The Ubiquity of Social Attention – a Detailed Investigation of the Underlying Mechanisms

Die Allgegenwärtigkeit Sozialer Aufmerksamkeit – eine Detaillierte Erforschung Zugrundeliegender Mechanismen

DISSERTATION

For a Doctoral Degree (Dr. rer. nat.)

At the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Neuroscience

Submitted by

ALEYA FELICIA FLECHSENHAR

Born in Stuttgart

Würzburg, 2019

MEMBERS OF THE COMMITTEE:

Chairperson:	Prof. Dr. Paul Pauli
Primary Supervisor:	Prof. Dr. Matthias Gamer
Supervisor (Second):	Prof. Dr. Lynn Huestegge
Supervisor (Third):	PD Dr. Angelika Schmitt-Böhrer

Date of Public Defense: ____

Date of Receipt of Certificates: _________________________________

Submitted on: ____

Office stamp

TO MY FAMILY

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and support of several individuals who in one way or another have contributed to the preparation and completion of the work and words of this thesis.

First of all I offer my gratitude to my supervisor Prof. Dr. Matthias Gamer, who supported me throughout the work of my entire Ph.D. with his patience and expert knowledge. His suggestions were always thorough and valuable and he offered sufficient freedom for creativity and efficiency. I am sincerely grateful to have been able to work in such a great, supportive and insightful environment.

I would further like to thank my supervisors Prof. Dr. Lynn Huestegge and PD Dr. Angelika Schmitt-Böhrer for their patience, support, encouragement, and feedback.

An important anchor for technical, as well as emotional support were given by my colleagues through the ups and downs, and the question marks during the past four years. I always considered our work environment to be constructive and invigorating, and enjoyed the translation to the personal level.

I would also like to thank my friends, without whom life in Würzburg would not have been the same. I am thankful for the great conversations, the recurrent words of motivation, but also the shared times of adversity and solace. I could always rely on an open ear and an open heart.

Special thanks also goes to my partner for the outlasting vote of confidence for my academic life and giving me an equilibrium in my personal life. Thank you for your patience in listening to preparations of talks, reading through documents, weighing and supporting decisions, and making me smile.

Last, but not least I am eternally grateful to my family for not only giving me the possibility and guidance to achieve this goal, but also for the courage, the belief in myself, the stability, and the unconditional love throughout my life. Words are not enough to describe my gratitude.

FINANCIAL SUPPORT

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

ETHICAL APPROVAL

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 declaration of Helsinki.

CONFLICT OF INTEREST

The authors declare no competing interests.

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

ANOVA	Analysis of variance
AUC	Area under the receiver-operating curve
CORM	Center of recurrent mass
DGP	German Psychological Society
D_{KL}	Kullback-Leibler divergence
EEG	Electroencephalography
ERP	Event-related potential
FFA	Fusiform face area
fMRI	Functional magnetic resonance imaging
FV	Free-viewing
GBVS	Graph-based visual saliency
GC	Gaze-contingent
HSF	High spatial frequency
Hz	Hertz
ICF	Intensity contrast features
IPS	Interparietal sulcus
IOR	Inhibition of return
IT	Inferior temporal cortex
ITI	Inter-trial interval
LSF	Low spatial frequency
M	Mean
MEG	Magnetoencephalography
MFG	Middle frontal gyrus
ms	Milliseconds
MT	Medial superior temporal area
PPA	Parahippocampal place area
PFC	Prefrontal cortex
pMTG	Posterior middle temporal gyrus
pt	Point
px	Pixel
OFC	Orbito-frontal cortex
r	Pearson's product-moment correlation coefficient
ROI	Regions of interest

RQA	Recurrent quantification analysis
SALICON	Saliency in context
SD	Standard deviation
SOA	Stimulus onset asynchrony
SSVEP	Steady-state visual evoked potential
STG	Superior temporal gyrus
STS	Superior temporal sulcus
ToM	Theory of Mind
TPJ	Temporal-parietal junction
VR	Virtual Reality
V1	Human striate visual cortex
V2	Human prestriate visual cortex
V3	Third visual complex with visual area V3
V4	Human extrastriate cortex
V5	Human middle temporal visual area

SUMMARY

This dissertation highlights various aspects of basic social attention by choosing versatile approaches to disentangle the precise mechanisms underlying the preference to focus on other human beings. The progressive examination of different social processes contrasted with aspects of previously adopted principles of general attention. Recent research investigating eye movements during free exploration revealed a clear and robust social bias, especially for the faces of depicted human beings in a naturalistic scene. However, free viewing implies a combination of mechanisms, namely automatic attention (bottom-up), goal-driven allocation (top-down), or contextual cues and inquires consideration of overt (open exploration using the eyes) as well as covert orienting (peripheral attention without eye movement). Within the scope of this dissertation, all of these aspects have been disentangled in three studies to provide a thorough investigation of different influences on social attention mechanisms.

In the first study (section 2.1), we implemented top-down manipulations targeting non-social features in a social scene to test competing resources. Interestingly, attention towards social aspects prevailed, even though this was detrimental to completing the requirements. Furthermore, the tendency of this bias was evident for overall fixation patterns, as well as fixations occurring directly after stimulus onset, suggesting sustained as well as early preferential processing of social features. Although the introduction of tasks generally changes gaze patterns, our results imply only subtle variance when stimuli are social. Concluding, this experiment indicates that attention towards social aspects remains preferential even in light of top-down demands.

The second study (section 2.2) comprised of two separate experiments, one in which we investigated reflexive covert attention and another in which we tested reflexive as well as sustained overt attention for images in which a human being was unilaterally located on either the left or right half of the scene. The first experiment consisted of a modified dot-probe paradigm, in which peripheral probes were presented either congruently on the side of the social aspect, or incongruently on the non-social side. This was based on the assumption that social features would act similar to cues in traditional spatial cueing paradigms, thereby facilitating reaction times for probes presented on the social half as opposed to the non-social half. Indeed, results reflected such congruency effect. The second experiment investigated these reflexive mechanisms by monitoring eye movements and specifying the location of saccades and fixations for short as well as long presentation times. Again, we found the majority of initial saccades to be congruently directed to the social side of the stimulus. Furthermore, we replicated findings for sustained attention processes with highest fixation densities for the head region of the displayed human being.

The third study (section 2.3), tackled the other mechanism proposed in the attention dichotomy, the bottom-up influence. Specifically, we reduced the available contextual information of a scene by using a gaze-contingent display, in which only the currently fixated regions would be visible to the viewer, while the remaining image would remain masked. Thereby, participants had to voluntarily change their gaze in order to explore the stimulus. First, results revealed a replication of a social bias in free-viewing displays. Second, the preference to select social features was also evident in gaze-contingent displays. Third, we find higher recurrent gaze patterns for social images compared to non-social ones for both viewing modalities. Taken together, these findings imply a top-down driven preference for social features largely independent of contextual information.

Importantly, for all experiments, we took saliency predictions of different computational algorithms into consideration to ensure that the observed social bias was not a result of high physical saliency within these areas. For our second experiment, we even reduced the stimulus set to those images, which yielded lower mean and peak saliency for the side of the stimulus containing the social information, while considering algorithms based on low-level features, as well as pre-trained high-level features incorporated in deep learning algorithms.

Our experiments offer new insights into single attentional mechanisms with regard to static social naturalistic scenes and enable a further understanding of basic social processing, contrasting from that of non-social attention. The replicability and consistency of our findings across experiments speaks for a robust effect, attributing social attention an exceptional role within the general attention construct, not only behaviorally, but potentially also on a neuronal level and further allowing implications for clinical populations with impaired social functioning.

ZUSAMMENFASSUNG

Diese Dissertation beschäftigt sich mit verschiedenen Aspekten grundlegender sozialer Aufmerksamkeitsprozesse. Insbesondere werden durch vielseitige Herangehensweisen einzelne Mechanismen untersucht, die der bevorzugten Betrachtung von Menschen zugrunde liegen. Die progressive Untersuchung unterschiedlicher sozialer Vorgänge widerspricht einiger zuvor angenommener Grundlagen allgemeiner Aufmerksamkeitsprozesse. So zeigen beispielsweise Probanden bei freier Betrachtung naturalistischer Bilder eine klare Präferenz für abgebildete Menschen, v.a. deren Gesichter. Allerdings beinhaltet die freie Betrachtung eine Kombination aus mehreren Vorgängen, wie automatische (engl. bottom-up) und zielorientierte, willentliche Aufmerksamkeitslenkung (engl. top-down), als auch den Einfluss von Kontextzusammenhängen. Dies bedingt weiter die Berücksichtigung offener (engl. overt; Exploration mittels Augenbewegungen), als auch verdeckter Aufmerksamkeit (engl. covert; periphere Erkundung ohne Augenbewegungen). Im Rahmen der Dissertation werden alle genannten Aspekte anhand von drei Studien behandelt, wodurch eine sorgfältige Untersuchung verschiedener Einflüsse sozialer Aufmerksamkeitsprozesse erfolgt.

In der ersten Studie (Abschnitt 2.1) wurden zielgerichtete Manipulationen in Form von Aufgabenstellungen vorgenommen, welche die Aufmerksamkeit innerhalb einer sozialen Szene spezifisch auf nicht-soziale Reize lenken sollten, um kompetitive Ressourcen von sozialer und zielgerichteter Aufmerksamkeit zu untersuchen. Interessanterweise überwog die Tendenz, soziale Aspekte zu betrachten, obwohl dies nachteilig für das Lösen der Aufgaben war. Diese Neigung erwies sich für die allgemeine Fixationsverteilung als auch für Fixationen, die unmittelbar nach Erscheinen der Stimuli auftraten. Dieser Befund impliziert, dass soziale Reize sowohl bei dauerhaften als auch frühen Aufmerksamkeitsprozessen bevorzugt werden. Obwohl es einen allgemeinen Konsens gibt, dass eine Implementierung von Aufgaben zu verändertem Blickverhalten führt, deuten unsere Ergebnisse lediglich auf subtile Abweichungen hin, wenn die Stimuli sozialer Natur sind. Abschließend indiziert dieses Experiment, dass auch mit steigender Bedeutung anderer top-down Modulationen bevorzugt soziale Aspekte betrachtet werden.

Die zweite Studie (Abschnitt 2.2) bestand aus zwei separaten Experimenten, welche verdeckte und offene Aufmerksamkeit auf soziale Reize untersuchten. Hierfür wurden in beiden Studien dieselben Bilder verwendet, in denen ein Mensch unilateral, entweder auf der rechten oder linken Hälfte abgebildet war. Das erste Experiment bestand aus einer modifizierten Variante des Dot-Probe Paradigmas, bei dem periphere Zielreize entweder kongruent auf der Seite des sozialen Stimulus erschienen, oder inkongruent auf der nicht sozialen Seite präsentiert wurden. Diese Zuteilung basierte auf der Annahme, dass soziale Merkmale auf ähnliche Weise fungieren wie

Hinweisreize in traditionellen Spatial-Cueing-Paradigmen, indem sie Reaktionszeiten auf Zielobjekte, die auf der sozialen Seite präsentiert werden, beschleunigen. Tatsächlich wiesen unsere Ergebnisse einen solchen Kongruenzeffekt auf. Das zweite Experiment überprüfte die reflexiven Vorgänge durch die Messung von Augenbewegungen mittels Spezifizierung der Sakkadenrichtung und Fixationsdichte für kurze als auch lange Präsentationszeiten. Wiederum stellte sich heraus, dass die Mehrzahl der initialen Sakkaden kongruent zum sozialen Reiz gerichtet waren. Darüber hinaus wurden die Ergebnisse für kontinuierliche Aufmerksamkeitsprozesse durch eine erhöhte Fixationsdichte auf Kopfregionen der abgebildeten Menschen repliziert.

Die dritte Studie (Abschnitt 2.3) behandelte den umgekehrten Mechanismus der Aufmerksamkeitsdichotomie, nämlich den bottom-up Einfluss. Durch die Verwendung eines blickkongruenten Paradigmas, konnte der kontextuelle Informationsgehalt der Szenen so reduziert werden, dass nur noch der aktuell betrachtete Bereich sichtbar war, während der Rest des Bildes maskiert blieb. Somit mussten die Teilnehmer willentlich ihren Blick verändern, um die Szene zu erkunden. Erstens zeigten die Ergebnisse eine Replikation des sozialen Bias bei freier Betrachtung. Zweitens scheint die Präferenz, soziale Aspekte zu selektieren, in der blickkongruenten Darstellung bestehen zu bleiben. Drittens zeigte sich ein erhöhtes wiederkehrendes Blickmuster bei sozialen im Vergleich zu nicht sozialen Bildern für beide Betrachtungsmodalitäten. Zusammenfassend implizieren diese Ergebnisse eine zielgerichtete Präferenz für soziale Reize, welche größtenteils kontextunabhängig ist.

Hervorzuheben ist auch, dass bei allen Experimenten Salienzprädiktoren verschiedener Algorithmen in Betracht gezogen wurden, um sicher zu stellen, dass die Tendenz soziale Reize zu bevorzugen nicht alleine durch hohe physikalische Salienz in diesen Bereichen bedingt wurde. Insbesondere für die zweite Studie (Abschnitt 2.2) wurden Algorithmen verwendet, die sowohl untergeordnete Merkmale als Prädiktoren integrierten als auch Deep Learning Algorithmen, welche vortrainierte, übergeordnete Merkmale definieren, um Vorhersagen zu treffen. So wurde das verwendete Stimulusmaterial reduziert, so dass nur Bilder mit niedriger mittlerer als auch maximaler Salienz auf der nicht-sozialen Seite analysiert wurden.

Diese Experimente geben Aufschluss auf einzelne Aufmerksamkeitsprozesse bei der Betrachtung von statischen, sozialen, naturalistischen Szenen und ermöglichen ein tiefergehendes Verständnis für grundlegende soziale Verarbeitung, welche sich von nicht-sozialer Aufmerksamkeit abhebt. Die Replizierbarkeit und Konsistenz der Experimente implizieren einen robusten Effekt und suggerieren eine gesonderte Rolle der sozialen Aufmerksamkeit innerhalb des allgemeinen Aufmerksamkeitskonstrukts. Dies basiert nicht nur auf Verhaltensparametern, sondern potentiell auch auf neuronaler Ebene und enthält darüber hinaus auch Implikationen für klinische Populationen mit beeinträchtigten sozialen Funktionen.

1. INTRODUCTION

"We call that psychical process, which is operative in the clear perception of a narrow region of the content of consciousness, attention." (Wundt, 1912, p. 16)

Wilhelm Wundt marks psychology as an independent field of research (Wundt, 1909) and introduced the study of attention to the field. Although attention generally refers to a certain focus or consideration, as implied by Wundt, it is a widely used concept and the core property of perception and cognition (Chun, Golomb, & Turk-Browne, 2011). In fact, it is so generic that it is arduous to conceive, define, or study.

"Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction, and Zerstreutheit in German." (James, 1890*,* p. 403-404)

This well-known description by William James nicely depicts the complexity and multitude of processes implied by attentional mechanisms. As such, attention is the foundation of all higher order processing, but also comprises of several basic features. It can be influenced externally by sensory factors, but can also be the result of internally generated intellectual information, such as adherence to tasks, responses to occurrences, experience or memory (Chun et al., 2011). It describes a general concept covering all factors that influence selection (Borji & Itti, 2013). Selection implies a certain capacity limitation to process only relevant information coherent with ongoing goals and behaviors (Pashler, Johnston, & Ruthruff, 2001). The aim of attention research is to understand which information is selected and how. As a measure, researchers have studied visual processes, as gaze reflects cognition involved in filtering irrelevant information to focus on a specific feature (McMains & Kastner, 2009). Eye movements are necessary for high acuity vision of a certain location, which in turn indicates where attention is directed (see Findlay & Gilchrist, 2003). A variety of theories have been proposed and studies have been conducted to measure behavioral correlates of attention, investigate neuronal representations, and build neural network models to explain it (Borji & Itti, 2013).

1.1 ATTENTIONAL MECHANISMS

Attention is be the ability to select information for perception, processing and action, while ignoring other aspects in one's surroundings. Therefore, classical attention theories have attempted to answer why something is selected, how it is selected and what in particular raises interest. An early description of attention was deduced from the *spotlight model* of James (Eriksen & Hoffman, 1972) stating that information of a scene can be extracted by focusing on an area with a high-resolution, while the remaining aspects build a fringe of attention in low-resolution. The *zoom-lens model* extends the spotlight model by suggesting a trade-off between the focus and the size of the spotlight presuming attentional resources to be fixed. Therefore, the focus of attention onto the area of the retina with highest visual acuity, called the *fovea centralis*, amounts to 1-2° of visual angle, whereas the remaining visual field of 180° horizontally and 130° vertically, are only crudely represented (Eriksen & Hoffman, 1972). Therefore, eye movements allow the oculomotor system to constantly select parts of the environment with this high acuity for processing in order to incorporate a whole scene (Wilming et al., 2017). The human visual system extracts relevant perceptual information and action-related goals and generates links between these domains (Huestegge & Koch, 2010), acting as a natural integrator of perception and action (Findlay & Gilchrist, 2003). The previously mentioned models rely on the assumption that the human attention system has a capacity limitation and therefore items for processing are selected at the expense of others, but also to filter out unwanted information (Broadbent, 1958; Desimone & Duncan, 1995). Further models have been suggested as a consequence of these theories. Treisman & Gelade (1980) for example have proposed the *feature integration theory* to describe the order of attentional resource distribution in order to perceive a given scene by selectively attending features of an object. Similarly, Jonides (1983) concluded that attention is first distributed equally over a given scene and the information is processed in parallel and only then concentrated to a certain region of the scene with serial character. Posner (1980) states that visual attention can only focus on one location at a time and introduced a very renowned theory for spatial attention, in which he postulates that the allocation of attention is not necessarily dependent on eye movements. Specifically, he investigated the difference between *overt* (shifts of the eyes) and *covert* (mental shifts through movement of attentional focus) attention with the aid of central or peripheral cues (Posner, Nissen, & Ogden, 1978). Overt attention can be measured by means of eye movements, mirroring different orienting mechanisms. Following the *selection for action theory* of Allport (1987), which describes that attention is deployed primarily to guide and control actions, the *premotor theory of attention* (Rizzolatti, Riggio, Dascola, & Umiltá, 1987) states that covert spatial attention derives from an activation of brain maps, which increase the readiness to execute a motor response, as well as to facilitate perceptual processing at that location (Huestegge & Koch, 2010).

The traditional theories of attention imply a dichotomization of two processes for allocation of attention, namely exogenously driven mechanisms (*bottom-up*) that reflect sensory stimulation and endogenously driven control (*top-down*), which comprise factors, such as knowledge, expectation and goals (e.g., Corbetta & Shulman, 2002; Müller & Rabbitt, 1989; Theeuwes, 2010). The dynamic interaction of both mechanisms determines where attention is deployed. Herein, exogenous attention may interrupt endogenous orienting, also known as the cocktail party effect (Joseph & Optican, 1996), while it is also possible to affect the extent of reflexive attention capture through top-down modulations (Folk, Remington, & Johnston, 1992). Early experiments, like the one of Posner, tested spatial attention using cues to target this dichotomous attention. Specifically, two probes (e.g. boxes) would be presented on the left and right side of a screen. One of these probes would reveal a stimulus (e.g. checkerboard) to which the participant was asked to react via key press. Before the onset of the stimulus, one of the probes would be cued centrally (e.g. arrow) or in the periphery (e.g. flash). Depending on the type of cue, attention would be allocated either endogenously (central cue), or exogenously (peripheral cue) to one of the probes. A cue can predict the location of the stimulus either correctly (valid cue) or incorrectly (invalid cue). The standard finding in Posner paradigms is that a target is detected more quickly when the cue is valid and attention is deployed to its location prior to its onset. Results showed that participants responded faster to exogenous cues (maximum facilitation achieved by peripheral cues 100-175 ms SOAs, see Müller & Rabbitt, 1989), reflecting an automatic attentional capture, whereas endogenous cues required longer processing (maximum facilitation achieved by central cues 400-725 ms SOAs, see Müller & Rabbitt, 1989) due to the engagement of cognitive capacities to suppress bottom-up mechanisms. Exogenous orienting seems to be less disrupted by secondary tasks and more resistant to suppression, compared to endogenous orienting (Jonides, 1981). This effect is not only evident on a behavioral basis, but also electrophysiologically, as shown by Gazzaniga, Ivry, and Mangun (2013). The first positive deflection (P100 of the event-related brain potential) depicts a higher amplitude for congruently presented targets compared to incongruent ones.

1.1.1 EYE MOVEMENTS

The core concept of visual attention is mostly related to the fact that we move our eyes to gather information about the world around us, rendering the study of eye movements an important method for cognitive psychology (Rayner, 1998, 2009; Risko, Anderson, Lanthier, & Kingstone, 2012). The development of eye trackers allowed for diagnostic (objective and quantitative information about overt attentional processes) as well as interactive applications (selective and gaze-dependent information gain; Duchowski, 2002). Therefore, eye tracking offers a powerful tool to investigate unique behavioral responses in time, accuracy, location, amplitude and duration (Anderson, Bischof, Laidlaw, Risko, & Kingstone, 2013). Furthermore, eye movements offer insight to ongoing cognitive processes as well as implicit, unconscious performance (Hayhoe, 2004), without having to rely on verbal reports or introspective judgements. Additionally, this ubiquitous method allows for cross-species comparison and studies in special populations, such as infants or patients (Hannula, 2010). Fixations are stationary eye movements between saccades, in which most of the visual input occurs. Saccades on the other hand, are fast (10 – 100 ms; Duchowski, 2017), ballistic eye movements that direct gaze to a region of interest (Gilchrist, 2011). Saccades can cover a margin of up to 20° of the visual field to orient attention to another location (Findlay & Gilchrist, 2003). Saccade direction can be an indication of early visual processing, while fixation patterns are an indication of considered features of high information quality useful for ongoing perceptual and cognitive analysis and therefore mirrors preferential attention allocation tendency. The available peripheral information during a fixation can determine the subsequent allocation. Thereby, attentional selection and initiation of the following saccade involves an interplay between visual properties of locations in a given scene and the goal of the observer (Gilchrist, 2011). The quantification of similarities between scan paths within or between observers can offer further insight to visual processing at different time points or when solving tasks. Such comparisons include string-edit distance (e.g., Foulsham & Underwood, 2008), or linear distance algorithms (e.g., Henderson, Brockmole, Castelhano, & Mack, 2007). However, these methods are limited to stimulus setups with high similarity, rendering it difficult to dissociate whether similarities between scan paths result from image properties or other factors, such as top-down information or knowledge as would be the case in real life (Anderson et al., 2013). Therefore, ubiquitous characterization of scan path structures can be accomplished by using recurrence quantification analyses, specifically for temporal characterizations of fixation sequences, revealing potential dependencies between temporal and spatial influences on visual attention, scene understanding and cognition (Anderson et al., 2013). Generally, fixations occurring directly after stimulus onset indicate early, almost reflexive processing (End & Gamer, 2017), while later fixations are subject to top-down influences. Initial saccades upon stimulus presentation allow for investigation of automatic, bottom-up driven mechanisms, usually induced by exogenous cues (Holmqvist et al., 2011) or short presentation times (e.g., Bindemann, Burton, Hooge, Jenkins, & de Haan, 2005; Driver et al., 1999; Rösler, End, & Gamer, 2017).

