. (1977). Eye contact and face scanning in early infancy. *Science*, *198*(4319), 853–855. https://doi.org/10.1126/science.918670

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In *Advances in neural information processing systems* (pp. 545–552).

Harnad, S. (1992). The turing test is not a trick: Turing indistinguishability is a scientific criterion. *ACM SIGART Bulletin*, *3*(4), 9–10.

Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLoS ONE*, *8*(8), e72467. https://doi.org/10.1371/journal.pone.0072467

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, *34*(3), 479–490. https://doi.org/10.1016/s0896-6273(02)00662-1

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (Vol. 1). https://doi.org/10.1007/b94608

Hayduk, L. A. (1981). The shape of personal space: An experimental investigation. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *13*(1), 87–93. https://doi.org/10.1037/h0081114

Hayes, P., & Ford, K. (1995). Turing test considered harmful. In *IJCAI (1)* (pp. 972–977).

Hayes, T. R., & Henderson, J. M. (2018). Scan patterns during scene viewing predict individual differences in clinical traits in a normative sample. *PLOS ONE*, *13*(5), e0196654. https://doi.org/10.1371/journal.pone.0196654

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

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*(13), 1707–1724. https://doi.org/10.1016/j.visres.2004.12.021

Hershler, O., & Hochstein, S. (2006). With a careful look: Still no low-level confound to face pop-out. *Vision Research*, *46*(18), 3028–3035. https://doi.org/10.1016/j.visres.2006.03.023

Hessels, R. S., Gijs A. Holleman, Cornelissen, T. H., Hooge, I. T., & Kemner, C. (2017). Eye contact takes two autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*. https://doi.org/10.5127/jep.062917

Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition*, *184*, 28–43. https://doi.org/10.1016/j.cognition.2018.12.005

Hietanen, J., Nummenmaa, L., Nyman, M., Parkkola, R., & Hämäläinen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMRI study. *NeuroImage*, *33*(1), 406–413. https://doi.org/10.1016/j.neuroimage.2006.06.048

Himmelbach, M., Boehme, R., & Karnath, H.-O. (2012). 20 years later: A second look on df's motor behaviour. *Neuropsychologia*, *50*(1), 139–144.

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. https://doi.org/10.1016/s0887-6185(97)00040-6

Hughes, S., Hussey, I., Corrigan, B., Jolie, K., Murphy, C., & Barnes-Holmes, D. (2016). Faking revisited: Exerting strategic control over performance on the implicit relational assessment procedure. *European Journal of Social Psychology*, *46*(5), 632–648. https://doi.org/10.1002/ejsp.2207

Hull, C. L. (1943). The problem of intervening variables in molar behavior theory. *Psychological Review*, *50*(3), 273–291. https://doi.org/10.1037/h0057518

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*(2), 231–255. https://doi.org/10.1207/s15327078in0602_5

Huppert, J. D., Pasupuleti, R. V., Foa, E. B., & Mathews, A. (2007). Interpretation biases in social anxiety: Response generation, response selection, and self-appraisals. *Behaviour Research and Therapy*, *45*(7), 1505–1515. https://doi.org/10.1016/j.brat.2007.01.006

Hutchins, E. (1995a). *Cognition in the wild*. MIT press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1

Hübler, O. (2009). The nonlinear link between height and wages in germany, 19852004. *Economics & Human Biology*, *7*(2), 191–199. https://doi.org/10.1016/j.ehb.2009.06.003

Iacobucci, D., Saldanha, N., & Deng, X. (2007). A mediation on mediation: Evidence that structural equation models perform better than regression. *Journal of Consumer Psychological*, *7*(2), 140–154. https://doi.org/10.1016/S1057-7408(07)70020-7

Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science*, *7*(5), 496–503.

Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, *12*(supplement 2), 35–51. https://doi.org/10.1162/089892900564055

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*(6), 1093–1123.

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

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.

Jenkins, J. J. (1974). Remember that old theory of memory? Well, forget it. *American Psychologist*, *29*(11), 785–795. https://doi.org/10.1037/h0037399

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

Johnson, D. M. (2005). Introduction to and review of simulator sickness research. American Psychological Association (APA). https://doi.org/10.1037/e456932006-001

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry*, *65*(8), 946. https://doi.org/10.1001/archpsyc.65.8.946

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 26-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

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

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*(1), B1–B11. https://doi.org/10.1016/s0010-0277(98)00035-3

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, *104*(1), 105–114. https://doi.org/10.1016/j.jecp.2009.01.006

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., … Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*(6), F31–F36. https://doi.org/10.1111/j.1467-7687.2005.0434a.x

Kendler, K., Karkowski, L., & Prescott, C. (1999). Fears and phobias: Reliability and heritability. *Psychological Medicine*, *29*(3), 539–553.

Kennedy, D. P., D'Onofrio, B. M., Quinn, P. D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic influence on eye movements to complex scenes at short timescales. *Current Biology*, *27*(22), 3554–3560.e3. https://doi.org/10.1016/j.cub.2017.10.007

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. *Archives of*

General Psychiatry, *62*(6), 593–602.

Kilteni, K., Normand, J.-M., Sanchez-Vives, M. V., & Slater, M. (2012). Extending body space in immersive virtual reality: A very long arm illusion. *PLoS ONE*, *7*(7), e40867. https://doi.org/10.1371/journal.pone.0040867

Kim, Y. S., Leventhal, B. L., Koh, Y.-J., Fombonne, E., Laska, E., Lim, E.-C., … others. (2011). Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry*, *168*(9), 904–912.

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

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

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

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100. https://doi.org/10.1037/0033-2909.100.1.78

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

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767–768. https://doi.org/10.1038/42842

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, *40*(5), 419–435. https://doi.org/10.1006/jhev.2001.0468

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

Körner, A., Geyer, M., & Brähler, E. (2002). Das neo-fünf-faktoren inventar. *Diagnostica*, *48*(1), 19–27.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421.