According to Henderson & Hollingworth (1999), human vision can generally be divided into low-level, intermediate-level and high-level vision. The first refers to the extraction of physical properties of a scene, such as depth, color, texture and edge representations (Marr, 1982, but see Henderson & Hollingworth, 1999). The second selectively or serially extracts shape and locations of aspects that can be determined without extracting their meaning (Ullman, 1996, but see Henderson & Hollingworth, 1999). The third concerns the mapping from visual representations to meaning, underlying cognition and perception (Henderson & Hollingworth, 1999). Early fixations are prone to image locations with certain low-level features such as high contrast, different orientation or colors (Parkhurst, Law, & Niebur, 2002; Tatler, Baddeley, & Gilchrist, 2005).

1.1.2 SALIENCY ALGORITHMS AND PREDICTION FAILURE

Several computational algorithms have been developed to quantify the probability of exogenous attention capture through conspicuous stimulus features, described by the term *physical saliency*. This is in accordance with the feature integration theory (Treisman & Gelade, 1980) and the pre-attentive capture of attention for certain pop-out targets, followed by the *feed-forward model* to combine these features (Koch & Ullman, 1987). The goal of saliency algorithms was not only to define salient features, but to further enable predictions of gaze patterns, as a higher number of fixations is said to correspond to salient image regions (Sharma, 2015), more than would be expected by chance (Foulsham & Underwood, 2008). The basic model of Itti, Koch, & Niebur (1998) acts as the foundation of later models. It subsamples a given stimulus into a Gaussian pyramid and decomposes it into different channels for color, intensity and local orientations, creating *feature maps*. After several normalizations, *conspicuity maps* are built and linearly combined to generate a *saliency map*. These computations describe a winner-takes-it-all principle indicating where attention will be allocated. Borji and Itti (2013) have reviewed and compared multiple algorithms and divided attention models into cognitive models, Bayesian models, decision as well as information theoretic models, graphic models and pattern classification models. These saliency maps have become an integral component of many subsequent models of gaze allocation, suggesting a correlation between low-level features in scenes and fixation selection by humans (for reviews see Borji & Itti, 2013; Tatler, Hayhoe, Land, & Ballard, 2011).

The majority of saliency predictions assume a task-free, stimulus-driven mode of attention employed by the viewer (free-viewing; e.g., Parkhurst et al., 2002) and mainly focus on bottom-up mechanisms of visual attention, yet lack the implementation of task-driven computational principles, although attempts have been made (e.g., Itti & Koch, 2001; Parkhurst, Law, & Niebur, 2002). Other implementations concern modelling prior spatial knowledge of targets, or contextual guidance, which have ameliorated the performance of such models (Torralba, Oliva, Castelhano, & Henderson, 2006). Generally, viewers also depict a tendency to fixate the center of an image, possibly because this location allows for optimal early information processing of the scene, or initiation of oculomotor responses (Tatler, 2007). This *central bias* has also been taken into account in many saliency models (for review see Borji & Itti, 2013). By explicitly asking participants to select the most salient areas, Borji, Sihite, and Itti (2013) found that these selected regions indeed mirrored fixation behavior, indicating a top-down judgement for physical saliency. However, introducing top-down control in the form of tasks, yields different results, as shown by a very early attention experiment by Yarbus (1967). His experiment demonstrated that tasks can have a strong influence on the deployment of attention. In this study, seven different tasks were given to the observer while viewing a complex painting or photograph (e.g., "They Did Not Expect Him" from I.E. Repin (1884): "Estimate the material circumstances of the family in the picture", "Give the ages of the people", "Surmise what the family had been doing before the arrival of the 'unexpected visitor'", "Remember the clothes worn by the people", "Remember the position of people and objects in the room", and "Estimate how long the 'unexpected visitor' had been away from the family"). Herein, salient regions did not draw attention, instead, observers fixated the most informative aspects in order to fulfill task demands (see also Einhäuser, Spain, & Perona, 2008). This demonstrates a top-down component of visual selection and oculomotor function.

In contrast to experiments employing simple, impoverished displays for which feature changes mark a considerable alteration, gaze behavior for naturalistic scenes is not necessarily salience driven. While initial saccades seem to be attracted to salient regions, later ones are more affected by image context during free-viewing (Parkhurst et al., 2002), encoding, or visual search (see Anderson, Ort, Kruijne, Meeter, & Donk, 2015). While low-level features may be good correlates of fixation probability to some extent, they do not necessarily drive attention causally (Einhäuser & König, 2003; Tatler, 2007; but see Einhäuser, Spain, et al., 2008). Anderson and colleagues (2015) suggest that there might be different degrees of top-down involvement during the influence of low-level features on saccadic targeting (*strategic divergence*), others propose an early dissipation of saliency effects replaced by object representations (Donk & Soesman, 2010), while Henderson et al. (2007) imply a hierarchical saccade selection depending on the task at hand and prior knowledge from episodic and semantic memory that is largely independent of saliency. These theories indicate different levels of correlation between saliency and gaze behavior. To quantify the fit between predicted and actually fixated locations, certain measures can be calculated, such as the area under the receiver-operator curve (see Wilming, Betz, Kietzmann, & König, 2011). While these measures are generally above chance-level, image saliency only moderately accounts for the fixation behavior when the viewer is given a task (Foulsham & Underwood, 2008; Henderson et al., 2007), when stimuli are related to personal expertise (Humphrey & Underwood, 2009), or when a human being is present (End & Gamer, 2017). In fact, aspects like text, depicted human beings, or animals are often fixated, even if the low-level saliency in these areas is low (Elazary & Itti, 2008). Einhäuser, Spain, and Perona (2008) ague that fixated locations are better described by object-level information, rather than image saliency. Therefore recently developed models have started to implement task-driven computational principles or high-level features (e.g., sensitivity to objects) to predict human fixations. Particularly deep neural models have received increasing interest for predicting gaze behavior. These models represent algorithms that are designed to recognize certain features through input of actual fixation data recorded from human viewers to predict eye movements for other stimuli, e.g., faces. Others have also started combining low-level saliency models with face detection algorithms to predict viewing behavior for naturalistic scenes, as observers will select faces with a probability of over 80% within their first fixations and elicit similar scan paths when stimuli contain a human being (Cerf, Harel, Einhäuser, & Koch, 2008). This may suggest that eye movements are influenced by higher order scene properties and scene interpretation, or semantics (Einhäuser, Spain, et al., 2008). Generally, it seems that the physical environment is different from that of a social one, as the latter is more complex, less predictable and more responsive to the own behavior (Adolphs, 2001).

1.2 SOCIAL ATTENTION

Social attention is a term that is used to describe general interest in other human beings and what they are attending. One view is that there are specialized processes for social cognition, whereas others think that attention to social features stems from basic components that are not specialized themselves (Adolphs, 2003). Basic concepts of social cognition may be the foundation of higher order processing, such as memory, decision-making, attention, motivation and emotions, that may be recruited in the presence of socially relevant events (Adolphs, 2001).

There are many different research aspects surrounding the term social attention, as human communication entails a multitude messages, including speech, gestures, gaze, or facial expressions. Concerning non-verbal expression, the face represents an exceptional class of stimuli (Kanwisher, McDermott, & Chun, 1997). The perception of a face is usually holistic or configural (Pelphrey et al., 2002), and is seen as a transmitter of expression signals that allows for interpretation of a state of mind that are typically arranged into six basic categories: fear, happiness, sadness, disgust, anger, and surprise (Smith, Cottrell, Gosselin, & Schyns, 2005). These expression signals have low correlations with one another in order for the brain to decode and further decorrelate them for rapid processing, such that we can quickly recognize a smile from a frown (Smith et al., 2005). To correctly interpret a facial expression, we use diagnostic features that are differentially distributed and attended across the face with respect to a certain expression. To systematically investigate the importance of the eyes in characterizing basic emotions and complex mental states, Baron-Cohen, Wheelwright, & Jolliffe (1997) tested adult subjects in their ability to characterize facial expressions through visibility of the whole face, the eyes alone or the mouth alone. Results attested that humans were better at assigning emotions to facial expressions when seeing the eyes alone than seeing the mouth alone, which sparked the idea of the *language of the eyes*. However, further studies refined this finding as such that the eyes seem to be an especially valuable diagnostic feature for recognizing fear, whereas the mouth was more important to categorize happy expressions (Scheller, Büchel, & Gamer, 2012). Furthermore, the facial characterizations were most accurate, when the whole face was visible. Accordingly, a face advantage is found for distinctions of other concurrent stimuli and implies that our attention may be affected by the presence of a face. This notion is supported by the study of Bindemann, Burton, & Jenkins (2005) who had participants discriminate between faces, inverted faces or objects and showed that the presence of an upright face significantly delayed target response times in a central go/no-go task. This *face inversion effect* (for review see Valentine, 1988) disturbs holistic processing, as the configuration of the features changes, leading to a more segmental processing used in object recognition (Pelphrey et al., 2002). De Haas and Schwarzkopf (2018) refine that inversion effects rest on orientation sensitivity for isolated facial regions, as a rotation of features only becomes evident for upright faces, while the detection of an alteration is at chance level for inverted faces ("Thatcher Illusion").

A facial bias is not only evident for overt, but also for covert attention mechanisms: dot-probe paradigms have been modified, presenting a face and an object cue on either side of a central fixation cross, requiring a discrimination response of a probe appearing on one of these respective locations (Bindemann, Burton, Langton, Schweinberger, & Doherty, 2007). Even though the presented cues were of no directional value or of relevance to the task, results depicted faster reaction times if probes appeared on the side of the face. Changes within a scene are also detected more rapidly and accurately for faces than other objects, implying that faces seem to specifically draw attention and may be processed differently than stimuli of less biological significance (Ro, Russell, & Lavie, 2001). Moreover, they may also be adept at retaining visual attention (Bindemann et al., 2007).

Eye gaze is one very important social stimulus to extract relevant information from faces, and following a gaze allows for interpretation of other's attentional state and their potential actions within a given situation. The direction of attention to a certain feature through the indication of gaze by another person results in *joint attention*. This mutual orientation requires knowledge about the perspective of a conspecific (Scaife & Bruner, 1975). The trait to adopt the point of view of another individual is unique among human beings and termed *theory of mind* (ToM; Premack & Woodruff, 1978). Evidence for this can be derived from newborns mimicking facial expressions or the ability to discriminate averted and direct gaze at an age of 4 months (Farroni, Johnson, & Csibra, 2004). Children at four years of age are able to attribute emotions to another person's facial expression from the direction of their gaze (Baron-Cohen & Cross, 1992). Gaze cues can be perceived both overtly, through eye movements, as well as covertly through reflexive shifts of spatial attention (Frischen, Bayliss, & Tipper, 2007; but see Ricciardelli, Carcagno, Vallar, & Bricolo, 2013). A typical gaze-following paradigm will present a picture or schematic of a face with eyes looking either to the left or the right side of the screen, drawing attention to that location, similar to arrow cues in dot-probe paradigms. A subsequent target will hence be detected faster at the gazed at location, but slower at the opposite location, respective of the attentional focus. Gaze following seems to be intuitive for human beings, as these effects occur very rapidly (less than 100 ms after the appearance of the cue; see Friesen & Kingstone, 1998; Frischen, Bayliss, & Tipper, 2007, but see Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012). Furthermore, it occurs even when the direction of the gaze is counterpredictive of the following target (Driver et al., 1999; Chris Kelland Friesen, Ristic, & Kingstone, 2004; but see Ristic & Kingstone, 2005). Ricciardelli and colleagues (2013) concluded that gaze following is likely to be a product of both automatic and goal-driven orienting mechanisms, as joint attention does not simply involve following someone's gaze, but rather represents an orienting to a common object of interest. Mutual gaze can modulate joint attention (e.g., Böckler, Knoblich, & Sebanz, 2011), suggesting that an individual is not inflexible to automatic orienting, but is susceptible to context and relevance of a given gaze cue. Specifically, if a gaze is directed to an empty or non-matching space, gaze following does not occur even with strong directional eye movements (Ricciardelli et al., 2013). The extent of attending a face or a gaze elicited by one, also depends on the emotional valence. Affective facial expressions, will capture attention more readily and induce stronger reactions than neutral expressions, in terms of reaction times (e.g., Eastwood, Smilek, & Merikle, 2001), electrophysiological early sensory components of event-related brain potentials (e.g., Krolak-Salmon, Fischer, Vighetto, &

Mauguière, 2001), as well as neural activity (e.g., Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001; Vuilleumier, Armony, Driver, & Dolan, 2001). This effect suggests that people are susceptible to emotional meaning and are unable to ignore emotionally valent stimuli (Vuilleumier, 2005).

1.2.1 NATURALISTIC SCENES

Cueing paradigms, however, may not pick up on basic differences in the social relevance of competing stimuli and may not reflect natural behavior (Kingstone, Smilek, Ristic, Friesen, & Eastwood, 2002). Returning to the comparison of directional cues, Birmingham, Bischof, & Kingstone (2009) presented gaze and arrows within complex stimuli while allowing free viewing of the scene. Their results not only confirmed that participants tended to fixate eyes more than arrows, but also that this tendency prevailed even when the size of the arrows was altered to the detriment of eyes, and also when the saliency (Itti et al., 1998) was controlled for. Moreover, as the study used complex naturalistic scenes with holistic features instead of faces in isolation, they also observed that heads and eyes were equally prioritized and always fixated first in comparison to the arrows. This raises another important aspect in attention research: presenting stimuli in isolation, may bypass critical aspects of attention and the inherent selection processes, whereby the sense of the relative importance placed on the stimulus is lost (Kingstone et al., 2002). Furthermore, contextual changes may account for further distortion of the selection process influencing eye-movements through scene gist (e.g., Oliva & Torralba, 2006) or objects (e.g., Einhäuser, Spain, & Perona, 2008), which are often diminished in impoverished visual displays (see also Anderson et al., 2015). In fact, while research has focused largely on the detection of local features (i.e., saliency maps), global structures are processed prior to local ones for real-world scenes (Kimchi, 1992). This is in line with the fact that a region of the brain (parahippocampal place area; PPA) is sensitive to intact or holistic scenes, but not single objects alone (Epstein & Kanwisher, 1998; Epstein, Stanley, Harris, & Kanwisher, 2000). When presenting a face without the respective body, the eyes have a high contrast in relation to the other features (Kobayashi & Kohshima, 1997), which may lead to a bias to fixate this region founded by saliency. Therefore, it is important to investigate social cues embedded in a natural surrounding within the context of competing objects. Obviously, less controlled environments may cause uncontrollability of variables, however, several recent studies have shown that components of attention can be successfully studied using naturalistic scenes: herewith, the notion of preferential attention towards eyes could be manifested (Birmingham, Bischof, & Kingstone, 2008b; Smilek, Birmingham, Cameron, Bischof, & Kingstone, 2006) and investigated within different contexts, such as that of deception during magical tricks (Kuhn &

Land, 2006), as well as in different social settings, where the tendency to fixate the eyes increased with higher social content, especially when the subjects interacted with one another (Birmingham et al., 2008b). By addressing the influence of saliency in these complex scenes, one may conclude that a bias for eyes is selected due to the social meaning they convey and not as a result of low-level saliency (Birmingham, Bischof, & Kingstone, 2009b; Ro et al., 2001). In fact, Cerf, Frady, & Koch (2008) showed that a model that considered saliency with face-detection, best described fixations within natural social scenes, providing support for a specialized "face" channel in the visual system (Viola & Jones, 2001). Furthermore, there seems to be a very early onset of this bias as it emerges within the first fixations after stimulus onset (End & Gamer, 2017; Henderson et al., 2007), the time at which saliency is said to be most influential (see Tatler et al., 2005). Social aspects are in fact reflexively selected when presented for a very brief amount of time (< 200 ms) and thereby circumvent top-down processes (for review see Carrasco, 2011; but also Rösler et al., 2017). Automatic processes are said to be rapid, non-conscious, mandatory and capacity-free (for review see Palermo & Rhodes, 2007). Cerf and colleagues (2008) conclude, that faces attract attention and gaze relatively independent of a given task and should therefore be considered as part of the bottom-up saliency pathway. However, a bias towards social features is not only evident for initial eye movements, but also across sustained viewing durations (End & Gamer, 2017). As bottom-up and top-down control act on different timescales (Anderson et al., 2015), a sustained bias for human stimuli indicates that there is also a top-down component to social attention.

1.3 NEURAL MECHANISMS OF SOCIAL ATTENTION

Visual information enters the retina and project onto the optic chiasm and further reaches three major subcortical targets: the pretectum, the superior colliculus, and the lateral geniculate nucleus, which pass on information to the primary visual cortex (V1) onto higher cortical regions (Wurtz & Kandel, 2000). The receptive fields of neurons can be categorized into separate functional modules and depict a hierarchical organization of visual information processing, from simple (low-level) to more complex, detailed and selective (high-level) features, such as objects (Ishai, Ungerleider, & Haxby, 2000), human faces (McCarthy, Puce, Gore, & Allison, 1997), as well as their bodies (Downing, 2001).

Considering previous findings, it is imaginable that the special status of social aspects may be processed differently or with higher priority than general physical aspects in the human brain. In order to adapt and react to a given social situation, neural systems have to process the perception of social signals and connect these to a behavioral response, guiding both automatic and voluntary responses (Adolphs, 2001). While perceptual representations of stimuli are related to higher-order sensory cortices, emotional responses and behavioral motivation are associated with the amygdala, striatum and orbitofrontal cortex, and aggregated internal models involve higher cortical regions (Adolphs, 2003).

So how does social attention fit into the dichotomization of bottom-up and top-down mechanisms? Before answering this question, neural correlates of general attentional mechanisms need clarification. Several approaches have yielded different results: Neuropsychological findings differentiate between three neuronal networks involved in directing attention: the orienting (posterior), the detecting (anterior), and the alerting system (Posner & Petersen, 1990). The first prioritizes sensory input by selecting modality and location, the second detects relevant stimuli and modulates awareness for attentional capture of these, and the third prepares and sustains alertness to process high-priority signals (Kaspar & König, 2012). Others consider four different sub-systems, responsible for initial attentional orientation, followed by mediation of selective attention (biased competition; Coull, 1998), a division of attentional resources (covert and overt; Kramer, 1998) and finally sustenance of attention (Müller, Malinowski, Gruber, & Hillyard, 2003). Corbetta and Shulman (2002) defined a dorsal network for selective attention, associated with top-down processes, and a ventral network for behaviorally relevant stimuli occurring in unattended channels or locations subserving bottom-up processes. The dorsal network comprises the posterior parietal and frontal cortex, while the ventral network is lateralized to the right hemisphere and is centered on the temporo-parietal and ventral frontal cortex. However, orienting to sensory stimuli is modulated by both and this interaction is central to theories of visual attention (Corbetta & Shulman, 2002). The extent of this interaction is partly influenced by the behavioral relevance of the stimuli. While the dorsal pathway engages in preparing and applying goal-directed selection for stimuli and responses, the ventral system is involved in the detection of behaviorally salient, or unexpected stimuli and may act as a ‘circuit breaker’ for the dorsal system (Corbetta, Patel, & Shulman, 2008). Herein, a key function of this right ventral fronto-parietal network is to direct attention to sensory stimuli that are outside the focus of processing. However, a coupling between these networks would result in ‘switch costs’ (Corbetta & Shulman, 2002) and require relatively long processing time that is insufficient to explain reflexive orienting for social stimuli (Rösler et al., 2017). Therefore, this “cortico-centric” view has been put into question concerning social attention mechanisms. Gaze following studies provide another confound: the intraparietal sulcus (IPS) of the dorsal network and the superior temporal sulcus (STS) of the ventral network, are both involved in gaze direction processing and gaze shifts (Pelphrey, Singerman, Allison, & McCarthy, 2003; but see Ricciardelli et al., 2013).

Although visual information is primarily processed by a succession of visual areas (V1, V2, V3, V4, V5), some projections also reach the amygdala via the superior colliculus and the pulvinar (Tamietto, Pullens, Gelder, Weiskrantz, & Goebel, 2012). These brain regions are subcortical structures that have been associated with the interpretation of social scenes and facial expressions (Adolphs et al., 2005). On the other hand, cortical structures like the temporo-parietal junction (TPJ), which is part of the ventral network, is also involved in directing attention towards social features (Nardo, Santangelo, & Macaluso, 2011), particularly in association with the reasoning about the contents of another person's mind (Saxe & Kanwisher, 2003). The inferior temporal cortex (IT), which is considered the final stage of the ventral cortical visual system (Gross, 1992), is also associated with processing gaze direction (Langton, Watt, & Bruce, 2000). The previously mentioned study (1.1.2 Saliency Algorithms and Prediction Failure) of Ristic and Kingstone (2005) examined if social orienting elicited by gaze cueing was purely bottom-up driven or whether it could be modulated by top-down manipulations. The authors found that reflexive social attention could be influenced by top-down mechanisms, but only in an asymmetrical manner, such that they were effective for triggering IT involvement if the ambiguous stimulus was first perceived as a car, but not when IT involvement had already been triggered through the interpretation of a face. Similarly, Dolan and colleagues (1997) only observed IT activation when a stimulus was recognized as a face. In an attempt to combine these approaches, researchers have suggested that cortical and subcortical activations for social scenes (1) are in opposition to one another with a possibility of override (Frith & Frith, 2008), (2) have distinct temporal components of social information processing, or (3) may be considered as dynamically iterative (Adolphs, 2010; Cunningham, Zelazo, Packer, & Van Bavel, 2007).

Neuroimaging studies have investigated how conceptual knowledge about people is organized in the brain by introducing tasks or altering stimuli of social nature (Mitchell, Heatherton, & Macrae, 2002). Results indicate that people and objects elicit functionally dissociable neural substrates, including brain regions associated with social-cognitive functioning (medial prefrontal cortex, superior temporal cortex, intraparietal sulcus, and fusiform gyrus)(Mitchell et al., 2002). Especially the fusiform gyrus has shown distinct activations for faces (e.g. Kanwisher, McDermott, & Chun, 1997), so robustly that this region is now termed the *fusiform face area* (FFA; Kanwisher & Yovel, 2006). Neurons in this area seem to be selective for faces and their corresponding features, as the response was diminished for objects like houses or even other human body parts, such as hands (George, Driver, & Dolan, 2001). Furthermore, dissociable neural systems within the right anterior STS have been found for different directions of gaze, supporting the componential nature of face perception with different functional and neural mechanisms (Calder et al., 2007; Haxby, Hoffman, & Gobbini, 2000). Other brain regions may also be involved in computing gaze direction are the parietal cortex (Haxby et al., 2000) and medial prefrontal cortex (Calder et al., 2002), as well as the amygdala (Kawashima et al., 1999).

Generally, subcortical structures (amygdala, pulvinar, superior colliculus) are sensitive to coarse visual information of low spatial frequency and may be involved in the perception of emotional faces (Frischen, Eastwood, & Smilek, 2008). Moreover, research groups of Whalen (1998) and Vuilleumier (2001; 2002) support the notion that faces expressing certain emotions, are rapidly processed in the brain, independent of where attention is deployed, or whether the face is consciously perceived. Especially fearful faces elicit robust activation of the amygdala, even when they are unattended or irrelevant to a postulated task, suggesting that the amygdala constitutes a dedicated system for processing emotionally-relevant events (Vuilleumier et al., 2002). Even earlier neuronal responses of the amygdala have been found for positive facial expressions (Liu, Ioannides, & Streit, 1999). In contrast to the previously suggested involvement of the TPJ for orienting attention towards social stimuli, activation in this area only occurs after more than 300 ms (Menon, Ford, Lim, Glover, & Pfefferbaum, 1997), whereas saccades towards social features are already evident after 100-150 ms of showing the stimulus (Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Rösler et al., 2017). Furthermore, studies like those of Morris and colleagues (1998) suggest that projections from the amygdala to occipital cortex may lead to the involvement of enhanced visual processing of emotional stimuli, leading to pre-attentive detection and only subsequent attention allocation. In line with this, Gamer & Büchel (2009) could prove a direct role of the amygdala in reflexive gaze initiation towards fearfully widened eyes. An involvement of the amygdala with the integration of information about space and value has also been implied through a stimulus-reward task, in which subjects showed increased amygdala activation and faster reaction times for trials in which a reward was dependent on a spatial specific response as compared to trials in which spatial specificity was less required (Ousdal et al., 2014). This also indicates that the amygdala is not simply involved in emotion discrimination, but also in the detection of salient information, acting as a 'tag' for relevant features (Adolphs & Spezio, 2006; Gamer & Büchel, 2009). There are however, researchers who argue against a rapid, automatic and partially unconscious processing of emotional and social stimuli through a subcortical pathway, instead viewing the cortex as the key player with the aid through contributions of the amygdala and the pulvinar through their extensive connectivity (for review see Pessoa & Adolphs, 2010).

1.5 OBJECTIVE

Our goal was to investigate basic mechanisms of social attention with a variety of different paradigms. Initial findings of End & Gamer (2017) depicted that attention was preferentially deployed to social features during free-viewing and that this allocation was independent of saliency. To extend this line of research, we conducted several studies to disentangle separate processes and assess the influence of bottom-up and top-down mechanisms on social attention. We aimed to uncover whether social attention may be part of a bottom-up or top-down driven control, or if it is in fact a process that cannot be described by the suggested dichotomy. Further examination was dedicated to potential differences between overt and covert attention mechanisms and potential differences or similarities of processing social information. Throughout our experiments, we aimed to target the transference from simple paradigms to more realistic setups and detect potential discrepancies. Additionally, we considered physical properties in the form of controlled saliency distributions to support our conclusions. Indications of aberrant processing of social aspects compared to e.g., general attention would imply distinct mechanisms with possible implications concerning neural processes and higher cognition. The multitude of literature on higher social processing is complex and difficult to fathom without the foundation of understanding basic underlying functions. Segregating single processes of social attention allows for profound deductions and implications unique from those of other mechanisms. This understanding may offer greater insight into social processing, higher social functioning and further transfer to alterations in social cognition and behavior with regard to patient populations.

2. PUBLICATIONS

Coherent with our objectives, three different publications will be discussed in the following, each representing experiments conducted to disentangle specific mechanisms of social attention. The content of the studies corresponds to the published form in the respective research journals, merely the format and the figure and table captions were edited to suite this dissertation.

The studies will be presented in the following order:

Study 1 – Flechsenhar, A. F., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PloS one*, *12*(8), e0183799.

http://dx.doi.org/10.1371/journal.pone.0183799

Study 2 – Flechsenhar, A., Larson, O., End, A., & Gamer, M. (2018). Investigating Overt and Covert Shifts of Attention within Social Naturalistic Scenes. *Journal of Vision*, 18(12):11.
http://dx.doi.org/10.1167/18.12.11

Study 3 - Flechsenhar, A., Rösler, L., & Gamer, M. (2018). Attentional Selection of Social Features Persists Despite Restricted Bottom-Up Information and Affects Temporal Viewing Dynamics. *Scientific reports*, 8(1), 12555.

http://dx.doi.org/10.1038/s41598-018-30736-8

After each publication, the relevance of the current findings are discussed and implications made to introduce the subsequent research question.

TOP-DOWN INFLUENCE ON GAZE PATTERNS IN THE PRESENCE OF SOCIAL FEATURES

ALEYA F. FLECHSENHAR, MATTHIAS GAMER

JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG

Corresponding author: Aleya F. Flechsenhar, Department of Psychology, Julius Maximilian University Wuerzburg, Marcusstrasse 9-11, 97070 Würzburg, Germany.

E-mail address: aleya.flechsenhar@uni-wuerzburg.de

Published 2017 in PLoS ONE 12(8): e0183799

ABSTRACT

Visual saliency maps reflecting locations that stand out from the background in terms of their low-level physical features have proven to be very useful for empirical research on attentional exploration and reliably predict gaze behavior. In the present study we tested these predictions for socially relevant stimuli occurring in naturalistic scenes using eye tracking. We hypothesized that social features (i.e. human faces or bodies) would be processed preferentially over non-social features (i.e. objects, animals) regardless of their low-level saliency. To challenge this notion, we included three tasks that deliberately addressed non- social attributes. In agreement with our hypothesis, social information, especially heads, was preferentially attended compared to highly salient image regions across all tasks. Social information was never required to solve a task but was regarded nevertheless. More so, after completing the task requirements, viewing behavior reverted back to that of free-viewing with heavy prioritization of social features. Additionally, initial eye movements reflecting potentially automatic shifts of attention, were predominantly directed towards heads irrespective of top-down task demands. On these grounds, we suggest that social stimuli may provide exclusive access to the priority map, enabling social attention to override reflexive and controlled attentional processes. Furthermore, our results challenge the generalizability of saliency-based attention models.

INTRODUCTION

According to traditional models, visual attention is guided by both cognitive endogenous (top-down) factors, such as knowledge, expectation and current goals, and exogenous (bottom-up) factors that reflect sensory stimulation. This insight is based on numerous experimental laboratory paradigms using simple stimuli to disentangle and identify mechanisms underlying attention control (for reviews see Knudsen, 2007; Theeuwes, 2010). However, the conclusions of these setups are not necessarily transferable to the real world, where sensory signals continuously compete for the brain's limited processing resources and stimuli and responses are inextricably linked. As such, prominent aspects in our environment are not only important due to their features, but also through their behavioral relevance. In turn, distinctive sensory stimuli attract attention more effectively when they are relevant or contingent to the task at hand (Folk et al., 1992). A large body of research that was devoted to predicting gaze behavior, relied on the calculation of so-called saliency maps, which filter early features, such as orientation, contrast intensity and color. For example, the prominent, graph-based saliency (GBVS) algorithm by Harel, Koch & Perona (2006) acts by decomposing the image into a series of feature maps and generating activation maps on certain feature channels based on graph theory. These maps are then combined into a single saliency map revealing locations that stand out in terms of their low-level features from the background. Accordingly, attention should be allocated to locations in the scene depending on the saliency in the computed map using a winner-takes-it-all mechanism. These maps have become an integral component of many subsequent models of gaze allocation, suggesting a correlation between low-level features in scenes and fixation selection by humans (for reviews see Borji & Itti, 2013; Tatler, Hayhoe, Land, & Ballard, 2011). However, they do not necessarily drive attention causally, but contingent on higher-order statistics (Nuthmann & Einhäuser, 2015) as visual saliency seems to provide a poor account of eye fixation patterns in complex visual scenes Cerf et al., 2008; Henderson et al., 2007 and lacks prediction of overt spatial orienting for long exposure (Elazary & Itti, 2008), where intrinsic and strategic aspects are thought to be prominent. Other studies concentrating on social aspects of visual attention, have shown that saliency models were also inaccurate in describing fixation selection for socially relevant stimuli, such as human faces (Scheller et al., 2012), as well as scenes with whole bodies (Birmingham et al., 2009b; End & Gamer, 2017; Fletcher-Watson et al., 2008), showing a gaze behavior that was uncorrelated to low-level image statistics even from the first saccade onwards. These findings indicate that social stimuli may engage special perceptual processing and provide exclusive access to the priority map, enabling social attention to override reflexive and controlled attentional processes. To explicitly disentangle to what degree this social override occurs, we conducted an eye tracking study introducing top-down demands in form of tasks with different complexity to non- social aspects of the stimuli, rendering the social aspects to be uninformative for the viewer. Most findings concerning preferential viewing of social information were conducted under free-viewing conditions or have introduced tasks encouraging fixations onto social information (Birmingham, Bischof, & Kingstone, 2008a; Fletcher-Watson et al., 2008). To our knowledge, the influence of varying degrees of top-down demands that explicitly require the direction of attention away from the social content has not yet been investigated. Early studies by Yarbus (1967) already suggest that the selection of gaze information may depend on the task that is assigned to participants and the social content of the scene. A replication study of DeAngelus and Pelz (2009) using modern eye tracking, confirmed these results, revealing different scanpath patterns for distinct tasks. We hypothesized that attention will be devoted to social scene elements due to their behavioral relevance for human beings even when they do not need to be scanned to accomplish the experimental task and this preference may already be evident in very early fixations.

MATERIALS AND METHODS

PARTICIPANTS

The study was approved by the ethics committee of the German Psychological Society (DGPs) and conducted according to the principles expressed by the Declaration of Helsinki. All participants provided written informed consent and received payment for their participation. Forty five subjects voluntarily took part in the experiment. Five were excluded due to missing responses in more than 25% of all trials or low eye tracking data quality (>20% missing base- line values or baseline outliers, see below), respectively. This sample size allowed us to detect medium effect sizes (Cohen's $f = 0.25$) for fixation differences between the four experimental tasks (see below) with a power of at least 95% when assuming a correlation of $r = .50$ between factor levels. The age of the final set of participants (21 women) ranged from 18–35 years ($M = 24.45$ years, $SD = 4.15$ years) and consisted of 31 students from various disciplines, as well as 9 employed subjects. All participants had normal or corrected-to-normal vision with the option of wearing contact lenses. Women were asked to refrain from using eye make-up. Three participants were left-handed (all women). The average amount of school years was no lower than a General Certificate of Secondary Education with a minimum duration of 10 school years ($M = 12.33$ years, $SD = 0.92$ years) and did not differ between men and women ($t(38) = 1.04, p = .304$).

A total of 120 naturalistic stimuli showing negative, neutral and positive scenes, were used for the experiment. Half of them included human beings (referred to as "social images" in the current article) and the other half did not depict human beings, but showed landscapes, objects, animals, etc. (referred to as "non-social images" here). Four different tasks were included with tasks 2 – 4 specifically addressing non-social attributes: (1) free viewing condition, in which participants were allowed to look wherever they pleased, (2) definition, which entailed naming the color of a four-wheeled vehicle in the picture (e.g. a car, a bus, a truck), (3) counting, which required determining the number of blue objects in the picture and 4) estimation, where the percentage of white in the picture had to be estimated. We decided to use these additional three different tasks instead of merely a single one to examine potential influences of complexity on social attentional mechanisms. Furthermore, we chose tasks that also aimed at global (estimation) as compared to local (definition, counting) scanning to investigate differences herein. Therefore, harder tasks requiring higher top-down employment may interfere more with attention towards social features than easier tasks and tasks demanding local gaze distributions may affect social attention more than those encouraging global scanning. One of these tasks had to be completed during the total presentation time of 10s for every stimulus. Participants were explicitly told that the blue objects they had to count were non-social and not part of human beings (e.g., no clothes). Notably, not all objects within a stimulus were always uniform, but were chosen as such that they could be clearly classified as individual objects. Also, no vehicle was ever occupied to ensure the non-social aspect of the task. Furthermore, stimuli were chosen as such that the number of humans depicted in the scene varied, as well as the aspect of high saliency of faces and bodies. Moreover, depicted subjects were distributed across the whole scene to reduce a central bias and their size varied. All stimuli were applicable to all tasks (with the exception of the color definition tasks, which was limited to 36 pictures) and randomized for each task and subject to avoid stimulus-specific effects. Thus, in total, each subject accomplished each task for 15 social and 15 non-social images, respectively.

STIMULI AND TASKS

Stimuli were presented on a 24" LG 24MB65PY-B screen (516.9 x 323.1 mm) with a resolution of 1920 x 1200 pixels and a refresh rate of 60 Hz. The viewing distance was 50 cm for each participant to view stimuli of 1200 x 900 pixels, resulting in a visual angle of 35.81° x 27.24°. The stimuli showed real-life situations that were either photographed or chosen from the internet (e.g. Google picture search, Flickr) and had to fulfill the requirements of all four tasks to be used randomly for different participants. Some pictures were modified to avoid writings or labels and adjusted in luminance and contrast to better match the rest of the stimulus set. Further, their resolution had to be sufficient in quality to rescale them accordingly. Image editing was performed with the software GIMP (version 2.8.16; GNU Image Manipulation Program, The GIMP Team).

Each trial began with the written command (black letters on a uniform grey background) defining the task for the subsequent stimulus, shown for 2s. This was followed by a fixation cross lasting for 2s. Afterwards, the stimulus was presented for a total of 10s, during which the participant had to gather an answer to the task and indicate this by clicking the left mouse button. This was used as a reaction time measure. Even after the click, the stimulus would stay on screen until the 10s passed. During this time participants were free to look wherever they pleased. At stimulus offset, an answer scale would appear on screen, requiring participants to choose one of eleven possible options for each task (defining: color palette, counting: from ≤ 5 to ≥ 15 in steps of 1; estimating: between 0% and 100% in steps of 10%) with the exception of the free-viewing condition, which depicted an OK button that had to be clicked to continue with the next trial. Inter-trial-intervals (ITI) amounted to 3-7s randomly drawn from a uniform distribution. A uniform grey screen was shown during the ITI. Eye tracking data were recorded during the tasks with a sampling rate of

1000 Hz using a tower mounted EyeLink 1000 plus system (SR Research Ltd., Ottawa, Canada) with a 25mm lens.

PROCEDURE

Upon arrival to the laboratory, participants were informed about the experiment, were asked to sign the consent form and then completed a short questionnaire concerning sociodemographic data (age, sex, profession and handedness). Afterwards, instructions concerning the tasks were given verbally and by means of printed examples, as well as for the eye tracking. To become acquainted with the procedure, 8 training trials were conducted with a separate set of pictures including every task twice. The fixation cross was to be fixated during the whole duration of its presence. During stimulus presentation, reading of task instructions, times for which the screen was blank or the time after their response, participants should feel free to change their gaze and look wherever they pleased. However, blinks during the recording time of the stimulus presentation should be avoided. Starting the experiment, the eye tracking system was calibrated using nine points and subsequently validated, followed by the actual task. Stimulus and task order was randomized across participants and the experiment was divided into four blocks to ensure small breaks in between and opportunities to rest the eyes. After completion of the experiment, participants were asked to rate the perceived difficulty of the tasks they had completed.

The experiment was programmed with the Experiment Builder Software (version 1.10.1630; SR Research Ltd., Ottawa, Canada) and data processing and analysis was performed using the open-source statistical programming language R (www.r-project.org) and Matlab® R2011b (Mathworks, Inc., Natick, MA, USA). A univariate analysis approach as implemented in the *ez* package (version 4.3; Lawrence, 2016) was used for all repeated-measures analyses of variance (ANOVAs). The a-priori significance level was set to $\alpha = .05$ for all statistical tests and general η^2 is reported as an effect size index. Huynd-Feldt's ε is reported for all repeated-measures ANOVAs containing more than one degree of freedom in the numerator to account for potential violations of the sphericity assumption. Cohen's *d* is reported as an effect size estimate for two-sample and paired *t*-tests (Cohen, 1988).

DATA PROCESSING AND ANALYSIS

BEHAVIORAL DATA

Reaction times were calculated as the difference between stimulus onset and the first mouse click during stimulus presentation. If no click was given, the reaction time was set to the total presentation time of 10s. Differences in reaction times were analyzed as a function of stimulus content (social vs. non-social) and task (defining, counting, estimating) using a 2 x 3 repeated-measures ANOVA. Complexity of each task was rated by participants after completion of the experiment on a scale of 1 (very easy) to 6 (very hard). Differences in difficulty ratings were analyzed in a one-way ANOVA with the factor task.

EYE TRACKING DATA

For the analysis of the eye-tracking data, we first calculated visual saliency maps. A large number of different algorithms have been suggested for such purpose (Borji & Itti, 2013). We decided to primarily rely on the graph-based visual saliency algorithm (GBVS) by Harel, Koch & Perona (2006). It uses graph algorithms for saliency computations, forming activation maps on certain feature channels and then normalizing them to highlight conspicuity, admitting combinations with other maps. The GBVS algorithm is available as Matlab® source code, it is applicable without initial training and it performs well in predicting fixations in free viewing conditions (Judd, Ehinger, Durand, & Torralba, 2009; Sharma, 2015). It is important to note that the currently reported results do not strictly depend on this saliency algorithm since similar results were obtained for the algorithm by Itti and colleagues (Itti et al., 1998). The GBVS algorithm generates maps that range between 0 to 1 and depict the distribution of visual saliency across the image. The current set of social and non-social scenes was comparable regarding mean saliency ($t_{(118)} = 0.086$, $p = .93$, $d = .016$, average mean for social scenes: .25, average mean for non-social scenes: .25) and saliency variation ($t_{(118)} = 0.031$, $p = .98$, $d = .006$; average *SD* for social scenes: .20, average *SD* for non-social scenes: .20). Saliency maps were used for determining image regions with higher as compared to lower visual saliency (see below).

In a second step, eye movements were segmented into saccades and fixations using velocity and acceleration thresholds of 30°/s and 8000°/s^2, respectively, for saccade detection. Time intervals between saccades were defined as fixations. For all eye movement measures, fixations were drift corrected with reference to a baseline period of 300 ms before stimulus onset (i.e. when the central cross was fixated). Outliers of baseline coordinates were identified using a recursive outlier removal procedure that was applied separately to x- and y-baseline-coordinates. For each participant the highest and lowest baseline coordinates were temporarily removed and the mean and standard deviation were calculated for the remaining data. If either of the two values fell outside an interval bounded by 3 standard deviations from the mean, it was removed completely. If the data points fell within the interval, they were returned to the data set. This procedure was continued until no more data points were discarded. Trials with invalid baseline position data were replaced by the means of all valid baseline positions, including a removed x or y baseline coordinate or missing baseline data (proportion of social and non-social scene trials: M = 5.54 %, SD = 5.68 % and M = 4.96 %, SD = 4.44 %, respectively). For further analyses, trials containing too many blinks were excluded (trials with a blink-free time period of less than 80% of the whole trial: M = 13.0 %, SD = 0.07 %) and fixations were drift corrected and then used to create fixation density maps. The first fixation was not considered since it usually overlapped from the fixation period before stimulus onset. An empty two-dimensional map (1200 x 900 pixels) was generated for each participant and stimulus. The respective fixations were weighted by their fixation durations in milliseconds, which were added at the pixel position of the fixation. The resulting map was then smoothed with a two-dimensional isotropic Gaussian kernel with a standard deviation of 36 pixels or 1° of visual angle using the R package *spatstat* (version 1.45.0; Baddeley, A., Rubak, E. & Turner, 2005). The total smoothing kernel amounted to 2° of visual angle (one standard deviation in positive and one in negative direction) to resemble the functional field of the human fovea centralis. The fixation maps were then normalized to range from 0 to 1. Eye movements up to the point of the mouse click were analyzed, to investigate task-relevant fixations, as well as the period after the mouse click to examine potential changes in viewing behavior after completion of the task.

In a third step, we introduced regions of interest (ROIs) to investigate the distribution of fixations onto the social and non-social features of the stimuli. Specifically, pixel coordinates were defined for head, body and areas with lower and higher saliency to all stimuli containing social information (n = 60) and areas with lower and higher saliency to all non-social stimuli (n = 60). Head and body ROIs were manually drawn in GIMP and each ROI pixel was assigned a certain color. In order to define the ROIs for saliency, the saliency maps for social scenes were considered for those image regions, which had not already been assigned to the head or body ROI, while the whole scene was considered for non-social stimuli. Saliency values smaller or equal to the eighth percentile of the saliency distribution were defined as areas of lower saliency. Although the criterion was arbitrary, this cut-off allowed identification of image regions which were highly salient, but contained no social information, which deemed essential to disentangle potential effects of attentional allocation to social versus physically salient information. To ensure that social features did not represent the most salient image regions, we calculated the mean saliency of each ROI for each social scene and divided the resulting values by the mean saliency of the whole scene. Further, we calculated a one-way ANOVA with the factor ROI to confirm that social features were less salient than areas of high saliency (main effect of ROI: $F_{(3,117)} = 113.18$, $\varepsilon = .53$, $p < .001$, $\eta^2 = .52$). Additionally, we determined the number of ROIs per category (head, body, low saliency, high saliency) per image across all social images. These values are also displayed in Figure 1.1. Importantly, the numbers of the most relevant social ROI (i.e., heads, $M = 5.78$, $SD = 5.95$) and non-social ROI (i.e., highly salient regions, $M = 5.35$, $SD = 3.60$) did not differ significantly ($t_{(59)} = 0.59$, $p = .56$, $d = .004$).

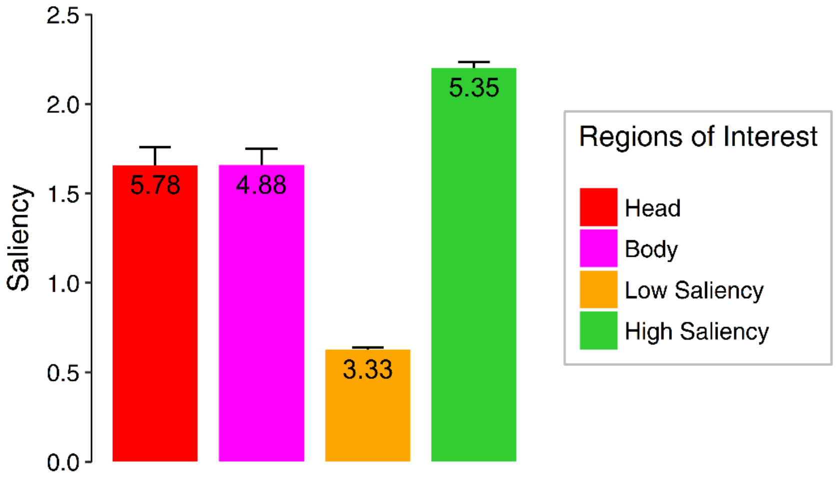

Figure 1.1. The graph depicts the mean saliency values for the different regions of interest (head, body, areas of low and high saliency) for all social images according to the graph-based visual saliency algorithm (Harel et al., 2007). The values depicted within the bars describe the mean number of ROIs per category and image across all social stimuli. Error bars represent standard errors of the mean.