Kruthiventi, S. S. S., Ayush, K., & Babu, R. V. (2017). DeepFix: A fully convolutional neural network for predicting human eye fixations. *IEEE Transactions on Image Processing*, *26*(9), 4446–4456. https://doi.org/10.1109/tip.2017.2710620

Kuhn, G., Teszka, R., Tenaw, N., & Kingstone, A. (2016). Don't be fooled! Attentional responses to social cues in a face-to-face and video magic trick reveals greater top-down control for overt than covert attention. *Cognition*, *146*, 136–142. https://doi.org/10.1016/j.cognition.2015.08.005

Kulik, A. (2018). Virtually the ultimate research lab. *British Journal of Psychology*, *109*(3), 434–436. https://doi.org/10.1111/bjop.12307

Kummerer, M., Wallis, T. S., Gatys, L. A., & Bethge, M. (2017). Understanding low- and high-level contributions to fixation prediction. In *2017 IEEE international conference on computer vision (ICCV)*. IEEE. https://doi.org/10.1109/iccv.2017.513

Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2015). Information-theoretic model comparison unifies saliency metrics. *Proceedings of the National Academy of Sciences*, *112*(52), 16054–16059. https://doi.org/10.1073/pnas.1510393112

Laidlaw, K. E., Rothwell, A., & Kingstone, A. (2016). Camouflaged attention: Covert attention is critical to social communication in natural settings. *Evolution and Human Behavior*, *37*(6), 449–455. https://doi.org/10.1016/j.evolhumbehav.2016.04.004

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*, *108*(14), 5548–5553. https://doi.org/10.1073/pnas.1017022108

Land, M. F., & Fernald, R. D. (1992). The evolution of eyes. *Annual Review of Neuroscience*, *15*(1), 1–29.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

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*, 39–58.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273.

Lange, W.-G., Rinck, M., & Becker, E. S. (2014). Behavioral deviations: Surface features of social anxiety and what they reveal. *The Wiley Blackwell Handbook of Social Anxiety Disorder*, 344–365.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology*, *23*(4-5), 727–738.

Lee, Y.-J., Greene, H. H., Tsai, C. W., & Chou, Y. J. (2016). Differences in sequential eye movement behavior between taiwanese and american viewers. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00697

Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research*, *47*(19), 2483–2498.

Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., … Park, S. (2000). Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Third international workshop on presence, delft, the netherlands* (Vol. 240, pp. 2–4).

Lundh, L.-G., & Öst, L.-G. (1996). Recognition bias for critical faces in social phobics. *Behaviour Research and Therapy*, *34*(10), 787–794. https://doi.org/10.1016/0005-7967(96)00035-6

Macdonald, R. G., & Tatler, B. W. (2018). Gaze in a real-world social interaction: A dual eye-tracking study. *Quarterly Journal of Experimental Psychology*, *71*(10), 2162–2173. https://doi.org/10.1177/1747021817739221

Madary, M., & Metzinger, T. K. (2016). Real virtuality: A code of ethical conduct. Recommendations for good scientific practice and the consumers of vr-technology. *Frontiers in Robotics and AI*, *3*, 3.

Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: Owning another body affects social cognition. *Trends in Cognitive Sciences*, *19*(1), 6–12. https://doi.org/10.1016/j.tics.2014.11.001

Maner, J. K., DeWall, C. N., & Gailliot, M. T. (2008). Selective attention to signs of success: Social dominance and early stage interpersonal perception. *Personality and Social Psychology Bulletin*, *34*(4), 488–501. https://doi.org/10.1177/0146167207311910

Mattick, R., & Clarke, 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

Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in v4 during free viewing visual search. *Neuron*, *40*(6), 1241–1250. https://doi.org/10.1016/s0896-6273(03)00764-5

McAdams, D. P., & Pals, J. L. (2006). A new big five: Fundamental principles for an integrative science of personality. *American Psychologist*, *61*(3), 204–217. https://doi.org/10.1037/0003-066x.61.3.204

McCrae, R. R., & Costa, P. T. (1994). The stability of personality: Observations and evaluations. *Current Directions in Psychological Science*, *3*(6), 173–175. https://doi.org/10.1111/1467-8721.ep10770693

Meehan, M., Insko, B., Whitton, M., & Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. In *Proceedings of the 29th annual conference on computer graphics and interactive techniques - SIGGRAPH 02*. ACM Press. https://doi.org/10.1145/566570.566630

Mellings, T. M., & Alden, L. E. (2000). Cognitive processes in social anxiety: The effects of self-focus, rumination and anticipatory processing. *Behaviour Research and Therapy*, *38*(3), 243–257. https://doi.org/10.1016/s0005-7967(99)00040-6

Milner, A. D. (2017). How do the two visual streams interact with each other? *Experimental Brain Research*, *235*(5), 1297–1308. https://doi.org/10.1007/s00221-017-4917-4

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley. https://doi.

org/10.4324/9780203763643

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246–268. https://doi.org/10.1037/0033-295x.102.2.246

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.

Mo, C., He, D., & Fang, F. (2017). Attention priority map of face images in human early visual cortex. *The Journal of Neuroscience*, 1206–1217. https://doi.org/10.1523/jneurosci.1206-17.2017

Mogg, K., Philippot, P., & Bradley, B. P. (2004). Selective attention to angry faces in clinical social phobia. *Journal of Abnormal Psychology*, *113*(1), 160–165. https://doi.org/10.1037/0021-843x.113.1.160

Molden, D. C. (2014). Understanding priming effects in social psychology: What is “Social priming” and how does it occur? *Social Cognition*, *32*(Supplement), 1–11. https://doi.org/10.1521/soco.2014.32.supp.1

Monti, A., & Aglioti, S. M. (2018). Flesh and bone digital sociality: On how humans may go virtual. *British Journal of Psychology*, *109*(3), 418–420. https://doi.org/10.1111/bjop.12300

Mood, C. (2010). Logistic regression: Why we cannot do what We think we can do, and what we can do about it. *European Sociological Review*, *26*(1), 67–82. https://doi.org/10.1093/esr/jcp006

Moor, J. H. (2001). The status and future of the turing test. *Minds and Machines*, *11*(1), 77–93.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, *98*(2), 164–181. https://doi.org/10.1037/0033-295x.98.2.164

Moscovitch, D. A., & Hofmann, S. G. (2007). When ambiguity hurts: Social standards moderate self-appraisals in generalized social phobia. *Behaviour Research and Therapy*, *45*(5), 1039–1052.

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

Muir, D. W., Humphrey, D. E., & Humphrey, G. K. (1994). Pattern and space perception in young infants. *Spatial Vision*, *8*(1), 141–165. https://doi.org/10.1163/156856894x00288

Murias, M., Major, S., Davlantis, K., Franz, L., Harris, A., Rardin, B., … Dawson, G. (2017). Validation of eye-tracking measures of social attention as a potential biomarker for autism clinical trials. *Autism Research*, *11*(1), 166–174. https://doi.org/10.1002/aur.1894

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., ... 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. (2011). Stimulus-driven orienting of visuo-spatial attention in complex dynamic environments. *Neuron*, *69*(5), 1015–1028. https://doi.org/10.1016/j.neuron.2011.02.020

Nelson, E. C., Grant, J. D., Bucholz, K. K., Glowinski, A., Madden, P., Reich, W., & Heath, A. C. (2000). Social phobia in a population-based female adolescent twin sample: Co-morbidity and associated suicide-related symptoms. *Psychological Medicine*, *30*(4), 797–804.