The defined ROIs and the fixation density maps were used to determine the relative extent to which each ROI was fixated. For this, the sum of fixation density values was calculated for each ROI and then divided by the sum of fixation density values for the whole scene. An example stimulus with respective ROIs, saliency and fixation density maps is depicted in Figure 1.2 (these images were adapted from End & Gamer (2017) whose analysis was followed closely in this study). The proportion score was then normalized by taking the size of the ROI into account to control for increased fixations onto larger rather than smaller areas (see also Birmingham et al., 2009b). The mean for this relative area-normalized sum of fixation density was calculated for each ROI across all social scenes for each participant as a function of time window (i.e., before acknowledging task completion and after the mouse click). To analyze fixation densities as a function of the experimental manipulations, several ANOVAs were calculated: A 4 x 4 repeated-measures ANOVA with the factors task (free-viewing, color definition, object counting and estimation) and ROI (head, body, areas of low saliency and high saliency) was run for the task-relevant time window (for the free-viewing condition, this amounted to the whole presentation duration). Task-relevant and task-irrelevant time windows were compared by a 2 x 3 x 4 repeated-measures ANOVA with factors time window, task (now excluding the free viewing task) and ROI. Similar analyses were conducted for the non-social stimuli, using a 4 x 2 repeated-measures ANOVA for the task-relevant time window with the factors task and ROI (low and high saliency) and a 2 x 3 x 2 repeated-measures ANOVA with the factors time window, task (again excluding the free viewing task) and ROI. Additionally, since we differentiated the background of each scene into areas of low and high saliency, we conducted a further analysis in which we compared areas of low and high saliency also within social ROIs using the same criteria for definition as for the background. Thus, head and body ROIs were dissociated into regions with low saliency (saliency values less or equal than the eighth percentile of the saliency distribution) and high saliency (remaining areas). Herewith, we could examine fixation densities across different regions in the scene (head, body, background) depending on the saliency distribution (low and high) for each task. This was accomplished using a 4 x 3 x 2 ANOVA with repeated measures using the factors task, regions and saliency.

Fourth, we carefully examined the initial fixation changes towards ROIs to reinforce our findings, but also to gain more insight into the processing speed and potential automatic attention-grabbing effect of social stimuli. For each participant, the relative frequency that each ROI was fixated across all social scenes was determined by dividing the frequency that each ROI was fixated by the frequency that any ROI was fixated for each of the first three fixations after stimulus onset. The relative frequency scores were normalized by considering the mean area of the respective ROI across all social scenes when represented in the according relative frequency score. Social scenes, were analyzed by a 4 x 4 x 3 repeated-measures ANOVA with the factors task, ROI and fixation number. To directly compare initial fixations between both social and non-social images, we conducted a 2 x 4 x 2 x 3 ANOVA incorporating the additional factor stimulus category and comparing ROIs for areas of low and high saliency (factors stimulus category, task, ROI and fixation).

It is of note that the trials of which baseline position data were replaced by mean baseline position data for drift correction were removed from the analyses of the first fixations, as participants may not have fixated the central cross directly before the onset of the scene in these trials. Hence, the starting position may differ between these trials as compared to trials in which the cross was fixated. The construction of fixation density maps would not be influenced by this disparity, since it comprises the fixations of the entire viewing duration of the scene. However, it may influence the locations of fixations occurring very early after the onset of a scene which would specifically effect analyses of the first three fixations.

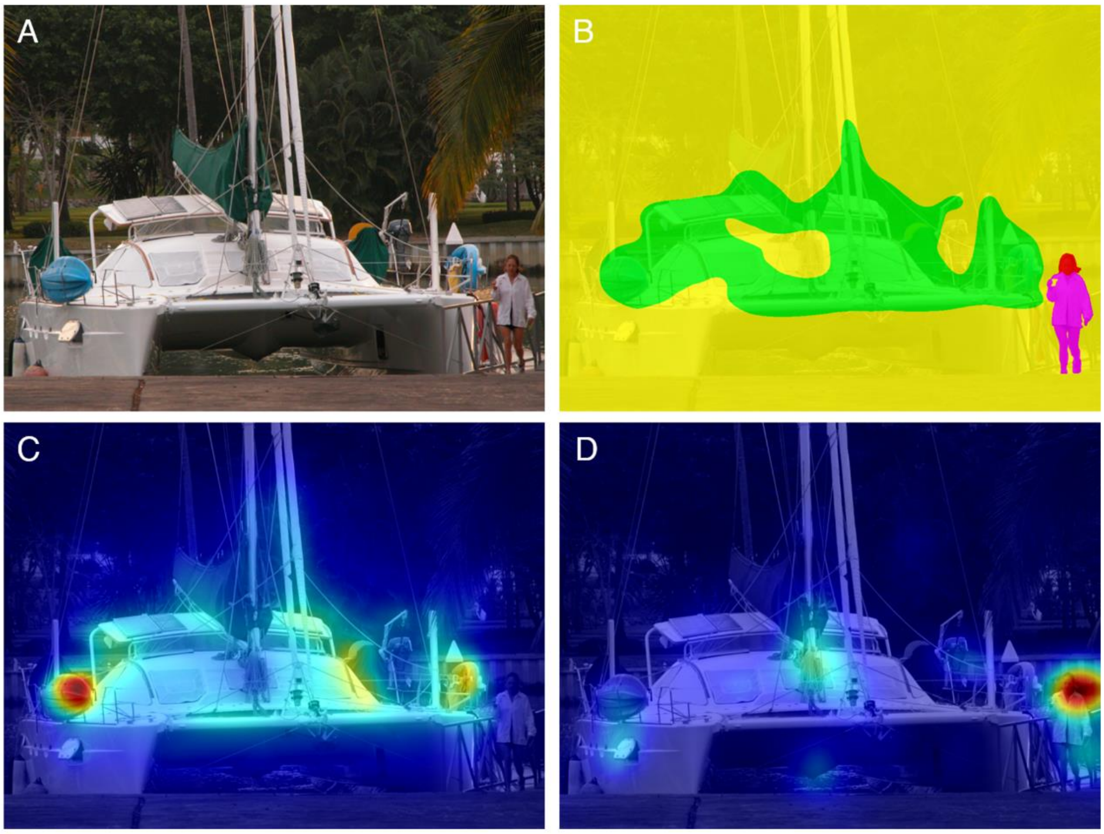

Figure 1.2. Example of a social stimulus as published in End & Gamer (2017). (A) Original scene. (B) Example of defined regions of interest for head (red), body (magenta), low saliency (yellow) and high saliency (green). (C) Overlay of a saliency map according to the Graph-Based Visual Saliency algorithm (Harel et al., 2007) with cool colors representing low salient regions and warm colors defining areas of high saliency. (D) Overlay of a fixation density map derived from fixation patterns of 31 participants who viewed the stimulus for 10s under free-viewing conditions. Image taken with permission from the Nencki Affective Picture System (Marchewka, Żurawski, Jednoróg, & Grabowska, 2014). Please note that the stimulus shown here was not used in the current study and is only depicted to illustrate the current data analysis strategy.

RESULTS

BEHAVIORAL DATA

Reaction Times were longest for counting objects followed by estimating and color definition (Figure 1.3A) resulting in a main effect of task ($F_{(2,78)} = 193.29$, $\varepsilon = .96$, $p < .001$, $\eta^2 = .56$). When comparing social and non-social stimuli, a main effect of stimulus category becomes apparent ($F_{(1,39)} = 21.57$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .008$), showing slightly higher reaction times for social stimuli. A two-way interaction between task and stimulus category ($F_{(2,78)} = 8.17$, $\varepsilon = .99$, $p < .001$, $\eta^2 = .008$) depicts higher reaction times for object counting and color definition for social stimuli than non-social ones, whereas estimation tasks required slightly longer reaction times for non-social stimuli. Concerning the complexity ratings, estimating was generally rated the hardest and color definition the easiest resulting in a main effect of task ($F_{(2,78)} = 55.02$, $\varepsilon = .99$, $p < .001$, $\eta^2 = .46$). This evaluation is somewhat reflected in the reaction time data, as the easiest definition task was also the one with the lowest response times. However, counting objects seemed to require a longer focus than estimation, even though estimation was rated the hardest (Figure 1.3B).

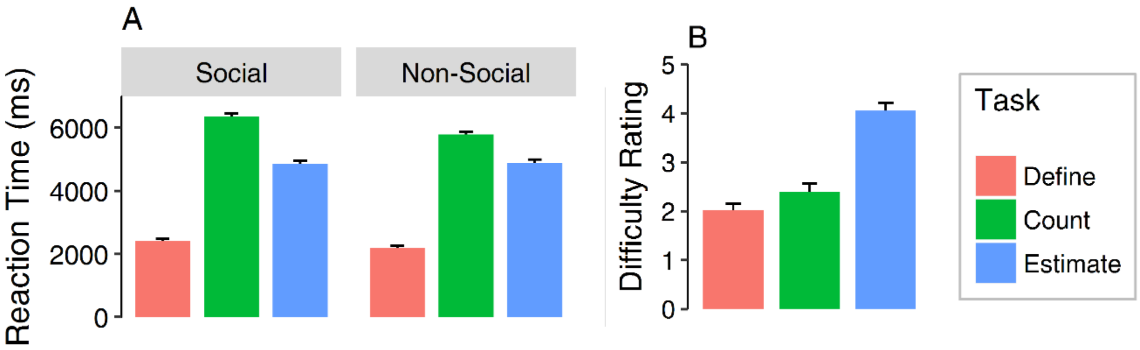

Figure 1.3. Reaction times (ms) as a function of task and stimulus category (A) and difficulty ratings for the different tasks (B). Error bars represent standard errors of the mean.

EYE-TRACKING

TASK-RELEVANT VIEWING BEHAVIOR

Analyses of the relative area-normed fixation density on social stimuli revealed a main effect of ROI ($F_{(3,117)} = 302.51$, $\varepsilon = .46$, $p < .001$, $\eta^2 = .67$), showing a higher fixation densities on heads than on areas of high saliency, a main effect of task ($F_{(3,117)} = 53.70$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .22$), depicting higher fixation densities for free-viewing than for estimation, then color definition and lowest for object counting. Furthermore, we also found an interaction between ROI and task ($F_{(9,351)} = 66.66$, $\varepsilon = .44$, $p < .001$, $\eta^2 = .42$) emphasizing that heads were fixated the longest in the free-viewing condition, followed by the estimation task, then the color definition and the smallest difference between heads and areas of high saliency was found in the counting task (Figure 1.4, left panel). Attention towards heads was most diminished in the easiest task, namely the color definition, but remained superior to image regions with high physical saliency (heads: $M = 0.26$, areas of high saliency: $M = 0.23$). Separate analyses were conducted for selection behavior in terms of fixation frequency as opposed to the reported analyses of fixation densities. Additionally, relative fixation densities were further analyzed as well, taking into account not only the area of the ROI, but also the number of ROIs within each category for each stimulus. Since both of these analyses provided highly similar results to the ones presented here, they are not described further.

TASK-IRRELEVANT VIEWING BEHAVIOR

A comparison of the two time windows (before and after the participants' response to the task) with the factor time confirmed these effects with an additional main effect of time ($F_{(1,39)} = 37.19$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .04$) with different fixation durations before and after the response to the given task. A main effect of task remained ($F_{(2,78)} = 4.98$, $\varepsilon = 1$, $p = .009$, $\eta^2 = .01$), which largely resembles the free-viewing condition before the click with an additional main effect of ROI ($F_{(3,117)} = 192.79$, $\varepsilon = .40$, $p < .001$, $\eta^2 = .57$) depicting a preference to focus on heads compared to other aspects of the picture. Furthermore, interactions between all three aspects was found: task and ROI ($F_{(6,234)} = 13.21$, $\varepsilon = .45$, $p < .001$, $\eta^2 = .059$), task and time ($F_{(2,78)} = 8.86$, $\varepsilon = .99$, $p < .001$, $\eta^2 = .01$), as well as ROI and time ($F_{(3,117)} = 45.78$, $\varepsilon = .56$, $p < .001$, $\eta^2 = .09$), revealing higher fixation durations for heads as compared to other ROIs in the tasks of color definition and object counting after completion of the task compared to before the response. An additional three-way interaction ($F_{(6,234)} = 8.09$, $\varepsilon = .51$, $p < .001$, $\eta^2 = .03$) confirms the tendency to prioritize social information when given the choice, even if a task preceded that required attention towards other aspects. In direct comparison, the attention towards heads is slightly diminished by the given task (before response), yet regains higher fixation durations after its completion (Figure 1.4).

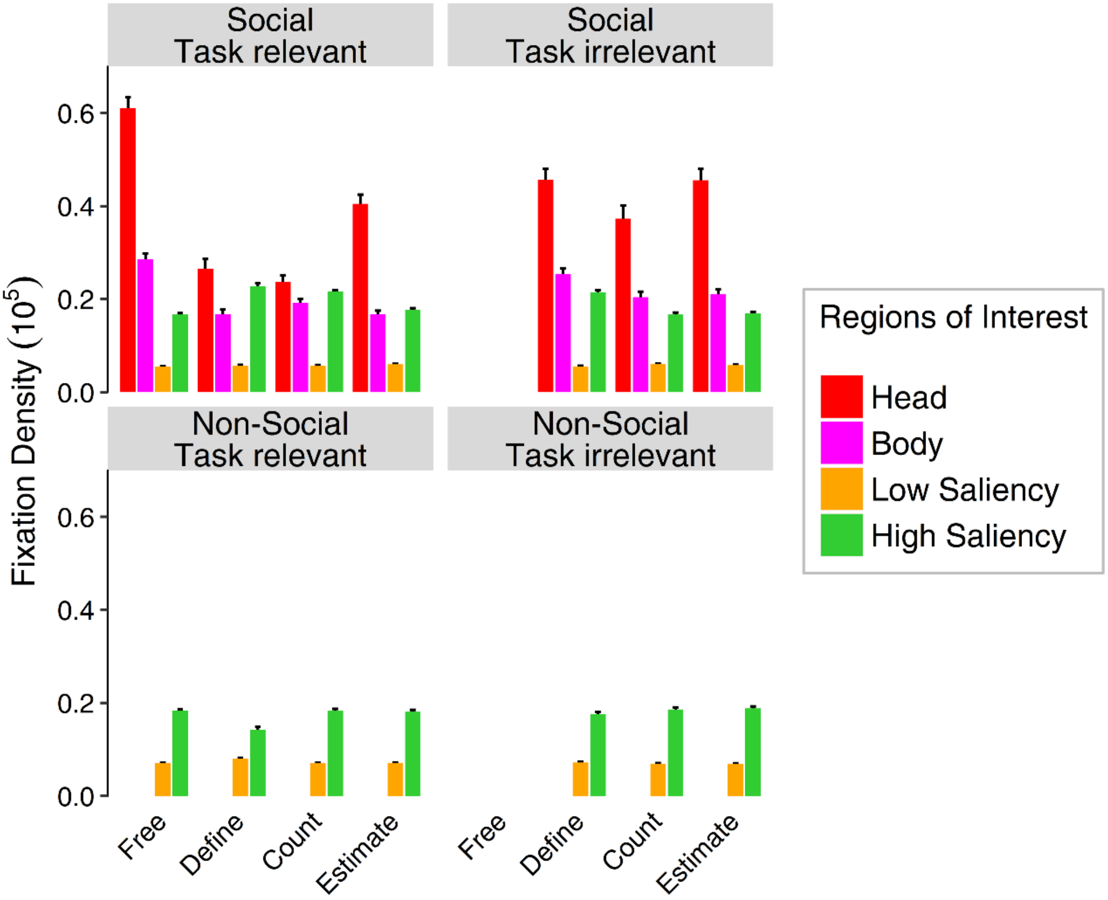

Figure 1.4. Direct comparison of relative area normed fixation density for social (above) and non-social stimuli (below) on different regions of interest for two different time windows (before and after a given response) as a function of task. The time window before the mouse click is referred to as "task relevant" (left panels), as it reflects possible top-down strategies on viewing behavior that were induced by the explicit tasks. The time window after completing the task is termed "task irrelevant" (right panels), as attentional allocation in this period should be less governed by task instructions. As there was no postulated task for the free-viewing condition, the task irrelevant time window does not contain any data for this modality. Viewing behavior for non-social stimuli was analyzed without social regions of interest, but allowed for comparing relative area normed fixation densities for areas of low and high saliency. Error bars indicate standard errors of the mean.

SALIENCY EFFECTS WITHIN SOCIAL REGIONS OF INTEREST

The ANOVA taking into account the saliency distribution also within social ROIs, resulted in a main effect of salience ($F_{(1,39)} = 611.98$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .20$), indicating that participants rather fixated areas of high saliency as compared to areas of low saliency across all ROIs. This effect was qualified by an interaction of saliency and ROI ($F_{(2,78)}$ = 22.16, ε = .97, p < .001, η^2 = .01) (see Figure 1.5A), demonstrating larger effects of saliency on fixation densities for the background as compared to social ROIs. Furthermore, a main effect of task ($F_{(3,117)}$ = 44.79, ε = .95, p < .001, η^2 = .22) again described differences between tasks with generally higher fixation densities for free-viewing and estimating compared to defining and counting. A main effect of ROI was evident ($F_{(2,78)}$ = 259.71, ε = .66, p < .001, η^2 = .52), as head ROIs were still viewed for longer durations than other ROIs. Interactions between saliency and task ($F_{(3,117)}$ = 9.50, ε = .83, p < .001, η^2 = .01), as well as task and ROI ($F_{(6,234)}$ = 55.87, ε = .59, p < .001, η^2 = .28) were also significant but qualified by an additional three-way interaction of saliency, task and ROI ($F_{(6,234)}$ = 3.31, ε = .71, p = .004, η^2 = .006). These effects indicate different fixation densities onto highly salient and less salient regions depending on the content (i.e., social vs. non-social regions) as well as the task. To further illustrate the actual influence of saliency onto viewing behavior towards these regions, we calculated the difference between high and low salient regions of each ROI. These results showed that difference values for head and body ROIs were lower than those for the background for all conditions except the definition task (see Figure 1.5B). Hence, saliency generally played a larger role in drawing attention for non-social aspects than for social ones, especially heads.

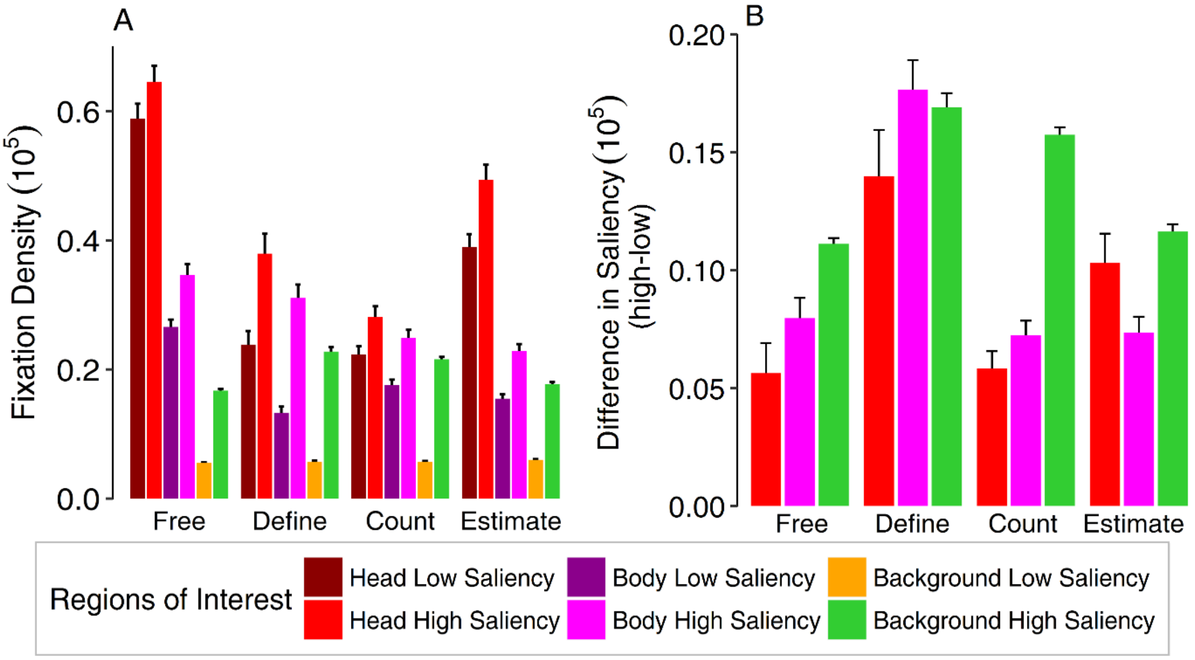

Figure 1.5. (A) Relative area normed fixation densities as a function of ROI (head, body and background) and saliency (high vs. low) of these regions across different tasks. (B) Difference in relative area normed fixation densities between highly salient and less salient areas for each region of interest (head, body and background). Please note that the color coding of ROIs was adapted to the values of the highly salient regions. Error bars represent standard errors of the mean.

FIRST FIXATIONS

The first three fixations were more closely analyzed with regard to the preference of specific ROIs (Figure 1.6). Over all tasks, a distinction of initial fixations on the head can be seen compared to the other areas resulting in a main effect of ROI ($F_{(3,111)} = 163.55$, $\varepsilon = .36$, $p < .001$, $\eta^2 = .41$). Additionally, a main effect of task ($F_{(3,111)} = 48.30$, $\varepsilon = .71$, $p < .001$, $\eta^2 = .09$) depicts slightly different viewing behavior for the different tasks, especially color definition and a significant interaction between ROI and task ($F_{(9,333)} = 44.09$, $\varepsilon = .28$, $p < .001$, $\eta^2 = .24$) can be attributed to changes across tasks driving attention away from heads and onto aspects potentially relevant for the tasks. This is also underpinned by a main effect of fixation number ($F_{(2,74)} = 14.61$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .008$) and an interaction between fixation number and ROI ($F_{(6,222)} = 18.57$, $\varepsilon = .38$, $p < .001$, $\eta^2 = .04$) as well as fixation number and task ($F_{(6,222)} = 11.44$, $\varepsilon = .91$, $p < .001$, $\eta^2 = .02$) implying changes from initial fixations on heads to subsequent fixations onto task-relevant areas of high saliency for color definition after the first fixation and object counting after the second fixation. This observation resulted in a significant three-way interaction for fixation number, ROI and task ($F_{(18,666)} = 9.92$, $\varepsilon = .35$, $p < .001$, $\eta^2 = .07$).

To disentangle this observation with respect to the separate tasks, post hoc analyses revealed the following pattern: in the free viewing condition, participants clearly favored heads, fixating them for almost three times as long as all other ROIs and with an increasing tendency over the number of fixations (main effect of ROI: $F_{(3,111)} = 125.58$, $\varepsilon = .35$, $p < .001$, $\eta^2 = .56$, main effect of fixation: $F_{(2,74)} = 23.55$, $\varepsilon = .97$, $p < .001$, $\eta^2 = .04$, as well as an interaction of both factors: $F_{(6,222)} = 24.87$, $\varepsilon = .37$, $p < .001$, $\eta^2 = .17$). The tasks of estimating and counting showed a slightly diminished fixation duration on heads, but displayed a similar pattern concerning the progression over fixations, particularly for heads (*estimate*: main effect of ROI: $F_{(3,111)} = 64.41$, $\varepsilon = .39$, $p < .001$, $\eta^2 = .43$, main effect of fixation: $F_{(2,74)} = 3.21$, $\varepsilon = 1$, $p = .047$, $\eta^2 = .008$, and interaction: $F_{(6,222)} = 5.58$, $\varepsilon = .40$, $p < .001$, $\eta^2 = .05$; *count*: main effect of ROI: $F_{(3,111)} = 35.78$, $\varepsilon = .37$, $p < .001$, $\eta^2 = .26$, main effect of fixation: $F_{(2,74)} = 6.55$, $\varepsilon = .96$, $p = .002$, $\eta^2 = .02$ and interaction: $F_{(6,222)} = 5.06$, $\varepsilon = .35$, $p < .001$, $\eta^2 = .05$). For color definition, fixations towards the vehicle (mainly included in the areas of high saliency) were quickly made from the second fixation onwards, however, the initial fixations remained directed towards the head of actors in the scene also resulting in a main effect of ROI ($F_{(3,111)} = 41.19$, $\varepsilon = .43$, $p < .001$, $\eta^2 = .22$), a main effect of fixation ($F_{(2,74)} = 10.21$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .03$) and a significant interaction ($F_{(6,222)} = 3.88$, $\varepsilon = .39$, $p = .001$, $\eta^2 = .05$). Similar to the results we found for overall fixation durations, the tasks elicit different viewing behaviors for the first fixations. That is, the more specific a task was, i.e. defining the color of the only vehicle in the picture or counting distinct blue objects, resulted in a faster drift of fixations from heads onto areas of high saliency, which largely included those features than for trials allowing broader scanning of the stimulus, i.e. to estimate the total amount of white.