Neufeld, K. J., Swartz, K., Bienvenu, O., Eaton, W., & Cai, G. (1999). Incidence of dis/dsm-iv social phobia in adults. *Acta Psychiatrica Scandinavica*, *100*(3), 186–192.

Norman, D. A. (1980). Twelve issues for cognitive science. *Cognitive Science*, *4*(1), 1–32. https://doi.org/10.1207/s15516709cog0401_1

Norton, E. C. (2012). *Log Odds and Ends.* National Bureau of Economic Research. https://doi.org/10.1002/0471725331.ch16

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.

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How well can saliency models predict fixation selection in scenes beyond central bias? A new approach to model evaluation using generalized linear mixed models. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00491

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).

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi, Ş., & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: A quantitative meta-analysis. *Depression and Anxiety*, *29*(2), 85–93.

Ortiz, C. (2016). Why we need a physically embodied turing test and what it might look like. *AI Magazine*, *37*(1), 55–62.

Otto, M. W., Pollack, M. H., Maki, K. M., Gould, R. A., Worthington, J. J., Smoller, J. W., & Rosenbaum, J. F. (2001). Childhood history of anxiety disorders among adults with social phobia: Rates, correlates, and comparisons with patients with panic disorder. *Depression and Anxiety*, *14*(4), 209–213.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detect-ing the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466.

Pan, X., Gillies, M., Barker, C., Clark, D. M., & Slater, M. (2012). Socially anxious and confident men interact with a forward virtual woman: An experimental study. *PloS One*, *7*(4), e32931.

Pan, X., & Hamilton, A. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

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.

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.

Parsons, T. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00660

Patterson, M. L. (1976). An arousal model of interpersonal intimacy. *Psychological Review*, *83*(3), 235–245. https://doi.org/10.1037/0033-295x.83.3.235

Patterson, M. L. (1982). A sequential functional model of nonverbal exchange. *Psychological Review*, *89*(3), 231–249. https://doi.org/10.1037/0033-295x.89.3.231

Patton, J., Dawe, G., Scharver, C., Mussa-Ivaldi, F., & Kenyon, R. (2006). Robotics and virtual reality: A perfect marriage for motor control research and rehabilitation. *Assistive Technology*, *18*(2), 181–195.

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, *22*(3), 779–787. https://doi.org/10.1016/j.concog.2013.04.016

Pereira, E. J., Birmingham, E., & Ristic, J. (2019). The eyes do not have it after all? Attention is not automatically biased towards faces and eyes. *Psychological Research*. https://doi.org/10.1007/s00426-018-1130-4

Perenin, M.-T., & Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms: I. Different aspects of the deficit in reaching for objects. *Brain*, *111*(3), 643–674.

Perilloux, C., Cloud, J. M., & Buss, D. M. (2013). Women's physical attractiveness and short-term mating strategies. *Personality and Individual Differences*, *54*(4), 490–495. https://doi.org/10.1016/j.paid.2012.10.028

Pessoa, L. (2010). Emotion and cognition and the amygdala: From "what is it?" To "what's to be done?". *Neuropsychologia*, *48*(12), 3416–3429.

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.

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

Petkova, V. I., & Ehrsson, H. H. (2008). If i were you: Perceptual illusion of body swapping. *PLoS ONE*, *3*(12), e3832. https://doi.org/10.1371/journal.pone.0003832

Pine, D. S., Grun, J., Maguire, E. A., Burgess, N., Zarahn, E., Koda, V., … Bilder, R. M. (2002). Neurodevelopmental aspects of spatial navigation: A virtual reality fMRI study. *Neuroimage*, *15*(2), 396–406.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2009). Lme: Linear and nonlinear mixed effects models. R package version 3.1-92.

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

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2010). Does it make a difference if i have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, *6*(4), 486–494. https://doi.org/10.1093/scan/nsq068

Pönkänen, L. M., Peltola, M. J., & Hietanen, J. K. (2011). The observer observed: Frontal eeg asymmetry and autonomic responses differentiate between another person's direct and averted gaze when the face is seen live. *International Journal of Psychophysiology*, *82*(2), 180–187.

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. (1995). Descriptive psychopathology of social phobia. *Social Phobia: Diagnosis, Assessment, and Treatment*, 41–66.

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

Rapee, R. M., & Lim, L. (1992). Discrepancy between self- and observer ratings of performance in social phobics. *Journal of Abnormal Psychology*, *101*(4), 728–731. https://doi.org/10.1037/0021-843x.101.4.728

Rapee, R. M., & Spence, S. H. (2004). The etiology of social phobia: Empirical evidence and an initial model. *Clinical Psychology Review*, *24*(7), 737–767.

Read, J. C. A., Phillipson, G. P., Serrano-Pedraza, I., Milner, A. D., & Parker, A. J. (2010). Stereoscopic vision in the absence of the lateral occipital cortex. *PLoS ONE*, *5*(9), e12608. https://doi.org/10.1371/journal.pone.0012608

Redding, R. E. (2001). Sociopolitical diversity in psychology: The case for pluralism. *American Psychologist*, *56*(3), 205–215. https://doi.org/10.1037/0003-066x.56.3.205

Richter, S. H., Garner, J. P., Auer, C., Kunert, J., & Würbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nature Methods*, *7*(3), 167.

Richter, S. H., Garner, J. P., & Würbel, H. (2009). Environmental standardization: Cure or cause of poor reproducibility in animal experiments? *Nature Methods*, *6*(4), 257.

Rinck, M., Rörtgen, T., Lange, W.-G., Dotsch, R., Wigboldus, D. H., & Becker, E. S. (2010). Social anxiety predicts avoidance behaviour in virtual encounters. *Cognition and Emotion*, *24*(7), 1269–1276.

Risko, E. F., & Kingstone, A. (2010). 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., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science. *Current Directions in Psychological Science*, *25*(1), 70–74. https://doi.org/10.1177/0963721415617806

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, *9*(3), 507–513. https://doi.org/10.3758/bf03196306

Ristic, J., Wright, A., & Kingstone, A. (2007). Attentional control and reflexive orienting to gaze and arrow cues. *Psychonomic Bulletin & Review*, *14*(5), 964–969. https://doi.org/10.3758/bf03194129

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, *17*(1), 31–35. https://doi.org/10.1111/j.1467-8721.2008.00543.x

Rosati, A. G., & Hare, B. (2009). Looking past the model species: Diversity in gaze-following skills across primates. *Current Opinion in Neurobiology*, *19*(1), 45–51. https://doi.org/10.1016/j.conb.2009.03.002

Ross, N. M., & Kowler, E. (2013). Eye movements while viewing narrated, captioned, and silent videos. *Journal of Vision*, *13*(4), 1–1.

Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PloS One*, *12*(7), e0182037.

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

Rubo, M., & Gamer, M. (2019). Visuo-tactile congruency influences the body schema during full body ownership illusion. https://doi.org/10.31234/osf.io/pe4v2

Rubo, M., Huestegge, L., & Gamer, M. (n.d.). Social anxiety modulates visual exploration in real life — but not in the laboratory.

Ruscio, A. M., Brown, T. A., Chiu, W. T., Sareen, J., Stein, M. B., & Kessler, R. C. (2007). Social fears and social phobia in the USA: Results from the national comorbidity survey replication. *Psychological Medicine*, *38*(01). https://doi.org/10.1017/s0033291707001699

Rutherford, M. D., & Krysko, K. M. (2008). Eye direction, not movement direction, predicts attention shifts in those with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(10), 1958–1965. https://doi.org/10.1007/s10803-008-0592-4

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4). https://doi.org/10.1515/revneuro.2003.14.4.303

Sanz, F. A., Olivier, A.-H., Bruder, G., Pettré, J., & Lécuyer, A. (2015). Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments. In *Virtual reality (vr), 2015 ieee* (pp. 75–80). IEEE.

Scheller, E., Büchel, C., & Gamer, M. (2012). Diagnostic features of emotional expressions are processed preferentially. *PloS One*, *7*(7), e41792.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). A second-person neuroscience in interaction. *Behavioral and Brain Sciences*, *36*(04), 441–462. https://doi.org/10.1017/s0140525x12002452

Schneider, B. W., & Turk, C. L. (2014). Examining the controversy surrounding social skills in social anxiety disorder: The state of the literature. *The Wiley Blackwell Handbook of Social Anxiety Disorder*, 366–387.

Schneider, G. E. (1969). Two visual systems. *Science*.

Schneier, F. R., Blanco, C., Antia, S. X., & Liebowitz, M. R. (2002). The social anxiety spectrum. *Psychiatric Clinics of North America*, *25*(4), 757–774.

Schneier, F. R., Johnson, J., Hornig, C. D., Liebowitz, M. R., & Weissman, M. M. (1992). Social phobia: Comorbidity and morbidity in an epidemiologic sample. *Archives of General Psychiatry*, *49*(4), 282–288.

Schneier, F. R., Rodebaugh, T. L., Blanco, C., Lewin, H., & Liebowitz, M. R. (2011). Fear and avoidance of eye contact in social anxiety disorder. *Comprehensive Psychiatry*, *52*(1), 81–87. https://doi.org/10.1016/j.comppsych.2010.04.006

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

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Schulze, L., Renneberg, B., & Lobmaier, J. S. (2013). Gaze perception in social anxiety and social anxiety disorder. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00872

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

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., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

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

Skarbez, R., Frederick P. Brooks, J., & Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Computing Surveys*, *50*(6), 1–39. https://doi.org/10.1145/3134301

Slater, M. (1999). Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *8*(5), 560–565. https://doi.org/10.1162/105474699566477

Slater, M., Brogni, A., & Steed, A. (2003). Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th annual international workshop on presence* (Vol. 157). Citeseer.

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., … Friedman, D. (2006). Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, *15*(5), 553–569. https://doi.org/10.1162/pres.15.5.553

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PloS One*, *5*(5), e10564.

Slater, M., Usoh, M., & Chrysanthou, Y. (1995). The influence of dynamic shadows on presence in immersive virtual environments. In *Eurographics* (pp. 8–21). Springer Vienna. https://doi.org/10.1007/978-3-7091-9433-1_2

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-n design. *Psychonomic Bulletin & Review*, *25*(6), 2083–2101. https://doi.org/10.3758/s13423-018-1451-8

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8), 16–16.

Sohn, K. (2015). The height premium in indonesia. *Economics & Human Biology*, *16*, 1–15. https://doi.org/10.1016/j.ehb.2013.12.011

Spector, I. P., Pecknold, J. C., & Libman, E. (2003). Selective attentional bias related to the noticeability aspect of anxiety symptoms in generalized social phobia. *Journal of Anxiety Disorders*, *17*(5), 517–531. https://doi.org/10.1016/s0887-6185(02)00232-3

Spence, S. H., Donovan, C., & Brechman-Toussaint, M. (1999). Social skills, social outcomes, and cognitive features of childhood social phobia. *Journal of Abnormal Psychology*, *108*(2), 211.

Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2006). 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. American Psychological Association (APA). https://doi.org/10.1037/t06496-000

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

State, M. W., & Sestan, N. (2012). The emerging biology of autism spectrum disorders. *Science*, *337*(6100), 1301–1303. https://doi.org/10.1126/science.1224989

Stein, M. B. (2006). An epidemiologic perspective on social anxiety disorder. *The Journal of Clinical Psychiatry*, *67*, 3–8.

Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *The Lancet*, *371*(9618), 1115–1125. https://doi.org/10.1016/s0140-6736(08)60488-2

Stein, M. B., Walker, J. R., & Forde, D. R. (1994). Setting diagnostic thresholds for social phobia: Considerations from a community survey of social anxiety. *The American Journal of Psychiatry*, *151*(3), 408.

Stoffregen, T. A., Hettinger, L. J., Haas, M. W., Roe, M. M., & Smart, L. J. (2000). Postural instability and motion sickness in a fixed-base flight simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *42*(3), 458–469. https://doi.org/10.1518/001872000779698097

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Stulp, G., Buunk, A. P., & Pollet, T. V. (2013a). Women want taller men more than men want shorter women. *Personality and Individual Differences*, *54*(8), 877–883. https://doi.org/10.1016/j.paid.2012.12.019

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2013b). Tall claims? Sense and nonsense about the importance of height of US presidents. *The Leadership Quarterly*, *24*(1), 159–171. https://doi.org/10.1016/j.leaqua.2012.09.002

Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., & Slater, M. (2017). Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-09497-3

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*(8), 764–769. https://doi.org/10.1038/77666

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4–4.