Figure 1.6. Relative area normed fixation density of the first three fixations after stimulus onset (before response) on the different regions of interest for all tasks separately. Top row: first fixations for social stimuli on four regions of interest (head, body, areas of low saliency and high saliency), bottom row: first fixations for non-social stimuli on two regions of interest (areas of low and high saliency). Error bars indicate standard errors of the mean.

COMPARISON OF SOCIAL AND NON-SOCIAL STIMULI

When comparing the gaze behavior with regard to saliency for social and non-social stimuli, we found a significant difference in fixation density between the categories ($F_{(1,39)} = 78.82$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .033$) and a two-way interaction between task and stimulus category ($F_{(2,78)} = 63.11$, $\varepsilon =$

.98, $p < .001$, $\eta^2 = .046$) as well as ROI and stimulus category ($F_{(1,39)} = 78.52$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .06$) and a three-way interaction of all three factors ($F_{(2,78)} = 62.73$, $\varepsilon = .98$, $p < .001$, $\eta^2 = .12$) confirming that that social scenes experienced longer fixation durations onto areas of high saliency for specific tasks like color definition and object counting, while non-social scenes experienced diminished fixations on areas of high saliency, specifically for the color definition task (Figure 2.4). Furthermore, a main effect of ROI ($F_{(1,39)} = 1537.84$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .88$) was significant, showing a preference for fixating areas of high saliency compared to those of low saliency throughout all tasks and stimuli, a main effect of task ($F_{(2,78)} = 3.23$, $\varepsilon = .98$, $p = .047$, $\eta^2 = .005$) and an interaction effect between task and ROI ($F_{(2,78)} = 3.18$, $\varepsilon = .98$, $p = .047$, $\eta^2 = .01$) depicting different fixation durations on areas of high saliency for different tasks across stimulus categories. With regard to task-relevant as opposed to task-irrelevant viewing behavior, we found a main effect of time ($F_{(1,39)} = 7.91$, $\varepsilon = 1$, $p = .008$, $\eta^2 = .006$). Further, an interaction between task and stimulus category ($F_{(2,78)} = 63.11$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .05$) and between ROI and stimulus category ($F_{(1,39)} = 78.52$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .09$) revealed different viewing behavior on salient aspects for social and non-social stimuli with regard to different tasks, which is confirmed by a three-way interaction of task, ROI and stimulus category ($F_{(2,78)} = 62.73$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .12$). The interaction between stimulus category and time did not reach statistical significance ($F_{(1,39)} = 0.75$, $\varepsilon = 1$, $p = .39$, $\eta^2 = .0002$). Accordingly, a ROI by stimulus category by time interaction was not significant either ($F_{(1,39)} = 0.73$, $\varepsilon = 1$, $p = .39$, $\eta^2 = .0006$). Considering different viewing behavior across tasks, an interaction between task and time was still evident ($F_{(2,78)} = 11.33$, $\varepsilon = .97$, $p < .001$, $\eta^2 = .01$) as well as a three-way interactions of task, ROI and time ($F_{(2,78)} = 13.17$, $\varepsilon = .97$, $p < .001$, $\eta^2 = .03$) and task, stimulus category and time ($F_{(2,78)} = 6.41$, $\varepsilon = 1$, $p = .003$, $\eta^2 = .003$). Furthermore, there was a two-way interaction of ROI and time ($F_{(1,39)} = 7.88$, $\varepsilon = 1$, $p = .008$, $\eta^2 = .02$) and a four-way interaction of task, ROI, stimulus category and time ($F_{(2,78)} = 6.32$, $\varepsilon = 1$, $p = .003$, $\eta^2 = .01$) indicating that fixations differed across tasks for both stimulus categories for low and high saliency depending on the time window.

Analyses of the first three fixations after stimulus onset revealed a main effect of ROI ($F_{(1,38)} = 2710.73$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .82$) clearly showing a preference for attending areas of high saliency compared to areas of low saliency, a main effect of task ($F_{(3,114)} = 12.49$, $\varepsilon = .97$, $p < .001$, $\eta^2 = .008$) that showed slight changes in viewing behavior across tasks and a significant interaction between the two ($F_{(3,114)} = 12.49$, $\varepsilon = .97$, $p < .001$, $\eta^2 = .02$). An additional main effect of fixation number ($F_{(2,76)} = 126.96$, $\varepsilon = .96$, $p < .001$, $\eta^2 = .06$) and an interaction between ROI and fixation number ($F_{(2,76)} = 127.09$, $\varepsilon = .96$, $p < .001$, $\eta^2 = .14$), as well as task and fixation number ($F_{(6,228)} =$

7.06, $\varepsilon = .88$, $p < .001$, $\eta^2 = .007$) was found, implying changes in initial and subsequent fixations with a strong bias towards areas of high saliency. A three-way interaction between ROI, task and fixation number ($F_{(6,228)} = 7.06$, $\varepsilon = .88$, $p < .001$, $\eta^2 = .02$) depicted that for specific tasks like color definition and object counting, areas of high saliency generally experienced longer fixations than tasks requiring broad scanning, which is similar to the data found for social stimuli. Comparing viewing behavior for social and non-social stimuli, we found no statistically significant main effect of stimulus category ($F_{(1,38)} = 3.43$, $\varepsilon = 1$, $p = .072$, $\eta^2 = .005$) and no significant interaction between ROI and stimulus category ($F_{(1,38)} = 4.07$, $\varepsilon = 1$, $p = .051$, $\eta^2 = .007$). However, an interaction between task and stimulus category ($F_{(3,114)} = 14.47$, $\varepsilon = .47$, $p < .001$, $\eta^2 = .06$) was significant, probably driven by different fixation progressions for social stimuli. Specifically, the tasks color definition and counting objects, experienced almost reversed fixation patterns for areas of high saliency for non-social stimuli as compared to social ones implied by an interaction for fixation number and stimulus category ($F_{(2,76)} = 8.52$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .002$), which was especially the case for areas of high saliency (interaction for fixation number, stimulus category and ROI: $F_{(1,76)} = 9.59$, $\varepsilon = 1$, $p < .001$, $\eta^2 = .006$). A statistically significant four-way interaction between ROI, task, stimulus category and fixation number was not found ($F_{(6,228)} = 1.65$, $\varepsilon = .73$, $p = .134$, $\eta^2 = .004$).

DISCUSSION

This study aimed at investigating differences in viewing behavior for naturalistic stimuli with social or non-social content during implementation of top-down demands to measure the degree to which social attention may be able to override top-down processes. By utilizing tasks of different complexity, we hoped to witness a gradation of potential influence on social information processing. Indeed, we found significant differences in viewing behavior across tasks. In the free-viewing condition, participants displayed a clear preference for social aspects, especially for heads, while tasks, which required gaze towards distinct non-social aspects of a scene (e.g. defining the color of a vehicle or counting objects of the same color) most effectively drew attention away from social features, yet without ever eliminating fixations on heads. Therefore, it seems that not necessarily the difficulty, but the specificity of a task interferes with social attention to a certain extent. These results are in line with Yarbus (1967) and a corresponding follow-up study by DeAngelus & Pelz (2009) using seven different tasks, in which those questions addressing global or multiple features resulted in more spatially distributed patterns of fixations, whereas free-viewing and tasks requiring assessment of the social feature itself produced fixations that fell primarily on faces and figures. Yarbus concluded that the observers' fixations reflected the most informative regions for the task at hand, which is central to the "bottom-up/top-down debate", with the invention of saliency maps supporting mere bottom-up models. However, most of these tasks necessitated scanning of social features (e.g., "Give the ages of the people") and therefore primary elements referring to faces and figures, were invariably fixated. The tasks used in our experiments deliberately focused on non-social aspects of a scene to investigate whether fixations indeed reflected the most informative regions. According to our results, this is not necessarily the case, as heads were preferentially attended despite task requirements. Specifically, even though visually salient image regions included task-relevant details, visual exploration towards heads remained superior to saliency in all tasks. This preference for social features, was present even though participants received the task instruction before stimulus onset and could therefore prepare by generating category-specific representations.

Additionally, the interference of social attention by top-down mechanisms was only manifested for the duration of task resolution and drastically reverted back to the pattern found in free-viewing, where social information, especially heads, regained their initial preference. This return was visible for all tasks and resulted in significant differences in favor of social aspects compared to salient areas of the scene, which agrees with findings of Birmingham, Bischof & Kingstone (2008b) implying a hierarchy of selection for social stimuli. Although heads gained the most attention, bodies represented the second most fixated regions in free-viewing and post-task gaze behavior. The more specific a task became, however, the more bodies forfeited their focus to task-relevant details. Nevertheless, these details did not influence viewing behavior after task completion such that bodies regained relevance, towering physical saliency, similar to the pattern seen for heads. For non-social stimuli however, viewing behavior did not change as drastically over different tasks or time windows and mainly revealed a preference for fixating highly salient image regions.

Generally, studies involving explicit tasks to modulate attentional orienting to social features have primarily used isolated facial stimuli in comparison to objects (e.g. Bindemann et al., 2007; Hershler & Hochstein, 2005; Ro et al., 2001). Herein, most results confirm faster detection of faces compared to non-social objects. However, isolation of a social cue avoids other potentially important aspects driving gaze selection as is the case in the real world. Hence, complex naturalistic scenes enable more general analyses of gaze patterns towards social features and consider potentially competing factors. A study of Kuhn and colleagues (2016) investigated top-down effects in the form of instructions to assess the modulation of gaze during both live presentation or that of a video depicting a magician as social feature. Herein, instructions aimed at directing attention away from the social aspect ("keep your eyes on the cards") modulated viewing behavior in some participants, indicating that viewers have certain top-down control, yet other participants' gaze persevered on the face despite the instruction, which implies a residual bias towards the face in competition with top-down control. These results are consistent with our findings, as we also find task effects, although the preference to attend faces remained across all tasks. Furthermore, gaze-following studies have shown that humans seem to have a strong predisposition to follow or imitate a social cue, even when this is obstructive to task performance (Koval, Thomas, & Everling, 2005; Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002). Coherently, the study of Birmingham, Bischof and Kingstone (2008) investigated the selection of gaze information as suggested by Yarbus in a more representative manner using multiple complex stimuli and three different task instructions. Their results substantiate our findings as such that attention was captured by eyes and heads of people in the scene regardless of the task. In non-social settings, such as outdoor scenes, however, task-demands counteracted sensory signals fully (Einhäuser, Rutishauser, & Koch, 2008). Although this top-down override appears to be rather strong in a non-social context, attention towards social aspects seems to withstand such counteraction, according to the results of Birmingham et al., as well as our own. However, the tasks chosen by Birmingham and colleagues encouraged scanning of the social aspects once again, as one goal was to investigate the influence of social content and activity in the scene on gaze behavior towards the eyes. Nevertheless, this makes our results all the more enlightening concerning the natural preference to select social aspects and the resilience of social attention when faced with behaviorally significant competition. Furthermore, attention towards salient image regions remained comparable between social and non-social scenes in the current study with a stronger focus on areas of high as opposed to those of low saliency. Hence, the results discussed above seem to mainly rely on attention processes towards the social scene aspects. We further addressed the influence of saliency within these social aspects and although results indicate that saliency does influence allocation of attention to more salient regions within head and body ROIs in a scene, the difference between high and low saliency was smaller for these social areas compared to that of the background. This is yet another indicator that social aspects experience a different viewing behavior than non-social ones.

According to Sharma (2015), who compared the performance of 10 different algorithms, most saliency models are statistically close to each other in their correspondence with human eye fixations. The fact that we observed a preference to regard human beings over physically salient aspects, implies that saliency algorithms may struggle to accurately describe viewing behavior in the presence of social information. Furthermore, this raises the question about which underlying processes are involved in the direction of and prioritized attention towards social features. It is possible that some stimuli attract attention because of some form of contingency that is hard wired in the brain by learning, development or genetics Corbetta & Shulman, 2002. Hence, some might argue that observers might have simply been interested to select faces voluntarily indicating that social attention is a special form of top-down mechanism that we have internalized through experience. However, a few studies demonstrating an early onset of preferential orienting towards social information argue against this notion. For example, Fletcher-Watson and colleagues (Fletcher-Watson et al., 2008) did not only find a bias towards looking at a social stimulus compared to a non-social one during free-viewing, but also reported that this bias was already evident in the first fixation occurring as early as 100ms after stimulus presentation. Another example is that of Crouzet, Kirchner & Thorpe (2010) who showed that saccades towards faces occur as early as 100-110ms, even in competition with simultaneously presented target stimuli, suggesting that this mechanism is not completely under instructional control. Also our own analyses are in accord with this, as the first fixations reveal an initial, possibly almost reflexive bias to attentionally select social information in a complex scene. According to our results, top-down goals of for example detecting a vehicle, seem to affect the second fixation at the earliest, however, the first fixations are primarily directed at heads. Similarly, electrophysiological (Pegna, Khateb, Michel, & Landis, 2004) and imaging studies (Liu, Harris, & Kanwisher, 2002) found faces to be categorized as early as 100ms after stimulus onset, whereas objects required around 200ms for categorization. Further support can be found in social orienting paradigms, as shifts towards gazed at locations occur rapidly, within a few hundred milliseconds after a gazing face is presented (e.g. Friesen & Kingstone, 1998) even if eye direction is counter-predictive of target location (Driver et al., 1999). Yet another example can be drawn from studies with infants who show looking preferences for faces, which led Johnson (Johnson, 2005) to conclude that the adult 'social brain' may be developmentally founded through a subcortical face-detection system involving the superior colliculus, pulvinar and amygdala. This route was initially proposed by clinical studies concerning patients with hemispacial neglect whose visual extinction towards stimuli in their neglected field was revoked if the elements were arranged in the pattern of a face (Vuilleumier, 2000; Vuilleumier & Sagiv, 2001). Also Dolan and colleagues (Dolan et al., 1997) found that ambiguous pictures activated face-processing regions only when observers recognized the pictures as depicting faces. These studies indicate that there may be a separate route for face detection (Johnson, 2005).

On the grounds of our study, we cannot conclude definitely whether mechanisms of social attention are voluntary or reflexive, nevertheless, this rationale suggests very early processing of social compared to non-social information that seem to precede top-down mechanisms. Further studies are needed to confirm this idea and differentiate between automatic shifts of attention and conscious selection of scene elements. This may be accomplished through a gaze-congruent paradigm to emphasize top-down control (e.g. Foulsham, Teszka, & Kingstone, 2011; Kennedy & Adolphs, 2011; Loschky & McConkie, 2002; van Diepen & d'Ydewalle, 2003) in comparison to very short presentation times of stimuli that preclude active exploration of visual scenes (e.g., 100ms) for stressing bottom-up mechanisms (see review Carrasco, 2011) or dot-probe variants for stimuli comprising a social "side" versus a non-social one (e.g. Bindemann et al., 2007). On the other hand, we do not state that social attention can be categorized as a clear-cut bottom-up process either that is driven by low-level physical features, supported by our finding that physical saliency is not an accurate predictor of attentional focus. Furthermore, studies investigating the "pop-out" effect of faces have been inconclusive (VanRullen, 2006). Instead, we suggest a potential additional attention mechanism that drives the social override. To examine the neural nature of this override, functional neuroimaging may offer insight into networks that drive social attention independently of bottom-up and top-down mechanisms. Herein, we suggest a subcortical route rapidly conveying information to the amygdala. This has previously also been suggested for fear recognition, as the amygdala responds rapidly to emotional faces, especially fearful expressions (e.g. Morris, DeGelder, Weiskrantz, & Dolan, 2001) and seems to be implicated in eliciting gaze shifts towards specific diagnostic facial features (e.g. Gamer & Büchel, 2009; Gamer, Schmitz, Tittgemeyer, & Schilbach, 2013). Furthermore, Fitzgerald and colleagues (Fitzgerald, Angstadt, Jelsone, Nathan, & Phan, 2006) concluded from an functional magnetic resonance imaging (fMRI) study that the amygdala is not selective of any particular emotion category, but instead, may have a more general-purpose function in processing facial information. Therefore, the amygdala might be involved in orienting attention towards social features in general and faces in particular in natural every-day environments similar to the suggestions of Johnson (2005) concerning a subcortical route for face detection and identification.

Although the current study revealed important insights into the influence of top-down demand on the extent of social attention, some limitations need to be acknowledged. First, we cannot exclude the possibility of carry-over effects in this study, as stimuli integrated blue objects or cars even for free-viewing tasks. Second, specific image regions were not entirely confined to one task (e.g. cars could have bright or bluish colors, therefore being relevant for estimation and counting tasks, respectively). Yet this further strengthens our findings as even though carry-over effects may have been present in visual exploration, social information was still preferentially selected, despite the fact that it was irrelevant and rather disadvantageous to solving the task. Third, some might argue about the implied "liveliness" of human beings in social versus non-social stimuli and we cannot fully negate this assumption with our data. However, some pictures included animals, which would have contributed to low or highly salient areas of the scene and challenging fixations towards human features. Furthermore, Crouzet, Kirchner & Thorpe (2010) explicitly investigated saccade velocity towards human faces compared to animals and found that while animals can be detected as early as 120-130ms after stimulus onset, saccades to human faces are even faster with the earliest saccades occurring at 100-110ms.

CONCLUSION

We have shown that while people will fixate other parts of a complex visual scene to extract relevant information, their preferential bias is to fixate faces of others. Furthermore, gaze selection seems to be driven by the goal to extract social information, even if this visual selection is irrelevant or even disadvantageous to solving a task. Our results therefore allow us to suggest that bottom-up driven saliency seems to have less influence on attentional orienting in a social context. Although top-down demands interfere with social attention, they do not extinguish fixations on conspecifics and they are strongly confined to the time of the task. This supports our hypothesis that social stimuli may engage special perceptual processing and provide exclusive access to the priority map, enabling a partial override of top-down as well as bottom-up mechanisms in attentional orienting. Future studies will have to reveal whether there is a qualitative instead of a quantitative difference between social and non-social attention and whether we have specific neural circuits that are supporting the attentional prioritization of human beings.

ACKNOWLEDGEMENTS

The authors thank Albert End for his assistance with data analyses.

AUTHOR CONTRIBUTIONS

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

IMPLICATIONS FOR STUDY 2

The first study investigated the influence of top-down control on social attention by introducing deviating tasks differing in complexity and showed that faces of depicted human beings in the scenes were preferentially attended independent of the given task. Hence, top-down demands cannot deter from the tendency to attend social aspects, so the subsequent deduction is to investigate the alternative, namely to what extent social attention is driven or influenced by bottom-up information. The second study examined the reflexive nature of social orienting and further differentiated between overt and covert attention. Although initial fixations after stimulus onset, which suggest an early onset of the observed social bias, were analyzed in study 1 (section 2.1), we wanted to limit the access of information to very brief durations to further isolate bottom-up influences. Simultaneously, we wanted to use the opportunity to replicate our findings and offer a direct comparison between uni-operational as well as bi-operational functioning of bottom-up and top-down processes in free-viewing, which is why we incorporated a longer presentation time in the overt condition.

INVESTIGATING OVERT AND COVERT SHIFTS OF ATTENTION WITHIN SOCIAL NATURALISTIC SCENES

ALEYA F. FLECHSENHAR[1], OLIVIA LARSON[2], ALBERT END[3] & MATTHIAS GAMER[1]

[1]JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG

[2]MCGILL UNIVERSITY MONTREAL

[3]UNIVERSITY MEDICAL CLINIC HAMBURG-EPPENDORF

Corresponding author: Aleya F. Flechsenhar, Department of Psychology, Julius Maximilian University Wuerzburg, Marcusstrasse 9-11, 97070 Würzburg, Germany.

E-mail address: aleya.flechsenhar@uni-wuerzburg.de

Published 2018 in Journal of Vision, 18(12):11.

ABSTRACT

Eye-tracking studies on social attention have consistently shown that humans prefer to attend to other human beings. Much less is known about whether a similar preference is also evident in covert attentional processes. To enable a direct comparison, this study examined covert and overt attentional guidance within two different experimental setups using complex naturalistic scenes, instead of isolated single features. In the first experiment, a modified version of the dot-probe paradigm served as a measure of covert reflexive attention towards briefly presented scenes containing a social feature in one half of the visual field compared to non-social elements in the other while controlling for low-level visual saliency. Participants showed a stable congruency effect with faster reaction times and fewer errors for probes presented on the social side of the scene. In a second experiment, we tracked eye movements for the same set of stimuli, while manipulating the presentation time to allow for differentiating reflexive and more sustained aspects of overt attention. Supportive of the first results, analyses revealed a robust preference for social features concerning initial saccade direction as well as fixation allocation. Collectively, these experiments imply preferential processing of social features over visually salient aspects for automatic allocation of covert as well as overt attention.

INTRODUCTION

Social interactions are part of our everyday life and thereby essential for communication, empathy and clarifying intent or needs for mutual understanding. Initial studies have confirmed a tendency to focus on other human beings in a social context, which has sparked extensive research in the field of social attention. Early studies frequently used isolated stimuli to compare allocation of attention towards social as opposed to non-social features (e.g., Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 1999; Pelphrey et al., 2002; Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002; Ro, Russell, & Lavie, 2001). One way to investigate this bias is by measuring the free allocation of eye movements to deduce attention distribution to these areas, also referred to as overt attention. However, some laboratory findings using simplified (e.g., iconic images or geometric shapes) or isolated stimuli (e.g., showing a face without the remaining body or contextual information) have yielded different results than studies using more naturalistic setups (Kingstone, 2009). When a face is presented in isolation, there is a clear preference to fixate the eye region of these images, but when a face is presented with its associated body parts, the tendency to scan the eyes disappears (Kingstone et al., 2002). This altered behavior may be due to a lack of contextual information and situational complexity in the former as compared to the latter condition.

Naturalistic scenes are rich in detail and provide meaningful contexts for objects influencing the selection process of eye-movements, e.g. through scene gist (e.g., Oliva & Torralba, 2006) or a multitude of competing objects (e.g., Einhäuser, Spain, & Perona, 2008; but see Borji, Sihite, & Itti, 2013). This aspect is often diminished in impoverished visual displays (see also Anderson, Ort, Kruijne, Meeter, & Donk, 2015). As a consequence, researchers have started to utilize naturalistic scenes to extend previous results to a more realistically complex, yet controlled setting. Recent studies have additionally introduced and investigated the presence of a human being within these scenes and its effects on gaze behavior (e.g., Birmingham, Bischof, & Kingstone, 2008a, 2008b; Cerf, Frady, & Koch, 2008; End & Gamer, 2017; Flechsenhar & Gamer, 2017; Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Rösler, End, & Gamer, 2017; Rubo & Gamer, 2018; Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014). These eye-tracking studies have shown that social features such as faces or bodies of other human beings are preferentially attended in free viewing conditions. However, the use of naturalistic stimuli has mainly been restricted to experiments addressing overt attention by means of eye gaze, but not for those that restrict eye movements in order to investigate covert shifts of attention. In addition to the distinction between covert and overt shifts of attention, the literature typically dissociates between reflexive, externally driven shifts of attention and voluntary shifts of attention. Traditional covert paradigms, such as variations of the Posner cueing paradigm (Posner, 1980) involve the presentation of cues that either address endogenous, top-down driven control or exogenous, bottom-up driven attentional orienting while restricting eye movements. In the original experiment, two boxes are presented on the screen and one of these boxes will reveal a checkboard. Before the onset of this probe, one of the boxes may be cued exogenously using a flash at the corresponding location, reflexively drawing attention to it. Alternatively, a central arrow cue may point to one of the boxes, triggering an endogenous shift of attention. The cues can either correctly indicate where a stimulus is going to appear (valid cue), or point to an incorrect location (invalid cue). Even when the cue does not reliably predict where the probe is going to appear, participants typically respond faster on trials where the probe is congruent with the cued and therefore attended location compared to a cue that was incongruent. This phenomenon is referred to as a congruency effect. Although the majority of these paradigms use directional cues, such as arrows (e.g., Tipples, 2002), some studies have introduced social endogenous cues, referring to centrally presented faces or eyes indicating a certain spatial location by changing the gaze direction (Deaner & Platt, 2003; Driver et al., 1999; Friesen & Kingstone, 1998; Kuhn & Kingstone, 2009; Ricciardelli et al., 2002). Such gaze-cues are able to induce automatic shifts of attention in the cued (i.e., gazed-at) direction and seem to resist voluntary control.

Dot-probe paradigms (MacLeod, Mathews, & Tata, 1986) make use of a similar principle to examine attentional biases. Participants fixate a central fixation cross and are presented with two peripheral stimuli for a brief amount of time. Subsequently, a probe appears at the location of one of these previously viewed stimuli and participants are instructed to respond to it quickly, for instance according to its location. If the probe succeeds the previously attended stimulus, reaction times will again be faster in line with the congruency effect. Herein, stimuli with emotional value (e.g., threatening words or angry faces) were shown to exert an attentional bias over simultaneously presented neutral stimuli. While previous studies conducted in clinical setups (e.g., investigating patients with anxiety disorders) yielded largely consistent evidence in favor of an attentional bias for emotional faces (e.g., Bradley, Mogg, White, Groom, & Bono, 1999), as well as threat-related words (MacLeod et al., 1986; Mogg, Mathews, & Eysenck, 1992) or stimuli (Kroeze & van den Hout, 2000), findings in healthy participants have been less robust (see e.g., Asmundson & Stein, 1994; Bar-Haim, Lamy, Bakermans-Kranenburg, & Ijzendoorn, 2007; Mogg et al., 2000). Schmukle (2005) even concluded that dot-probe paradigms in general are inadequate to measure attentional biases in non-clinical samples, as they have low test-retest reliability. However, Staugaard (2009) refined this statement by concluding that they might be unsuitable for individual differences research, but reliable for between-group designs to investigate different aspects of attention. Chapman, Devue and Grimshaw (2017) addressed these inconsistencies directly and state that electrophysiological evidence shows that healthy participants reliably attend to the emotional stimulus within such a dot-probe paradigm, but that this may only be a brief occurrence, potentially causing reaction times to vary across trials leading to low reliability. Their results depicted greater reliability for stimulus onset asynchronies (SOAs, i.e., delay between the onset of the cue and onset of the probe) below 300 ms. Bindemann and colleagues (2007) also investigated different SOAs and simultaneously presented participants with a face cue and an object cue, one on each side of a central fixation cross. A subsequently presented probe that required a manual response (either detection or discrimination) was presented at the same location as either the face or the object. When face and object cues were equally likely to predict a target, participants were faster to respond to probes that appeared in the same location as the face, suggesting that isolated faces capture covert attention. These results were largely independent of the SOA.

Although results of the previously mentioned studies investigating overt social attention using eye-tracking (e.g., Birmingham et al., 2008a; End & Gamer, 2017; Fletcher-Watson, Findlay, Leekam, & Benson, 2008) and covert social attention using spatial cueing paradigms (e.g., Bindemann et al., 2007; Kuhn, Tatler, & Cole, 2009) revealed comparable results regarding the preferential orienting towards social information, some researchers suggest that overt and covert shifts of attention are completely independent of one another (Posner, Cohen, & Rafal, 1982; Posner & Petersen, 1990) or are only partially overlapping (Corbetta et al., 1998). For instance, electrophysiological studies using steady-state visual-evoked potentials (SSVEPs) have shown worse task performance (i.e., higher error rates and worse classification) and lower SSVEP amplitudes when eye movements were disallowed as opposed to when they were permitted (e.g., Kelly, Lalor, Finucane, & Reilly, 2004; Treder & Blankertz, 2010), suggesting that overt and covert orienting processes can be dissociated. Specifically, there seems to be a decrease in signal power in covert compared to overt attention (Ordikhani-Seyedlar, Sorensen, Kjaer, Siebner, & Puthusserypady, 2014). Overt attention reflected respective SSVEPs frequency entrainment in the primary visual cortex, while covert attention initiated a shift towards parietal and frontal areas, that is, recruiting higher cognitive functions. Neuroimaging studies revealed inconsistent results, as some attested to activation of identical brain regions during covert and overt orienting (De Haan, Morgan, & Rorden, 2008), while others emphasized a dissociation of involved brain regions: cortical activations were similar for overt as well as covert attention, whereas the functional coupling changed as a function of task (goal-directed and stimulus-driven; Fairhall, Indovina, Driver, & Macaluso, 2016). These deviating results raise questions about the dissociation between these two attentional mechanisms. The previously mentioned studies which directly compared covert and overt attention mainly investigated general attentional mechanisms by using simplistic stimuli such as geometric shapes (De Haan et al., 2008; Fairhall et al., 2009; Ordikhani-Seyedlar et al., 2014) or checkerboards (Kelly et al., 2004). Only few studies have compared overt and covert attention in parallel in a social context (e.g., Domes et al., 2013) of which even less engaged naturalistic stimuli. An exception is the study of Kuhn, Tatler and Cole (2009), who manipulated overt attention using misdirection tricks in a video depicting a magician. However, the authors considered covert attention as a detection of the trick in this gaze-cueing paradigm, which only few participants depicted (for similar approaches, see also Kuhn, Teszka, Tenaw, & Kingstone, 2016; Gustav Kuhn & Teszka, 2017). Thus, in order to further our understanding of the mechanisms of social attention, its overt and covert aspects should be compared more directly than in previous research.

The present study aims to provide such a comparison by addressing the following issues: Firstly, we directly compared overt and covert social attention using identical stimuli in two independent studies in order to examine whether these aspects of attentional orienting are differentially sensitive to the presence of others (e.g., Kuhn, Teszka, Tenaw, & Kingstone, 2016; Risko, Richardson, & Kingstone, 2016). Secondly, our stimuli comprised naturalistic scenes containing a human being in different real-world situations in contrast to other studies (especially those investigating covert attention) that employed simplified (e.g., iconic images, geometric shapes) or isolated stimuli (lack of contextual information, e.g., face without a body). To the best of our knowledge, previous research has not yet examined covert shifts of attention towards social information within such more naturalistic viewing conditions. Finally, many studies indicating a social bias have not conclusively addressed the potential influence of low-level stimulus properties on attentional orienting (e.g., Castelhano, Wieth, & Henderson, 2007; DeAngelus & Pelz, 2009; Fletcher-Watson et al., 2008). Studies such as those of End and Gamer (2017) highlight the importance of accounting for low-level saliency, not only to rule out the possibility that a social bias is caused by higher visual saliency of the social features in a scene, but also to explain a worse prediction of gaze behavior by saliency in naturalistic scenes in which a social feature is depicted. Our stimulus set takes saliency distributions into account as we compared algorithms of center-surround low-level feature predictions (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998) as well as graph-based predictions (GBVS; Harel, Koch, & Perona, 2007), low-level isotropic contrast features (ICF; Kümmerer, Wallis, Gatys, & Bethge, 2017), and probabilistic models based on pre-trained deep neural network features (DeepGazeII; Kümmerer et al., 2016) to account for low-level and high-level image features in their efficiency to predict covert orienting as well as fixation patterns. In summary, we examined in two experiments whether social information is covertly and/or overtly selected over non-social information when embedded in real-world scenes.

EXPERIMENT 1: COVERT SOCIAL ATTENTION

In order to examine covert attentional orienting towards social information, we utilized the dot-probe task as an established experimental procedure. In the original version of the task (MacLeod et al., 1986), two cues are briefly presented simultaneously on the computer screen followed by a probe at one of the former cue locations that prompts a behavioral response. By manipulating cue categories (e.g., upright vs. inverted faces, see Langton, Law, Burton, & Schweinberger, 2008) and analyzing response times separately for cued and uncued locations, one can draw conclusions about which stimulus preferentially attracts attention. In the current study, we deviated from this standard procedure by presenting complex visual scenes that spanned the entire screen instead of separately stimulating both visual hemispheres. A single human being was present on one half of the picture (referred to hereafter as the *social side of the stimulus*) whereas the other half did not contain a human being (referred to hereafter as the *non-social side of the stimulus*). In the original version of the dot-probe task, a small dot is presented as the probe and participants have to indicate its location by button press (MacLeod et al., 1986). This so-called detection-variant has been criticized since participants could simply respond to the presence or absence of the probe at a specific location instead of distributing their attention to both locations (Bradley, Mogg, Falla, & Hamilton, 1998). Therefore, several studies have used a discrimination-variant of the task where participants do not simply respond to the probe location but instead have to respond to the identity of the probe, for example by deciding whether the probe matches a pair of dots in horizontal or vertical direction, or whether a letter displayed as a probe is an "E" or an "F" (e.g., Chen, Ehlers, Clark, & Mansell, 2002). The current experiment consisted of both, the detection and discrimination-variant to determine attentional biases triggered by briefly shown (200 ms) naturalistic scenes. The detection task required participants to indicate the location of a presented probe (*), while in the discrimination-variant, participants had to identify whether the letter "E" or "F" was presented at one of two locations. Participants were instructed to respond via keyboard to the probe as quickly and as accurately as possible, while their eyes were to remain centrally fixated throughout the experiment. They were also aware of the fact that their response times were measured and their fixations controlled via a camera. By restricting eye movements through instruction to fixate a central cross, this setup allowed us to examine covert attentional orienting towards social information under naturalistic viewing conditions. Although the given tasks differed in complexity (discriminating letters being more demanding than detecting of a location), we hypothesized that an attentional bias towards the side of a depicted human being will result in a robust congruency effect for both manipulations, yielding faster reaction times and lower error rates for probes presented on the social side, as opposed to the non-social side of the stimulus, independent of task variant.

METHODS

PARTICIPANTS

The study was conducted according to the principles expressed by the Declaration of Helsinki. The required sample size (*n* = 27) was estimated to allow for detecting a medium-sized advantage of the social stimulus side (*d* = 0.50) with a power of at least .80 at an alpha level of .05. To account for potential drop-outs, we recruited thirty-four participants through the University of Würzburg's Human Participant Pool and personal recruitment. Since all participants provided usable data, the final sample consisted of 9 males and 25 females with a mean age of 21.85 years (*SD* = 2.72 years). Each participant had normal or corrected-to-normal vision and provided written informed consent, and received course credit for participation.

APPARATUS

The experiment was programmed with Presentation® (Neurobehavioural Systems Inc., Version 18.1). Stimuli were presented on a 24” display (Fujitsu B24T-7 LED) with a size of 53.1 x 29.9 cm, a resolution of 1920 x 1080 pixels, and refresh rate of 60 Hz. A chin-rest with a forehead bar was mounted to the table and ensured a constant viewing distance of 72 cm from the eyes to the center of the monitor. A webcam (Microsoft LifeCam VX – 1000) was installed on the top of the display screen and the resulting video image could be observed on a second computer that was not visible to the participant. The experiment was conducted in a secluded room with constant lighting conditions.

STIMULI

Stimuli consisted of 60 naturalistic social scenes, which were acquired through the Nencki Affective Picture System (Marchewka et al., 2014), EmoPics (Wessa et al., 2010) as well as internet searches (e.g., Google picture search, flickr). Scenes were cropped if necessary, contrast and luminance were adjusted manually to increase the homogeneity of the stimulus set, and image sizes were adjusted to 1200 x 900 pixels (26.0° x 19.6° of visual angle). Each of the images contained a single person positioned on either the left side or the right side of a complex naturalistic scene (for an example image as well as the overlap of social regions of interest for all images, see Figure 2.1). All original images could be flipped horizontally without compromising their appearance (i.e., stimuli did not include text or other elements that change meaning when mirrored). The distribution of visual saliency across image sides was controlled to avoid confounds of the location of social elements and low-level conspicuity. Visual saliency maps were computed before the final selection of images using the procedure suggested by Itti and Koch (Itti & Koch, 2000; Itti et al., 1998). This algorithm extracts three different low-level properties (i.e., orientation, intensity and color contrast) to reveal locations that stand out from the background in terms of these features and combines them into a single topographic saliency map. Afterwards, the final set of images was compiled such that 30 pictures had higher visual saliency on the social side (i.e., the half containing a person) and 30 pictures higher visual saliency on the non-social side (i.e., the half opposite to the person) on the image. Six additional images were selected using the same criteria and served as a training set that was used to familiarize participants with the task.

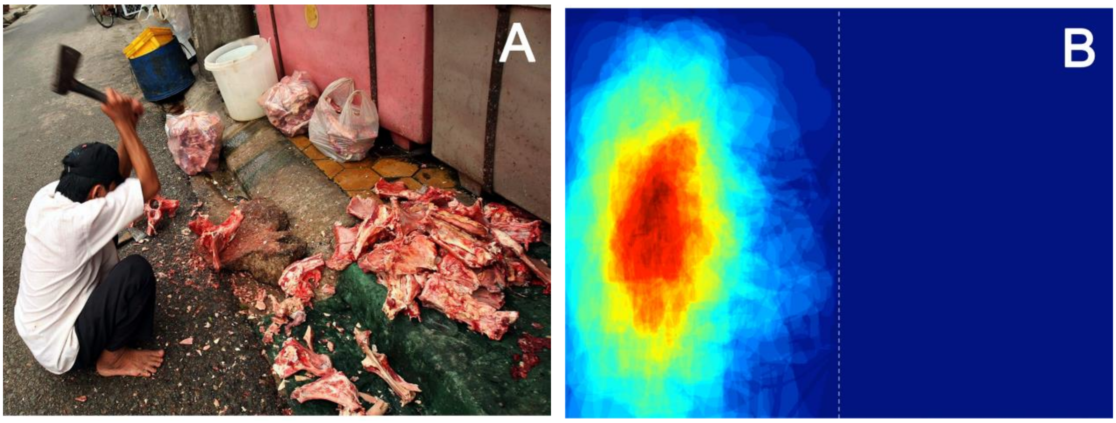

Figure 2.1. Illustration of stimulus characteristics. A) Original scene with the human being on the left half of the picture (image taken from the Nencki Affective Picture System (NAPS; Marchewka et al., 2014) with permission for fair use). B) Overlap of head and body regions of interest for all stimuli with the social information on the left half of the picture. Warm colors indicate high overlap of regions.

PARADIGM

In the current experiment, participants completed both a detection- and a discrimination-variant of the dot-probe task in separate blocks whose order was counterbalanced across participants. In each block, all 60 pictures were presented in their original orientation as well as in a horizontally flipped version, and each of these two versions was shown once with the probe on the social side and once with the probe on the non-social side of the stimulus. Thus, the side of the social feature, the relative visual saliency of the social side as well as the probe location were counterbalanced. This amounted to a total of 240 trials per condition. Trial order was randomized for each subject with the restriction that the same image could not appear in successive trials. After 120 trials, participants were given the opportunity to take a short self-paced break.

In the detection-variant, an asterisk (*) was used as probe and participants had to indicate its location using their index fingers which were placed on the keys “A” (left) or “#” (right) of a standard QWERTZ keyboard. In the discrimination-variant, the letter “E” or “F” was used as probe. The probe was presented 6.6° of visual angle from the screen’s center. Both types of probes (asterisk/letter) were shown in white on a dark grey background in the font Arial with a size of 35 pt resulting in a visual angle of 0.4° x 0.4° for the asterisk and 0.6 x 0.8° for the letters. We randomly determined which letter was assigned to which trial but ensured an equal distribution of both letters within each set of original and flipped picture versions. Participants had to identify the letter using the keys "H" and the space bar of a standard QWERTZ keyboard using the thumb or index finger, respectively, of their dominant hand. The assignment of keys to letters was counterbalanced across participants for the discrimination task. The keys for both dot-probe variants were deliberately chosen in horizontal and vertical alignment to avoid confounds related to the "Simon Effect" (Simon & Rudell, 1967), which results in shorter reaction times for spatially congruent keys. Responses had to be given as quickly and accurately as possible.

PROCEDURE

Participants were given task instructions and told to keep their eyes in the center of the display at all times (covert viewing) indicated by a fixation cross at the beginning of each trial. It was pointed out to them that a webcam was installed on top of the monitor to ensure continuous fixation. The experimenter monitored the webcam video during the experiment in order to exclude participants with excessive eye movements. According to this visual assessment, all participants adhered to the instructions and showed no illicit eye movements during the task.

Each trial started with a fixation cross presented for 1 s, followed by an image that was shown for 200 ms. This presentation duration was chosen to avoid inhibition of return, which describes a briefly enhanced attention at a cued peripheral location that is subsequently impaired for longer viewing durations (> 255 ms; see Klein, 2000), but also to emphasize exogenous attention by reducing the influence of later (>300 ms, see Carrasco, 2011) top-down mechanisms. Directly after image offset, the probe appeared and remained on the screen until participants pressed a response key. After the behavioral response, the fixation cross reappeared in the middle of the screen for a random duration between 500 and 2000 ms before the next trial started with the presentation of the fixation cross for 1 s, and a subsequent stimulus presentation. The variable inter-trial-interval was chosen to avoid anticipation effects that may occur with consistent time intervals, enabling a certain preparation for the onset of the stimulus. For this reason, we also refrained from using a blank screen, but upheld the presentation of a fixation cross to add to the unpredictability of the stimulus onset. Before starting with the detection- or discrimination-variant of the task, participants were familiarized with the procedure by completing 24 training trials with a different set of images. These trials were excluded from further analyses. Throughout the whole task, no error feedback was given but the experimenter observed the participants' responses during the training trials and gave advice if necessary (e.g., regarding the correct key assignment).

DATA ANALYSIS

Reaction times and error rates were averaged per participant across trials separately for each cell of the 2 x 2 design consisting of the factors dot-probe variant (detection- vs. discrimination-task) and congruency (congruent vs. incongruent, i.e., probe presented on the social vs. the non-social side of the stimulus). Erroneous responses as well as outliers of response latencies were excluded from the calculation of mean reaction times. A value was defined as outlier when it deviated more than 2.5 *SDs* from the mean per subject and dot-probe variant. For the detection variant, this amounted to *M* = 2.5% (*SD* = 1.0%) and for the discrimination variant to *M* = 2.9% (*SD* = 0.8%) outliers. The influence of the experimental manipulations on error rates and reaction times was analyzed using 2 x 2 repeated-measures analyses of variance (ANOVAs) with the factors dot-probe variant and congruency.

With our stimulus selection, we tried to control for systematic biases in visual saliency between the social and the non-social side of the images. Indeed, when relying on the algorithm of Itti and Koch (Itti & Koch, 2000; Itti et al., 1998), visual saliency was balanced between both sides as a binary variable (i.e., the social side had a higher mean saliency in 50% of the images) as well as numerically with mean saliency values amounting to *M* = 70.42 (*SD* = 15.81) for the social and *M* = 67.04 (*SD* = 20.99) for the non-social side ($t_{(59)}$ = 1.18, *p* = .24, Cohen's *d* = .15). However, recent algorithms proved to be more successful in the prediction of visual fixation patterns than the algorithm that was used here for stimulus selection (see e.g., Borji & Itti, 2013; Judd, Durand, & Torralba, 2012). In order to ensure stability of our results for better performing algorithms, we did a post-hoc analysis of our data using the Graph-Based Visual Saliency algorithm (GBVS; Harel et al., 2007) which was previously shown to outperform the Itti and Koch algorithm in the prediction of human gaze (Judd et al., 2009). Additionally, we incorporated the Intensity Contrast Features (ICF) model, which achieves top performance among all low-level baseline models (Kümmerer et al., 2017). Finally, we took into account the DeepGazeII model, which makes use of convolutional filters that have been pre-trained on object recognition (Kümmerer, Theis, & Bethge, 2014a). Herein, the algorithms based on low-level saliency (Itti, GBVS and ICF) are considered cognitive models that use basic feature channels to create conspicuity maps, which are combined to generate predictions of gaze behavior in the form of saliency maps (Borji & Itti, 2013). DeepGazeII is considered a state-of-the-art model for saliency prediction using features from the VGG-19 deep neural network (Simonyan & Zisserman, 2014) pre-trained on the SALICON dataset (Jiang, Huang, Duan, &

Zhao, 2015), showing high performance in the prediction of fixation patterns. The well-established MIT saliency benchmark offers a general ranking of saliency models (saliency.mit.edu, Bylinskii et al., 2016), while other comparisons were made with regard to information gain (quantification of the reduction in uncertainty of gaze allocation; see Kümmerer, Wallis, & Bethge, 2015; Kümmerer et al., 2017). Herein, the algorithms used in this study show increasing information gain in the following order: Itti, GBVS, ICF, DeepGazeII (for an illustration of the visual saliency maps generated by these algorithms see Figure 2.2B). Whereas the ICF algorithm did not reveal enhanced mean saliency on the social (*M* = 25.01, *SD* = 8.80) as compared to the non-social side of the stimulus (*M* = 24.27, *SD* = 11.28; $t_{(59)} = 0.61$, $p = .54$, $d = .08$), different results were obtained for the GBVS as well as the DeepGazeII algorithm. These latter two models both revealed higher mean saliency on the social (GBVS: *M* = 62.14, *SD* = 16.57; DeepGazeII: *M* = 24.74, *SD* = 11.28) as compared to the non-social side (GBVS: *M* = 56.77, *SD* = 15.90; DeepGazeII: *M* = 8.51, *SD* = 4.25) of our stimuli (GBVS: $t_{(59)} = 2.65$, $p = .01$, $d = .34$; DeepGazeII: $t_{(59)} = 10.66$, $p < .001$, $d = 1.38$). In order to ensure that the current results were not merely driven by differences in the saliency distribution between the social and the non-social side of our stimuli, we repeated all analyses of the behavioral data for a subset of pictures for which the mean saliency was lower as well as the area of peak saliency (highest 5% of salience values) was smaller on the social as compared to the non-social side (GBVS: *n* = 19, ICF: *n* = 24, DeepGazeII: *n* = 9 images; see Figure 2.2D). This measure not only compares saliency of the social versus the non-social stimulus sides, but also takes into account local saliency distributions that specifically drive early attentional selection.