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.

Taubert, J., Flessert, M., Wardle, S. G., Basile, B. M., Murphy, A. P., Murray, E. A., & Ungerleider, L. G. (2018). Amygdala lesions eliminate viewing preferences for faces in rhesus monkeys. *Proceedings of the National Academy of Sciences*, *115*(31), 8043–8048. https://doi.org/10.1073/pnas.1807245115

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review*, *9*(2), 314–318. https://doi.org/10.3758/bf03196287

Tomasello, M., Call, J., & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, *55*(4), 1063–1069. https://doi.org/10.1006/anbe.1997.0636

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, *52*(3), 314–320. https://doi.org/10.1016/j.jhevol.2006.10.001

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

Trivers, R., & others. (1972). *Parental investment and sexual selection* (Vol. 136). Biological Laboratories, Harvard University Cambridge.

Tseng, P.-H., Cameron, I. G. M., Pari, G., Reynolds, J. N., Munoz, D. P., & Itti, L. (2012). High-throughput classification of clinical populations from natural viewing eye movements. *Journal of Neurology*, *260*(1), 275–284. https://doi.org/10.1007/s00415-012-6631-2

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

Journal of Vision, *9*(7), 4–4.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.

Turner, S. M., Beidel, D. C., Dancu, C. V., & Stanley, M. A. (1989). An empirically derived inventory to measure social fears and anxiety: The social phobia and anxiety inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *1*(1), 35–40. https://doi.org/10.1037/1040-3590.1.1.35

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, *9*(5), 497–503. https://doi.org/10.1162/105474600566989

Van Buuren, S. (2018). *Flexible imputation of missing data*. Chapman; Hall/CRC.

VanRullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. *Vision Research*, *46*(18), 3017–3027. https://doi.org/10.1016/j.visres.2005.07.009

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160113. https://doi.org/10.1098/rstb.2016.0113

Velden, M., & Wölk, C. (1987). Depicting cardiac activity over real time: A proposal for standardization. *Journal of Psychophysiology*.

Veljaca, K.-A., & Rapee, R. M. (1998). Detection of negative and positive audience behaviours by socially anxious subjects. *Behaviour Research and Therapy*, *36*(3), 311–321. https://doi.org/10.1016/s0005-7967(98)00016-3

Vick, S.-J., & Anderson, J. R. (2000). Learning and limits of use of eye gaze by capuchin monkeys (cebus apella) in an object-choice task. *Journal of Comparative Psychology*, *114*(2), 200–207. https://doi.org/10.1037/0735-7036.114.2.200

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, *45*(1), 135–170.

Von Hippel, W., & Buss, D. M. (2017). Do ideologically driven scientific agendas impede the understanding and acceptance of evolutionary principles in social psychology? In *Politics of social psychology* (pp. 17–35). Psychology Press.

Vuilleumier, P., & Huang, Y.-M. (2009). Emotional attention: Uncovering the mechanisms of affective biases in perception. *Current Directions in Psychological Science*, *18*(3), 148–152.

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

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

Welsh, T. N., Lyons, J., Weeks, D. J., Anson, J. G., Chua, R., Mendoza, J., & Elliott, D. (2007). Within- and between-nervous-system inhibition of return: Observation is as good as performance. *Psychonomic Bulletin & Review*, *14*(5), 950–956. https://doi.org/10.3758/bf03194127

Wienrich, C., Gross, R., Kretschmer, F., & Muller-Plath, G. (2018). Developing and proving a framework for reaction time experiments in VR to objectively measure social interaction with virtual agents. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE. https://doi.org/10.1109/vr.2018.8446352

Wieser, M. J., McTeague, L. M., & Keil, A. (2011). Sustained preferential processing of social threat cues: Bias without competition? *Journal of Cognitive Neuroscience*, *23*(8), 1973–1986.

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

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wittchen, H.-U., & Fehm, L. (2004). Comorbidity in social anxiety disorder. In *Social anxiety disorder* (pp. 60–73). CRC Press.

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), 309–323.

Würbel, H. (2000). Behaviour and the standardization fallacy. *Nature Genetics*, *26*(3), 263–263. https://doi.org/10.1038/81541

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

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171–211). Springer.

Yee, N., & Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33*(3), 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*(1), 3–47.

Yin, J., Xu, H., Duan, J., & Shen, M. (2018). Object-based attention on social units: Visual selection of hands performing a social interaction. *Psychological Science*, *29*(7), 1040–1048. https://doi.org/10.1177/0956797617749636

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, *10*(6), 224–228.

Zeifman, D., Delaney, S., & Blass, E. M. (1996). Sweet taste, looking, and calm in 2- and 4-week-old infants: The eyes have it. *Developmental Psychology*, *32*(6), 1090–1099. https://doi.org/10.1037/0012-1649.32.6.1090

Zhao, Q., & Koch, C. (2011). Learning a saliency map using fixated locations in natural scenes. *Journal of Vision*, *11*(3), 9–9.

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., & Domes, G. (2019). Virtually stressed? A refined virtual reality adaptation of the trier social stress test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*, *101*, 186–192. https://doi.org/10.1016/j.psyneuen.2018.11.010

Zuckerman, M., Miserandino, M., & Bernieri, F. (1983). Civil inattention existsin elevators. *Personality and Social Psychology Bulletin*, *9*(4), 578–586. https://doi.org/10.1177/0146167283094007

CHAPTER 6

APPENDIX

Supplementary for Study 2

Supplemental Methods

Prescreening Procedure

Online pre-screening of participants was achieved using a set of five questions based on the definition of Social Anxiety Disorders according to the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013). The following questions were rated on a 5-point-Likert-scale ranging from 1 (do not agree at all) to 5 (strongly agree):

1. I pronouncedly fear showing behavior in social or performance situations which could be embarrassing or humiliating for me.
2. Social or performance situations almost always elicit strong fear in me.
3. I believe that my fear of social or performance situations is unfounded and exaggerated.
4. I avoid social or performance situations when possible. When I cannot avoid them, I bear them with intense fear and discomfort.
5. I feel impaired by my fear of social and performance situations in my daily life.