All data processing and statistical analyses were performed using the statistical programming language R (version 3.2.3; R Core Team, 2016). An a priori significance level of $\alpha = .05$ was applied and two-tailed testing was used throughout. For all ANOVAs, the generalized η^2 (Bakeman, 2005) is reported as effect size estimate.

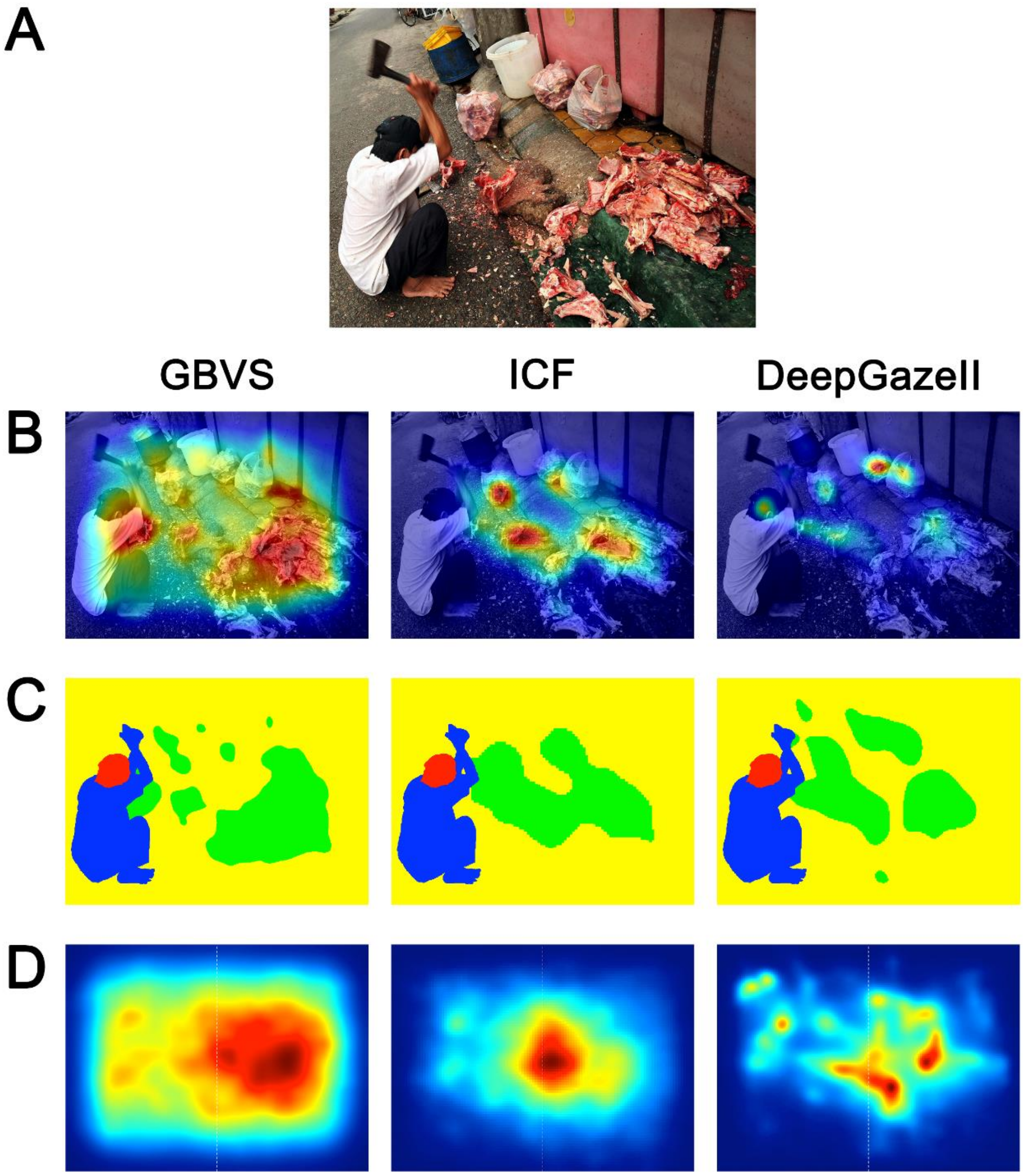

Figure 2.2. Illustration of image characteristics. A) Original scene (image taken from the Nencki Affective Picture System (NAPS; Marchewka et al., 2014) with permission for fair use). B) Saliency maps for three different algorithms used for visual saliency analyses (GBVS: Graph-Based Visual Saliency by Harel et al., 2007, ICF: Intensity Contrast Features by Kümmerer et al., 2017, DeepGazeII by Kümmerer, Theis, & Bethge, 2016). Warm colors indicate areas of high saliency. C) Exemplary regions of interest for head (red) and body (blue) as well as areas of higher (green) and lower visual saliency (yellow), see Experiment 2. D) Mean visual saliency map of all images with the human being on the left half of the picture with higher mean and peak visual saliency on the non-social right half of the image with warmer colors indicating higher saliency.

RESULTS

Reaction times for the full stimulus set (Table 2.1) were higher for the more challenging discrimination- as compared to the detection-variant of the dot-probe task, resulting in a main effect of dot-probe variant ($F_{(1,33)} = 517.11$, $p < .001$, $\eta^2 = .692$). In addition, we observed a significant main effect of congruency ($F_{(1,33)} = 7.55$, $p = .01$, $\eta^2 = .001$) depicting longer reaction times for incongruent trials in which the social feature appeared on the opposite side of the probe. The interaction effect was not significant ($F_{(1,33)} = 0.03$, $p = .87$, $\eta^2 = .000$), implying that the facilitation effect of social content was similar in both dot-probe variants.[1] These results were comparable when restricting the stimulus set to pictures that contained a non-social side with higher mean and peak saliency in terms of low-level features (Table 2.1). For the GBVS as well as the ICF algorithm, we obtained significant main effects of dot-probe variant ($F_{(1,33)} = 529.28$, $p < .001$, $\eta^2 = .697$ and $F_{(1,33)} = 493.41$, $p < .001$, $\eta^2 = .680$, respectively), and congruency ($F_{(1,33)} = 15.67$, $p < .001$, $\eta^2 = .009$ and $F_{(1,33)} = 16.45$, $p < .001$, $\eta^2 = .006$, respectively), but no statistically significant interaction of both factors ($F_{(1,33)} = 0.06$, $p = .81$, $\eta^2 = .000$ and $F_{(1,33)} = 2.07$, $p = .16$, $\eta^2 = .001$, respectively). However, for the reduced set determined by DeepGazeII, only a significant main effect of dot-probe variant emerged ($F_{(1,33)} = 464.35$, $p < .001$, $\eta^2 = .673$). Neither the main effect of congruency ($F_{(1,33)} = 2.38$, $p = .13$, $\eta^2 = .004$), nor the interaction of both factors ($F_{(1,33)} = 1.44$, $p = .24$, $\eta^2 = .002$) reached statistical significance. As depicted in Figure 2.3A, social information reliably modulated response times in both dot-probe variants, irrespective of its low-level visual saliency. When incorporating high-level feature predictions as modeled by the DeepGazeII algorithm, however, the congruency effect was no longer statistically significant. As indicated by the high variability of congruency effects in this condition, this might at least in part be related to the small number of suitable stimuli and correspondingly the small number of trials in the reduced stimulus set.

---

[1] As participants saw each figure twice for every condition, we conducted an explorative post-hoc analysis to investigate potential familiarity effects that may be reflected in reaction times. The 2 x 2 x 2 repeated-measures ANOVA incorporating the factor repetition in addition to the factors dot-probe variant and congruency did indeed reveal a main effect of repetition ($F_{(1,33)} = 9.35$, $p = .004$, $\eta^2 = .002$), indicating that reaction times were faster when the image had already been presented before. Importantly however, we obtained a main effect of congruency ($F_{(1,33)} = 8.33$, $p = .007$, $\eta^2 = .001$), but no statistically significant interaction between repetition and congruency ($F_{(1,33)} = .79$, $p = .378$, $\eta^2 < .001$). Thus, response times were reliably smaller on congruent trials and this pattern was not significantly affected by familiarity.

Table 2.1. Response times and error rates for the full as well as the reduced stimulus sets by algorithm (GBVS, ICF, DeepGazeII) as a function of dot-probe variant (detection vs. discrimination) and congruency of probe location and social information on the image (congruent vs. incongruent).

	Response times (ms)		Error rates (%)	
	Congruent	Incongruent	Congruent	Incongruent
Full stimulus set				
Detection task	407.25 (53.15)	411.41 (51.72)	2.03 (2.07)	3.16 (3.47)
Discrimination task	592.69 (72.44)	597.39 (71.30)	4.04 (3.31)	4.73 (3.13)
Reduced stimulus set				
GBVS				
Detection task	404.97 (47.31)	415.95 (51.71)	2.17 (3.46)	3.17 (3.48)
Discrimination task	590.16 (68.94)	602.44 (76.10)	4.64 (5.50)	4.26 (4.25)
ICF				
Detection task	407.58 (52.36)	414.12 (53.02)	2.27 (3.30)	2.94 (3.21)
Discrimination task	590.03 (71.59)	603.91 (78.29)	4.11 (4.63)	4.17 (3.28)
DeepGazeII				
Detection task	405.28 (49.84)	419.76 (60.21)	2.61 (4.98)	4.25 (5.65)
Discrimination task	602.84 (79.28)	605.76 (77.30)	4.90 (5.76)	4.08 (5.35)

Note. Mean values across participants are depicted with standard deviations in parentheses. Reduced stimulus set: Values were calculated for a subset of 19 out of 60 stimuli for the GBVS algorithm, 24 out of 60 stimuli for the ICF algorithm and 9 out of 60 for the DeepGazeII algorithm that exhibited higher mean and peak visual saliency on the non-social side of the image.

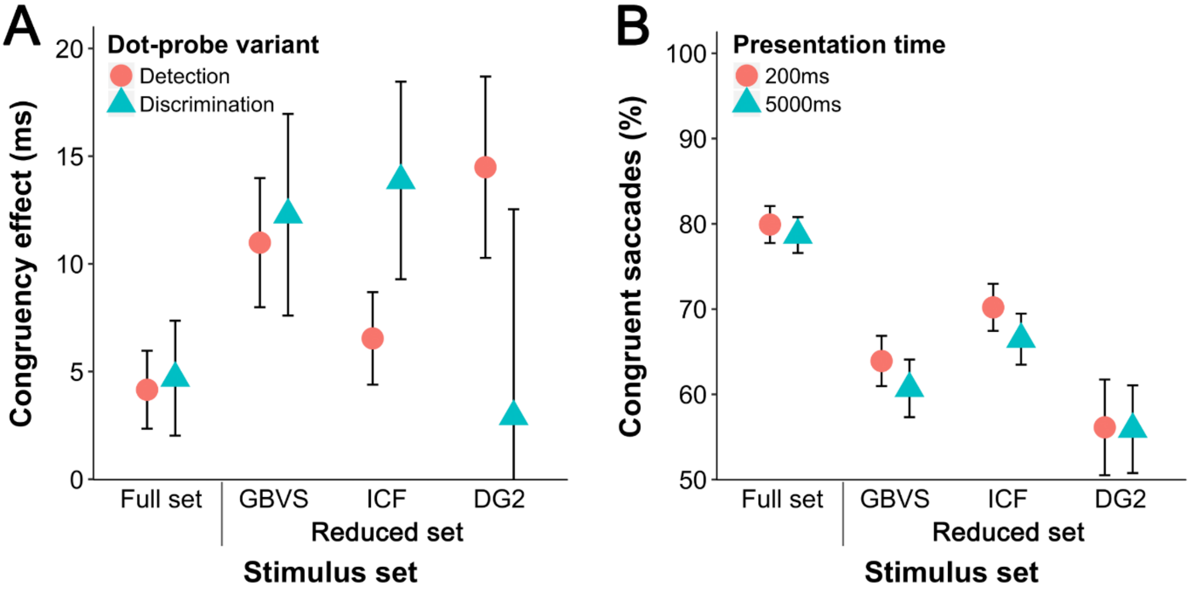

Figure 2.3. A) Mean response time differences between congruent (i.e., the probe appeared on the social half of the picture) and incongruent trials (i.e., the probe was presented on the non-social half of the picture) in Experiment 1 as a function of dot-probe variant (detection vs. discrimination). B) Relative frequency (%) of the first saccade being congruently directed towards the social half of the stimulus in Experiment 2 as a function of presentation time (200 vs. 5000 ms). Both effects are plotted for the whole set of pictures as well as for the reduced sets with a higher mean and peak saliency on the non-social image half according to three different algorithms (GBVS: Graph-Based Visual Saliency, ICF: Intensity Contrast Features, DG2: DeepGazeII). Error bars indicate standard errors of the mean.

Similar to the reaction time results, we also found a main effect of dot-probe variant on error rates for the full stimulus set ($F_{(1,33)} = 21.89, p < .001, \eta^2 = .082$), indicating a larger number of errors in the discrimination compared with the detection variant (Table 2.1). In line with the reaction time data, we also obtained a main effect of congruency ($F_{(1,33)} = 10.57, p = .003$, $\eta^2 = .022$), depicting lower error rates for congruent trials as opposed to incongruent trials. Again, the interaction of both factors did not reach statistical significance ($F_{(1,33)} = 0.53, p = .47$, $\eta^2 = .001$). For the reduced stimulus sets as determined by the GVBS and the ICF algorithms, only the main effect of dot-probe variant was found to be significant ($F_{(1,33)} = 7.29, p = .01, \eta^2 = .043$ and $F_{(1,33)} = 8.47, p = .006, \eta^2 = .043$, respectively). Neither the main effect of congruency ($F_{(1,33)} = 0.38, p = .54, \eta^2 = .001$ and $F_{(1,33)} = 0.85, p = .36, \eta^2 = .003$, respectively), nor the interaction of both factors ($F_{(1,33)} = 1.58, p = .22, \eta^2 = .007$ and $F_{(1,33)} = 0.45, p = .51, \eta^2 = .002$, respectively) reached statistical significance. For the reduced stimulus set as determined by the DeepGazeII algorithm, we did not obtain statistically significant effects: main effect dot-probe variant ($F_{(1,33)} = 1.32, p = .26, \eta^2 = .010$), main effect congruency ($F_{(1,33)} = 0.27, p = .60, \eta^2 = .001$), interaction of dot-probe variant and congruency ($F_{(1,33)} = 2.05, p = .16, \eta^2 = .013$).

DISCUSSION

This study supports the notion that covert attention is preferably allocated to social features in complex naturalistic scenes. We observed reliably shorter reaction times and fewer errors for probes appearing on the side of the stimulus depicting a human being than for the non-social half, suggesting that social elements capture covert attention. Even though reaction time differences between congruent and incongruent trials were relatively small for the whole stimulus set, this effect persisted for both variants of the social dot-probe paradigm with generally increasing demands of cognitive resources for the discrimination as compared to the detection variant. By carefully preselecting pictures with respect to the distribution of visual saliency and by restricting the data set to trials with higher mean and peak saliency on the non-social side of the pictures in post-hoc analyses, we furthermore dissociated between influences of low-level and high-level features in their efficiency to predict this social bias. Algorithms using low-level features only for saliency prediction demonstrated that preferential shifting of covert attention was largely independent of these image features. Moreover, the congruency effect in reaction times even seemed to be stronger for the reduced set of stimuli as compared to the full set, which discounts the influence of low-level saliency further. On these grounds, we suggest that social aspects are preferentially attended, even when bottom-up influences should draw attention towards the opposite side. Algorithms involving a pre-training of deep neural networks seem to result in better predictions, confirming findings of Kümmerer and colleagues (2017) that DeepGazeII performs particularly well on images containing faces. This further suggests that human gaze may be modelled more successfully using high-level features, especially with regard to reflexive social attention. However, it must be mentioned that the number of stimuli in the reduced set based on the DeepGazeII algorithm (*n* = 9) amounted to only 15% of the total dataset, necessitating a reexamination of the effects elicited here. Nevertheless, *o*ur data also replicate and extend the findings of studies investigating preferential attention towards faces in comparison with other objects (e.g., Bindemann et al., 2007; Devue, Belopolsky, & Theeuwes, 2012) and confirm that there is indeed a preferential selection of social information even when such features are irrelevant to the task at hand. Additionally, our findings illustrate that such effects can be observed even when using complex naturalistic stimuli instead of isolated or simplified ones (see Kingstone, Smilek, Ristic, Friesen, & Eastwood, 2002).

Although dot-probe paradigms have yielded consistent results regarding attentional biases towards emotionally salient stimuli in the past for clinical samples (Bradley et al., 1999; Kroeze & van den Hout, 2000; MacLeod et al., 1986; Mogg et al., 1992), findings in healthy participants have been shown to be less robust (see e.g., Mogg et al., 2000; Bar-Haim, Lamy, Bakermans-Kranenburg, & Ijzendoorn, 2007). As mentioned previously, Schmukle (2005) even concluded that dot-probe paradigms in general are inadequate to measure attentional biases in non-clinical samples. Our paradigm differs from other dot-probe paradigms in several ways, somewhat refuting these doubts. First, based on the results of Chapman and colleagues (2017) that the reliability of attentional biases as measured by the dot-probe paradigm diminishes quickly with increasing SOA, we deliberately kept the SOA below 300ms in the first experiment to specifically target initial and brief attention capture. Second, we did not investigate an attention bias towards threatening stimuli that may potentially generate an avoidance behavior in healthy subjects (Bar-Haim et al., 2007). Third, our paradigm did not use traditional directional or exogenous cues to guide attention that participants could have resisted by actively suppressing them when preparing for the upcoming probe presentation. In conclusion, the results of our experiment speak in favor of an early social bias for covert attention, which is evident without explicit cueing. Nevertheless, since our manipulation only allowed for automatic covert orienting, reflexive overt behavior as well as sustained attention processes should be further examined for this social stimulus set. In order to investigate this reliably and generalize the current findings, we conducted a second experiment using eye-tracking with the same set of images to examine whether patterns of overt visual exploration are consistent with the observation of covert attentional orienting towards social cues.

EXPERIMENT 2: REFLEXIVE AND SUSTAINED OVERT SOCIAL ATTENTION

In Experiment 2, a separate group of participants had their eye movements tracked as they freely viewed each of the scenes that were also used in Experiment 1. Participants were given no explicit task in this experiment and were instructed to freely view the stimuli like in a magazine. In order to differentiate between reflexive and more sustained aspects of visual attention, we showed half of the scenes for a very brief duration (200 ms) and quantified whether the first saccade was directed towards the social or non-social half of the image. The other stimuli were shown for longer durations (5000 ms) to allow for a detailed exploration of the scene. Participants were aware of the fact that the images would be shown to them for different durations and were shown practice trials with a different set of stimuli before the experiment started. We hypothesized to observe a bias towards social information in reflexive attentional orienting as well as an increased exploration of faces in measures of sustained attention (see Birmingham et al., 2008a; End & Gamer, 2017; Rösler et al., 2017).

METHODS

PARTICIPANTS

In order to account for potential drop-outs due to low data quality, we recruited 33 participants (10 males and 23 females) with a mean age of 25.64 years (*SD* = 4.59 years). Participation was compensated in the form of course credit or monetary compensation. Since all data were of sufficient quality, the analyses are based on the whole sample of 33 participants. All other details were similar to Experiment 1.

APPARATUS

Using Presentation® (Neurobehavioural Systems Inc., Version 18.1), stimuli were displayed on an LG 24MB 65PY-B (24") screen (51.7 x 32.3 cm) with a resolution of 1920 x 1200 pixels and a refresh rate of 60 Hz. Eye movements were tracked in a dimly lit laboratory using a mounted EyeLink 1000 Plus system (SR Research Ltd., Ottawa, Canada). The camera had a 25 mm lens and the system was set to a sampling rate of 1000 Hz. Although viewing was binocular, only the right eye was tracked with a viewing distance of 50 cm from the monitor. Participants' head position was stabilized using a chin rest and a forehead bar.

STIMULI

We used the same 60 stimuli (1200 x 900 pixels) with counterbalanced low-level saliency on the social and the non-social side as in Experiment 1. With the current setup, the viewing angle amounted to 35.8° x 27.2° for each stimulus.

PARADIGM

The experiment consisted of a 2 x 2 design with the within-subject factors presentation time (200 ms vs. 5000 ms) and laterality of the social information (left vs. right). As described in the methods section of Experiment 1, low-level visual saliency according to the algorithm of Itti and Koch (Itti & Koch, 2000; Itti et al., 1998) was balanced across stimuli. Each image was only shown once during the experiment and the association of each stimulus to the experimental conditions was determined randomly for each participant with an equal number of stimuli in each cell of the 2 x 2 design. Binary low-level saliency was balanced across presentation times such that half of the images had a higher saliency on the social and half on the non-social side. The order of all 60 trials was also randomized with the restriction that the same presentation time was not used in more than three successive trials to reduce expectation effects, which might potentially encourage rapid saccade initiation for short stimulus presentations or a lack thereof for consistently long presentation durations.

PROCEDURE

Participants were told that they would be shown a number of images following a fixation cross. They were instructed to look at the fixation cross whenever present on the screen, but that they were otherwise free to examine the images when presented. Before the experiment, participants were calibrated on the eye-tracker using a nine-point calibration sequence. After successful validation, the trials began. Prior to each trial, a fixation cross was displayed for 1 s. The fixation cross was followed by the presentation of one of the stimulus images for a duration of either 200 ms or 5000 ms. For brief stimulus durations, a uniformly gray screen was shown for 1800 ms before the fixation cross reappeared again. Between trials, the fixation cross was presented for a randomly chosen period between 1000 and 3000 ms. Participants were familiarized with the procedure by completing 6 training trials (3 with a 200 ms and 3 with a 5000 ms presentation time) with a different set of images. These trials were excluded from further analyses.

DATA ANALYSIS

Eye movements were parsed into saccades and fixations using thresholds of 30°/s for velocity and $8000°/s^2$ for acceleration for saccade detection. Fixations were defined as time intervals between saccades. A drift correction with reference to a baseline period of 300 ms before stimulus onset (i.e., during the presentation of the fixation cross) was accomplished for all fixations and saccades. To define outliers of baseline coordinates for fixation analyses, we used a recursive outlier removal procedure, which was applied separately for x- and y-coordinates (see also e.g., End & Gamer, 2017; Flechsenhar & Gamer, 2017). Hereafter, the highest and lowest baseline coordinates were temporarily removed, and the mean and standard deviation were calculated for the remaining data for each participant. An interval bounded by 3 *SDs* from the mean was set and if any of the two temporarily excluded values fell outside of this interval, the value was removed permanently. If a data point fell within the interval, it was returned to the data set. This procedure was continued until no more data points were discarded permanently. Subsequently, the baseline position data of all scenes including a removed x or y baseline coordinate or missing baseline data (117 trials in total; average number of trials per participant: *M* = 3.54, *SD* = 3.69) were replaced by the individual mean of all scenes with a valid baseline position.