These 5 questions were embedded into a brief questionnaire that also included questions on age, gender and current occupation. Participants could leave an email address or phone number if they agreed to be contacted for further psychological studies. In total, 814 participants completed the questionnaire. Since we were mainly interested in the effects of social anxiety on fixation patterns, we recruited participants based on their mean score across the corresponding 5 screening questions and specifically contacted participants who had a mean score of 3

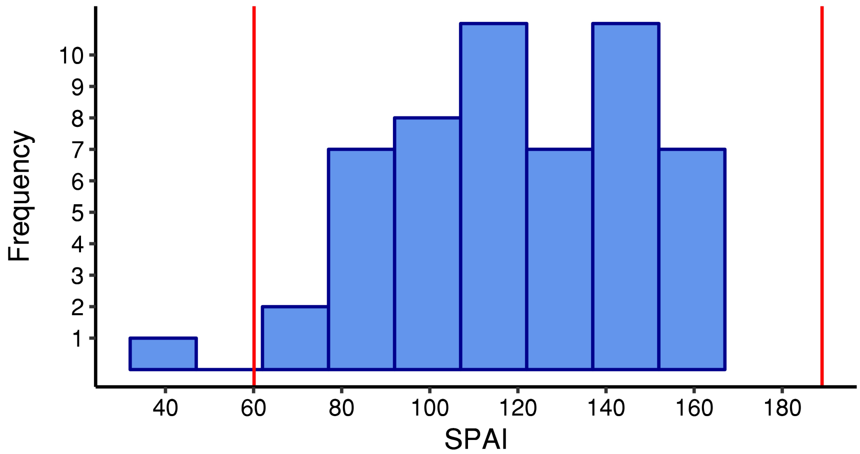

Figure 6.1: Histogram of SPAI scores of all participants in the real-life and the laboratory group. Vertical red lines mark scores two standard deviations below and above the mean. The participant with a SPAI score lower than two standard deviations below the mean was excluded from further analyses.

or higher.

After the experiment, participants completed questionnaires on depression (Beck Depression Inventory, BDI (Beck, 1961)), trait anxiety (trait version of the State Trait Anxiety Inventory, STAI (Spielberger, 1983)) and social anxiety (Social Interaction Anxiety Scale, SIAS, (Mattick & Clarke, 1998)), and Social Phobia and Anxiety Inventory, SPAI, (Turner et al., 1989)); for details on these instruments see below). Overall, prescreening scores corresponded well with the detailed assessment of social anxiety. However, we noted a substantial discrepancy between the prescreening and the detailed assessment of social anxiety in one participant in the real-life group. This participant had an unusual low SPAI score (SPAI, 2.67 SDs below the mean, see 6.1) but at the same time exhibited a rather high score on the pre-screening measure for social anxiety (1.51 SDs above the mean). Because of this incongruous pattern as well as the low data quality with less than 75% valid eye-tracking data, we decided to exclude this participant from the sample. Since we could not ensure accurate matching of this participant with his partner from the laboratory condition, we excluded the whole pair.

Questionnaires

Upon completing the experiment, participants completed the following questionnaires:

Beck Depression Inventory (BDI, (Beck, 1961)): The BDI is a self-report inventory consisting of 21 multiple-choice questions that address symptoms of depression. Each question has a set of four possible answers that range in intensity. To score the BDI, a value of 0 (no indication of depression) to 3 (severe depression) is assigned for each answer and values are added together for the whole test. BDI sum scores thus range from 0 to 63.

State Trait Anxiety Inventory (STAI, (Spielberger, 1983)): The STAI allows for an assessment of state and trait anxiety. In the current study, only the trait version was used. This questionnaire consists of 20 short descriptions of emotional states that should be rated regarding their habitual occurrence on a 4-point Likert scale ranging from 1 (almost never) to 4 (almost always). STAI sum scores range from 20 to 80 with higher values indicating higher levels of trait anxiety.

Social Interaction Anxiety Scale (SIAS, (Mattick & Clarke, 1998)): The SIAS is a self-report scale that measures fear in interactions with other people. It includes 20 items that are answered on a 5-point Likert scale ranging from 0 (not at all characteristic of me) to 4 (extremely characteristic of me). Individual answers are summed up to produce a total score that ranges from 20 to 80 with higher values reflecting higher social anxiety.

Social Phobia Anxiety Inventory (SPAI, (Turner et al., 1989)): The SPAI assesses specific somatic symptoms, cognitions, and behaviors across a range of potentially fear-producing social situations. It consists of 22 questions that partly include subsections to allow for a comprehensive assessment of social anxiety. Questions are answered on a 7-point Likert scale ranging from 1 (never) to 7 (always). The SPAI is scored by first averaging answers across subsections for each question and then summing up the resulting values. Since the suppressor function of the agoraphobia subscale that is part of the original version of the SPAI (Turner et al., 1989) could not be verified in a German sample (Fydrich, 2002), this subscale was not employed in the current study. Moreover, the German version of the questionnaire differs from the original version in the number of items (22 vs. 32). SPAI sum scores were therefore linearly transformed in the current study to be comparable to the original version. Resulting sum scores vary between 32 and 224 with higher scores indicating higher social anxiety. Internal consistency, measured as Cronbach's α, was high among all four questionnaires for the 56 participants who remained in the analysis (BDI: α = .91, STAI: α = .83, SIAS: α = .90, SPAI: α = .94).

Data Reduction and Analysis

For each participant, gaze in every sixth frame (i.e. every 200 ms) was coded based on a ring displayed in the video indicating the current gaze location 6.2. The coding scheme was adapted from previous studies using mobile eye-tracking (Foulsham et al., 2011). The category person was coded when the gaze ring touched any displayed person. Gaze was coded as being directed towards an object if the gaze ring did not touch a person, but any non-walkable, but touchable location in the image (e.g., bags, pillars, seats, walls). If the gaze ring was located on a walkable surface, but did not touch a person or an object, it was labeled as path. If the gaze ring could not be categorized as any of these categories, it was labeled as undefined (e.g., when it was directed towards the sky or the ceiling). A data point was labeled as valid only if gaze was tracked, the experimenter was not visible and the participant was not involved in an unplanned conversation. In addition to specifying the category of the attentional focus, we also determined whether gaze was located at a position within (< 8 m) or beyond the near-distant action space (> 8 m).

Since coding was accomplished by two different raters, we tested for their inter-

Figure 6.2: Examples of individual frames from recordings of the participants' point of view. The orange circle indicates gaze position concurrently tracked by the mobile eye-tracking device. In a) and b), the participant is currently looking at a person. In c), gaze is directed on the path, and in d), gaze is directed at an object. Gaze was coded to be directed in the vicinity in a), and in the distance in b), c) and d).

rater reliability for a subset of five matched pairs of videos that were additionally coded by both raters (31,461 frames in total). Interrater agreement was only tested for the coding categories that were used in the main analyses (i.e., invalid frames, undefined gaze locations and frames in the dataset in which no persons were visible were excluded). Interrater reliability was evaluated on the basis of their agreement as well as by calculating Cohen's kappa with the *irr*-package (Gamer, Lemon, Fellows, & Singh, 2012) for the R software for statistical computing (version 3.2; R Development Core Team, 2015). Kappa values between .61 and .80 are commonly considered as substantial agreement and values between .81 and 1.00 as almost perfect agreement (Landis & Koch, 1977). In our sample, the categorization of the fixation target (person, path or object) agreed between raters in 89.14% of all cases with a kappa of .81. The distance estimation (near vs. far) agreed in 89.09% and kappa amounted to .68. For the whole set of all 6 coding categories (person vs. object vs. path × near vs. distant), raters agreed in 80.06% of all frames and kappa was .72. These values reflect high interrater reliability.

Supplemental Results

Influence of Social Anxiety on Gaze at Objects

We compared the influence of social anxiety on gaze at persons with a possible influence on gaze at objects in order to validate a social specificity of the gaze modulation in socially anxious individuals. The effect of social anxiety could not be reproduced for gaze on objects. The respective linear mixed model only revealed a significant effect of distance ($F(1{,}74) = 245.37$, $p < .001$) as well as a group × distance interaction ($F(1{,}74) = 11.15$, $p = .001$). Figure 6.3 indicates that participants preferred to fixate near objects as compared to distant ones but this effect was more pronounced in the laboratory condition. No other main or interaction effect reached the level of significance (group: $F(1{,}74) = 2.65$, $p = .108$; SPAI: $F(1{,}74) = 0.01$, $p = .925$; distance × SPAI: $F(1{,}74) = 0.01$, $p = .922$;

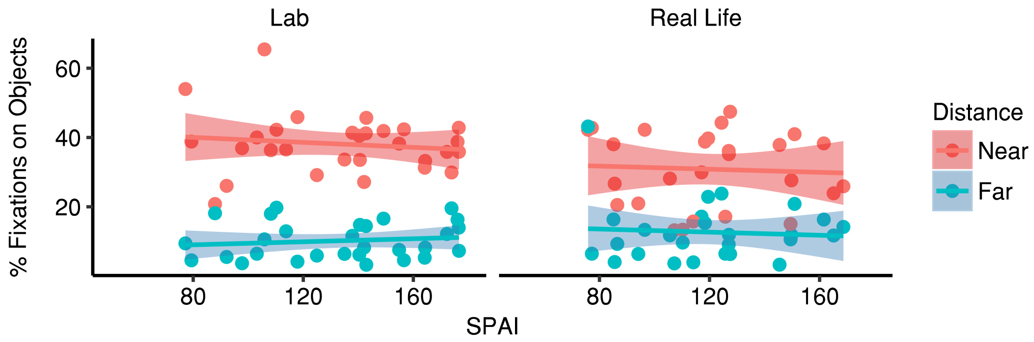

Figure 6.3: Frequency of gaze at objects as a function of group, distance and SPAI. Shaded areas represent 95% confidence intervals of the linear regression.

group × SPAI: $F(1{,}74) = 0.35$, $p = .558$; group × distance × SPAI: $F(1{,}74) = 0.30$, $p = .583$).

Influence of Social Anxiety (as measured by SIAS) and General Anxiety (STAI) on Gaze at Persons

As stated in the main part of the manuscript, the influence of social anxiety as measured by the SPAI could be qualitatively replicated using the SIAS questionnaire, while general anxiety measured by the STAI did not modulate gaze behavior. When incorporating SIAS scores into a linear mixed model predicting fixations on persons for the real-life and the laboratory group (Figure 6.4), we found a main effect of distance ($F(1{,}74) = 205.96$, $p < .001$), a group × distance interaction ($F(1{,}74) = 7.46$, $p = .008$) and a marginally significant group × distance × SIAS interaction ($F(1{,}74) = 2.82$, $p = .098$). None of the other main or interaction effects reached the level of significance (SIAS: $F(1{,}74) = 0.82$, $p = .367$, group: $F(1{,}74) = 2.40$, $p = .126$, distance × SIAS: $F(1{,}74) = 2.64$, $p = .109$; group × SIAS: $F(1{,}74) = 0.08$, $p = .780$). In the real-life group alone, we found a significant main effect for distance ($F(1{,}24) = 48.86$, $p < .001$) and a distance × SIAS interaction ($F(1{,}24) = 5.18$, $p = .032$), but no main effect for SIAS ($F(1{,}24) = 0.41$, $p = .528$). In the laboratory group alone, we found a main effect for distance ($F(1{,}27) = 191.07$, $p < .001$), but no main effect for SIAS ($F(1{,}27) = 0.17$, $p =$

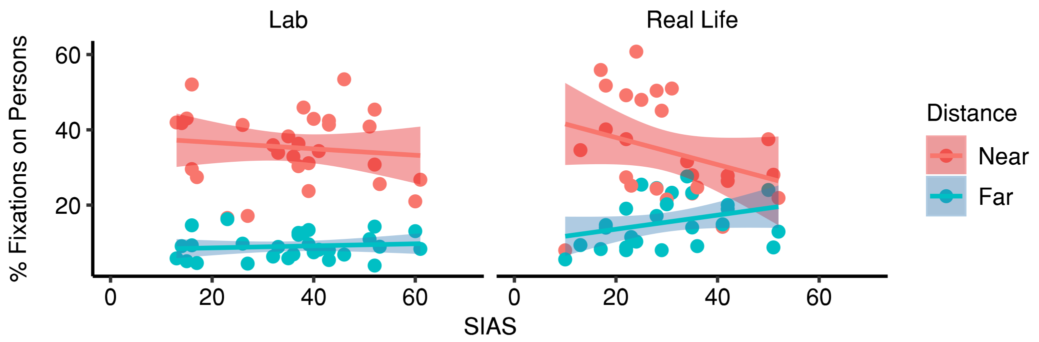

Figure 6.4: Frequency of gaze at persons as a function of group, distance and SIAS. Shaded areas represent 95% confidence intervals of the linear regression.

.681) and no distance × SIAS interaction ($F(1{,}27) = 0.65$, $p = .426$).

When incorporating STAI scores into a linear mixed model predicting fixations on persons for the real-life and the laboratory group (Figure 6.5), we found a main effect of distance ($F(1{,}74) = 194.44$, $p < .001$) and a group × distance interaction ($F(1{,}74) = 4.61$, $p = .035$). None of the other main or interaction effects reached the level of significance (STAI: $F(1{,}74) = 2.34$, $p = .131$, group: $F(1{,}74) = 2.11$, $p = .151$, distance × STAI: $F(1{,}74) = 0.56$, $p = .458$; group × STAI: $F(1{,}74) = 0.05$, $p = .829$, group × distance × STAI: ($F(1{,}74) = 0.40$, $p = .531$).

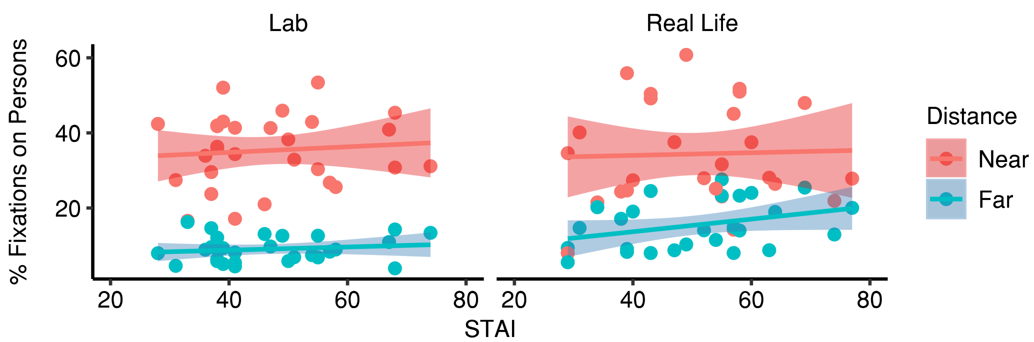

Figure 6.5: Frequency of gaze at persons as a function of group, distance and STAI. Shaded areas represent 95% confidence intervals of the linear regression.

Curriculum Vitae

Marius Rubo

Address: Rimparer Steig 7, 97080 Würzburg, Germany

Email: marius.rubo@uni-wuerzburg.de

Github: github.com/mariusrubo

EDUCATION

Dr.rer.nat. Psychology	*Expected: June 2019* *Würzburg University, Germany* *Social Attention in the Laboratory, in Real Life and in Virtual Reality*
M.Sc. Psychology	*Nov 2013* *Freiburg University, Germany* *Final grade: 1.0*
B.Sc. Psychology	*Aug 2011* *Freiburg University, Germany* *Final grade: 1.4*

RESEARCH AREAS

Primary: Social Attention, Human Body Image, Virtual Reality

Secondary: Social Anxiety, R Programming, C# Programming

PUBLISHED WORKS IN REFEREED JOURNALS

Rubo, M., & Gamer, M. (2019). Social Anxiety modulates Visual Exploration in Real-Life — but not in the Laboratory. *British Journal of Psychology.*

Rubo, M., & Gamer, M. (2018). Social Content and Emotional Valence modulate Gaze Fixations in Dynamic Scenes. *Scientific Reports, 8*(1), 3804.

Ben-Moussa, M., **Rubo, M.**, Debracque, C., & Lange, W. G. (2017). DJINNI: A novel technology supported exposure therapy paradigm for SAD combining virtual reality and augmented reality. *Frontiers in psychiatry, 8*(26).

OTHER PUBLICATIONS

Rubo, M., & Gamer, M. (2018, June). Tracing gaze-following behavior in virtual reality using wiener-granger causality. *In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (p. 86). ACM.

Rubo, M., & Gamer, M. (2018, June). Virtual reality as a proxy for real-life social attention?. In Proceedings of the 2018 *ACM Symposium on Eye Tracking Research & Applications* (p. 81). ACM.

UNDER REVIEW

Rubo, M., & Gamer, M. Visuo-tactile Congruency influences the Body Schema during Full Body Ownership Illusion.

Rubo, M., & Gamer, M. Natural Gaze Behavior and Stable Intra-individual Gaze Preferences in a Social Virtual Scenario with a

Reactive Virtual Agent.

AWARDS

- Ferry-Forsche price, 2008
- Karl-Steinbuch scholarship, 2013
- G.A.-Lienert scholarship, 2017

TEACHING EXPERIENCE

- Seminars in
  - General Psychology
  - Biological Psychology
  - Body Image
- Lecture (on the Human Visual System)
- Several theses
  - 6 Bachelor's theses
  - 2 Master's theses

Statement of Individual Author Contributions

Rubo, M. & Gamer, M. (2018). Social Content and Emotional Valence modulate Gaze Fixations in Dynamic Scenes. *Scientific Reports, 8(1).* doi:10.1038/s41598-018-22127-w					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	M.G. M.R.	M.R. M.G.			
Data Collection	M.R.	M.G.			
Data Analysis and Interpretation	M.R.	M.G.			
Manuscript Writing	M.R.	M.G.			

M. Rubo: conceptualization, data curation, formal analysis, visualization, writing original draft, writing review and editing.

M. Gamer: supervision, funding acquisition, conceptualization, validation, writing review and editing.

Rubo, M., Huestegge, L. & Gamer, M. (2019). Social anxiety modulates visual exploration in real-life – but not in the lab. *British Journal of Psychology*					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	M.G. M.R.	M.R. M.G.	L.H. L.H.		
Data Collection	M.R.	M.G.	L.H.		
Data Analysis and Interpretation	M.R.	M.G.	L.H.		
Manuscript Writing	M.R.	M.G.	L.H.		

M. Rubo: conceptualization, data curation, formal analysis, visualization, writing original draft, writing review and editing.

M. Gamer: supervision, funding acquisition, conceptualization, validation, writing original draft, review and editing.

L. Huestegge: supervision, conceptualization, validation, writing review and editing.

Marius Rubo		Würzburg	
Doctoral Researcher's Name	Date	Place	Signature

Prof. Dr. Matthias Gamer		Würzburg	
Primary Supervisor's Name	Date	Place	Signature

Affidavit

AFFIDAVIT

I hereby confirm that my thesis entitled “Social Attention in Dynamic Stimuli, Real-World Situations and Virtual Reality” 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 “Social attention in dynamic stimuli, real-world situations and virtual reality” eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters angefertigt zu haben und keine anderen als die von mir angegebenen Quellen uns Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form in einem anderen Prüfungsverfahren vorgelegt wurde.

Ort, Datum | Unterschrift