Three different analyses were carried out on these data. First, we analyzed the direction of the first saccade after stimulus onset as a measure of reflexive attentional orienting (see also e.g., Rösler, End, & Gamer, 2017; Scheller, Büchel, & Gamer, 2012). In general, saccade latencies varied between 1 and 3373 ms after stimulus onset (*M* = 328.61 ms; *SD* = 308.91 ms) with a minority of 26.48% starting before 200ms. Thus, pictures in the condition with a short presentation time only served to trigger saccades that mostly occurred after stimulus offset when a uniformly gray screen was shown. In order to ensure that saccades were related to initial processing of the stimulus, we only scored saccades when their latencies ranged between 150 ms and 1000 ms (cf., Rösler et al., 2017). Participants made saccades on the majority of trials (merely 4.72% of all trials did not contain saccade information), even though participants did not receive explicit instructions to initiate eye movements rapidly. Trials were excluded from the analyses when an invalid baseline was detected (i.e., when we could not ensure that participants fixated the center of the screen at stimulus onset) or when saccades were missing or occurring outside the scoring interval of 150 ms to 1000 ms (238 invalid trials in total; average number per participant: *M* = 8.58, *SD* = 6.12). Thereafter, the saccades were classified according to whether they were directed towards the social side of the stimulus (congruent saccades) or the non-social one (incongruent saccades). We then analyzed the proportion of congruent saccades as a function of presentation time (200 ms vs. 5000 ms) using an ANOVA. For this analysis, we subtracted 50% from the respective proportions such that a significant intercept in the ANOVA model indicates a deviation from a chance distribution of saccade directions.

Second, we analyzed fixation densities as a measure of sustained visual attention for all trials with a presentation time of 5000 ms (see also End & Gamer, 2017; Flechsenhar & Gamer, 2017). For these fixation analyses, trials containing too many blinks were excluded (192 trials in total with a blink-free time period of less or equal than 80% of the whole trial, average number of excluded trials per participant: *M* = 5.82, *SD* = 5.26). Blinks in trials that contained more than 80% data were also removed, so that the analysis of the remaining trials consisted of blink-free data. Fixation maps were generated for each participant and stimulus by adding fixations weighted by their fixation durations in milliseconds to an empty two-dimensional matrix (1200 x 900 pixels). The first fixation was excluded when overlapping from the baseline period. The resulting map was then smoothed with a two-dimensional isotropic Gaussian kernel with a standard deviation of 32 pixels or 1° of visual angle using the R package *spatstat* (version 1.47.0, Baddeley, Rubak, & Turner, 2015). The resulting smoothing kernel comprised a width of 2° of visual angle (i.e., one standard deviation in the positive as well as the negative direction) to resemble the functional field of the human fovea centralis. Finally, fixation density maps were normalized to a range from 0 to 1. In order to investigate the distribution of fixations onto the social and non-social features of the stimuli, we introduced four regions of interest (ROIs) similar to our previous studies (see also End & Gamer, 2017; Flechsenhar & Gamer, 2017). The head and body regions were manually drawn using the image manipulation software GIMP (version 2.8.14; GNU Image Manipulation Program, The GIMP Team). Additionally, we defined ROIs for areas with lower and higher visual saliency. This was achieved by partitioning saliency maps as calculated by the GBVS (Harel et al., 2007), the ICF (Kümmerer et al., 2017), as well as the DeepGazeII algorithm (Kümmerer et al., 2014a). Saliency values smaller or equal to the eighth decile of the saliency distribution were defined as areas of lower saliency and the remaining image regions as areas of higher visual saliency. This calculation was only accomplished for image regions which had not already been assigned to the head or body ROI. Although the eighth saliency decile represents an arbitrary choice, this cut-off proved useful in previous studies and allowed for the identification of image regions without social information, but high low-level saliency (End & Gamer, 2017; Flechsenhar & Gamer, 2017). Figure 2.2B and 2.2C illustrate the saliency distribution and the resulting ROIs for one example picture. For each participant and scene, we determined the relative amount each ROI was fixated by calculating the sum of fixation density values for each ROI divided by the sum of fixation density values for the whole scene. This proportion score was then normalized by dividing it by the size of the ROI (in pixels) to control for increased fixations onto larger image areas (cf. Birmingham, Bischof, & Kingstone, 2009a). The mean for this relative area-normalized sum of fixation density was calculated for each ROI and participant, and a one-way ANOVA with the factor ROI (head, body, high saliency, low saliency) was carried out to examine the influence of image regions on fixation densities.

Third, we analyzed the location of the first five fixations post stimulus onset for long stimulus presentations. The relative frequency for each ROI was determined individually by dividing the frequency of fixations directed towards a ROI by the frequency that any ROI was fixated. Similar to the fixation densities, these scores were normalized with regard to the area of the respective ROI (see Flechsenhar & Gamer, 2017). The resulting data were analyzed using a 4 x 5 repeated-measures ANOVA with the factors ROI (head, body, high saliency, low saliency) and fixation number.

In general, data analyses were conducted similarly as in Experiment 1. First analyses included the full set of all 60 stimuli that had similar visual saliency on the social and non-social side as determined by the Itti and Koch algorithm (Itti & Koch, 2000; Itti et al., 1998). In this case, ROI definition was based on the GBVS algorithm to allow for a comparison with our previous studies (End & Gamer, 2017; Flechsenhar & Gamer, 2017; Rösler et al., 2017). A second set of analyses only included a reduced set of scenes with lower mean visual saliency and smaller areas of peak saliency (5% highest saliency values) on the social stimulus side as determined by three different algorithms: the GBVS algorithm (Harel et al., 2007), the ICF (Kümmerer et al., 2017) and the DeepGazeII algorithm (Kümmerer et al., 2014a). For these analyses, we did not only rely on the visual saliency algorithms for stimulus selection but also used the respective saliency maps for ROI definition. Therefore, these analyses allow for a comprehensive examination of an attentional preference towards social information in competition with non-social image regions with high visual saliency.

All data processing and statistical analyses were performed using the statistical programming language R (version 3.2.3; R Core Team, 2016). An a priori significance level of $\alpha = .05$ was applied and the generalized η^2 (Bakeman, 2005) is reported as effect size estimate. Huynh-Feldt's ε is reported for all repeated-measures ANOVAs containing more than one degree of freedom in the numerator to account for potential violations of the sphericity assumption. Significant effects in the ANOVA were followed by post-hoc pairwise comparisons with Bonferroni-corrected p-values.

RESULTS

FIRST SACCADE

Reflexive attention was analyzed as the relative frequency of first saccades directed congruently towards the social half of the stimulus (Figure 2.3B). Results of the full stimulus set revealed a significant intercept ($F_{(1,32)} = 235.25, p < .001, \eta^2 = .854$) indicating enhanced orienting towards the social as compared to the non-social stimulus side. The main effect of presentation time was not statistically significant ($F_{(1,32)} = 0.41, p = .53, \eta^2 = .003$). Concerning the reduced stimulus sets, results for a main effect of presentation time also failed to reach statistical significance for all analyses across algorithms (GBVS: $F_{(1,32)} = .54, p = .47, \eta^2 = .008$; ICF: $F_{(1,32)} = .93, p = .34, \eta^2 = .013$; DeepGazeII: $F_{(1,32)} = .93, p = .93, \eta^2 = .000$). Concerning a proportion of congruent saccades above chance level, we found significant intercepts for the reduced set of the GBVS ($F_{(1,32)} = 28.61, p < .001, \eta^2 = .321$) as well as the ICF algorithm ($F_{(1,32)} = 73.79, p < .001, \eta^2 = .559$), but not for the DeepGazeII algorithm ($F_{(1,28)} = 3.29, p = .080, \eta^2 = .321$). Thus, irrespective of whether stimuli were shown only briefly or for a longer duration, social content quickly attracted attention even when its low-level visual saliency was lower than for the non-social content. When saliency was defined through pre-trained deep neural networks, however, the number of congruent saccades dropped to chance level.

FIXATION DENSITY

In order to investigate aspects of sustained visual attention, fixation densities were assessed and analyzed when stimuli were shown for 5000 ms (see Figure 2.4). Analyses revealed a main effect of ROI for the full stimulus set ($F_{(3,96)} = 165.99$, $\varepsilon = .35$, $p < .001$, $\eta^2 = .788$). The reduced sets defined by the GBVS ($F_{(3,96)} = 107.07$, $\varepsilon = .36$, $p < .001$, $\eta^2 = .702$), the ICF ($F_{(3,96)} = 136.58$, $\varepsilon = .37$, $p < .001$, $\eta^2 = .754$) and the DeepGazeII algorithm ($F_{(3,96)} = 76.70$, $\varepsilon = .36$, $p < .001$, $\eta^2 = .637$) also yielded significant main effects of ROI. Post-hoc tests indicated heads to be looked at significantly more than all other image regions (all $p < .001$). Although there was no significant difference between bodies and areas of higher saliency, both regions received significantly more attention than areas of lower saliency (both $p < .001$). These effects were comparable between the full stimulus set and the reduced ones.

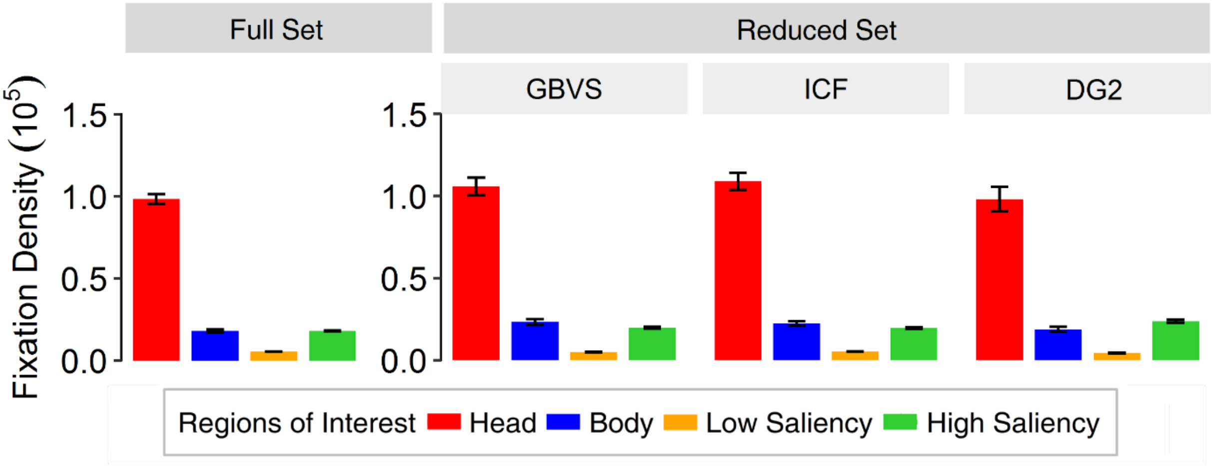

Figure 2.4. Relative area-normalized fixation density onto the different regions of interest (head, body, lower and higher saliency) during a stimulus duration of 5000 ms for the full stimulus set with ROIs defined according to the Graph-Based Visual Saliency (GBVS) algorithm, as well as the reduced sets as calculated by the three different algorithms: GBVS, Intensity Contrast Features (ICF) and DeepGazeII (DG2). Please note that ROI definition in the reduced stimulus sets was accomplished by the same algorithm that was also used for stimulus selection. Error bars represent standard errors of the mean.

FIRST FIXATIONS

To follow up on the direction of the first saccade, the first five fixations were analyzed concerning the preference to select certain ROIs. For the full stimulus set, we found a significant main effect of ROI ($F_{(3,96)} = 157.73$, $\varepsilon = .35$, $p < .001$, $\eta^2 = .665$), as well as a main effect of fixation number ($F_{(4,128)} = 30.13$, $\varepsilon = .63$, $p < .001$, $\eta^2 = .083$). An interaction effect between ROI and fixation number was also significant ($F_{(12,384)} = 30.08$, $\varepsilon = .23$, $p < .001$, $\eta^2 = .262$). As shown in Figure 2.5, fixations were primarily directed towards the head of the human being followed by areas of high saliency. This effect was most pronounced in the second fixation and declined afterwards. For the reduced stimulus sets, results for the GBVS algorithm revealed a main effect of ROI ($F_{(3,96)} = 48.60$, $\varepsilon = .34$, $p < .001$, $\eta^2 = .350$), a main effect of fixation number ($F_{(4,128)} = 9.35$, $\varepsilon = .65$, $p < .001$, $\eta^2 = .032$), as well as an interaction effect ($F_{(12,384)} = 10.89$, $\varepsilon = .24$, $p < .001$, $\eta^2 = .124$). The analysis on the reduced set of the ICF algorithm yielded similar effects, revealing a main effect of ROI ($F_{(3,96)} = 46.97$, $\varepsilon = .34$, $p < .001$, $\eta^2 = .367$) and fixation number ($F_{(4,128)} = 4.48$, $\varepsilon = .59$, $p = .010$, $\eta^2 = .015$), as well as an interaction effect of both factors ($F_{(12,384)} = 5.90$, $\varepsilon = .21$, $p = .002$, $\eta^2 = .064$), as did the DeepGazeII algorithm: main effect of ROI ($F_{(3,96)} = 31.19$, $\varepsilon = .35$, $p < .001$, $\eta^2 = .186$), fixation number ($F_{(4,128)} = 3.94$, $\varepsilon = .63$, $p = .016$, $\eta^2 = .019$), and interaction ($F_{(12,384)} = 4.66$, $\varepsilon = .22$, $p = .007$, $\eta^2 = .075$). In general, analyses on all reduced stimulus sets revealed a preference for fixating the head of the depicted human being that was largest for the second and third fixation. For the stimulus sets defined by the GBVS as well as the ICF algorithm, this preference was even present for the first fixation. Consistent with the direction of the first saccades, no such preferential orienting was observed for the first fixation in the reduced stimulus set as defined by the DeepGazeII algorithm.

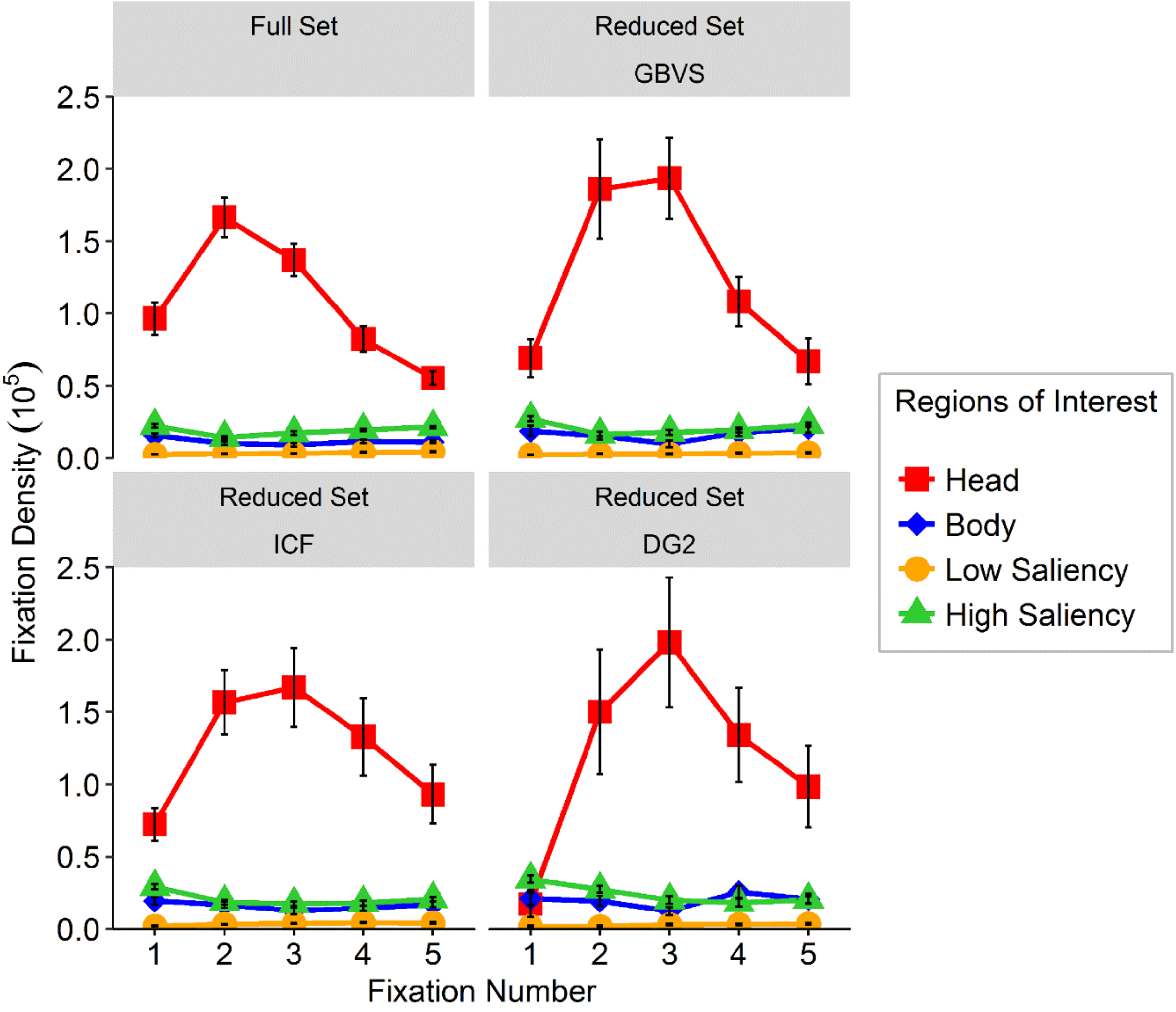

Figure 7. Relative area normed fixation densities of the first five fixations after stimulus onset on four different regions of interest (head, body, low-saliency, high-saliency) for the full stimulus set and the reduced sets defined by the three different algorithms (Graph-Based Visual Saliency, GBVS; Intensity Contrast Features, ICF, DeepGazeII, DG2) with higher mean and peak saliency on the non-social sides of the stimulus. Error bars represent standard errors of the mean.

DISCUSSION

In Experiment 2, we investigated whether viewers overtly attend to social information earlier and longer than to non-social information. By using the same stimulus set as Experiment 1, this analysis allowed us to directly compare covert and overt attentional orienting towards social information. Our data reveal two main findings: First, participants made significantly more first saccades towards the side of the image that included a social stimulus irrespective of the presentation time. This initial attentional orienting even for short viewing durations (200 ms) that do not permit for an active exploration of the stimulus indicates that social attention is reflexive (Rösler et al., 2017). Second, these saccades also predominantly landed on social features, specifically on heads of the depicted human beings during long presentations (5000 ms). Saccade target selection is therefore not only biased towards the social side in the beginning of scene presentation, but also specifically targets heads when allowed to further explore the scene. This attention bias was evident even for stimuli in which the social half of the stimulus was less salient in terms of low-level features than the non-social half (reduced stimulus sets for GBVS and ICF, respectively) and could be demonstrated even when comparing social elements (heads) against areas of high saliency. Furthermore, we took into account high-level information by using a stimulus selection based on the pre-trained deep neural network of the DeepGazeII algorithm. In this case, social information did no longer predict the direction of initial saccades. Our results thereby not only suggest that social information is overtly selected over non-social information largely independent of low-level saliency, but also that this social bias might be more adequately captured by deep learning models. This is also reflected in the initial fixations, which revealed similar patterns for algorithms based on low-level features (GBVS, ICF) with a high priority for attending heads throughout all five fixations post stimulus onset, yet different results for the DeepGazeII algorithm in which the first fixation was mostly allocated onto areas of high saliency. This again indicates that the DeepGazeII algorithm performed better in predicting potential fixation targets than the algorithms considering only low-level features, but only for very early attentional mechanisms, since subsequent fixations depict a reverted superior attention to heads, similar to those of the low-level feature algorithms. This suggests that pre-trained features may be more successful in explaining reflexive social attention, although its accountability deteriorates for later fixations. Concerning more sustained attention reflected by fixation densities over longer presentation durations, the reduced set of the DeepGazeII algorithm yielded similar results as those of the low-level saliency algorithms. A social bias persisted even when the mean and peak saliency was lower on the social side of the stimulus. Furthermore, the small sample of images in the DeepGazeII reduced set calls for further examination, as the current results do not necessarily generalize to a larger set of heterogeneous stimuli. Nevertheless, the comparison between saliency algorithms based on low-level and high-level features offers interesting insights for future approaches in modelling social attention. Another interesting aspect for future analyses is the additional consideration of local saliency distributions as well as the location of semantic information. Our method of reducing the stimulus set according to mean and peak saliency accounts for local and more distributed peaks of saliency, but could be extended by implementing a pattern-based framework and object categories to specify how various factors, such as semantic information, contribute to saliency and gaze behavior (see Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014).

Several studies have shown that low-level features are capable of predicting fixation locations well above chance level in naturalistic scenes (e.g., Foulsham & Underwood, 2008) with greatest accuracy for initial eye movements after stimulus onset (e.g., Parkhurst, Law, & Niebur, 2002; but see Tatler, Baddeley, & Gilchrist, 2005). Rösler, End and Gamer (2017) used a generalized linear mixed model to reveal that saliency was a significant predictor of gaze behavior for initial saccades in social settings, although social content surpassed saliency as a predictor in the model. On the one hand, our findings support the latter result, as we controlled for low-level saliency distributions across our stimuli and further selected those for which mean and peak saliency were lower on the social half of the stimulus than on the non-social half. On the other hand, our data further suggest the feasibility of using probabilistic models including transfer learning from other networks. The DeepGazeII algorithm, as well as its predecessor were shown to be sensitive to faces (Kümmerer, Theis, & Bethge, 2014, p. 8) and spatially-specific face and person detection (Kümmerer et al., 2016, p. 13), due to an increase in explained information through a combination of pre-training with features from a convolutional network. This seems to be the case in our data as well in accounting for reflexive (direction of the first saccade), but not for sustained social attention (prevailing high fixation densities on heads). Regarding results gained from purely low-level driven characteristics, our data using the GBVS and ICF algorithms are in line with the study of Birmingham, Bischof, & Kingstone (2009b), who showed that the low-level saliency model did no better at predicting first initial viewing behavior towards the eyes than would be expected by chance. As previously shown by End & Gamer (2017), who compared the predictability of fixation behavior through saliency for social and non-social naturalistic scenes, low-level features seem to allow for better fixation location predictions when considering scenes without any social content, but perform worse for images containing human beings. We conclude that low-level saliency does not account for the attention bias to select social features, while high-level features that might be implicitly tuned towards social information by means of pre-training perform better. Our study further suggests that social attention in naturalistic scenes is reflexive, which confirms the findings of Rösler, End, & Gamer (2017) and nicely ties in with the results of Experiment 1 demonstrating rapid covert orienting towards social information.

GENERAL DISCUSSION

This study investigated mechanisms of social attention within overt and covert viewing conditions of complex naturalistic scenes in two different paradigms. Results of both experiments revealed an attention bias towards social information. A social dot-probe paradigm addressing covert attention revealed faster and more accurate responses to probes presented on the side of the social feature (i.e., human being) in the scene, whereas a subsequent eye-tracking study tested reflexive and sustained overt attention in accordance with these probe response data. Results derived from gaze behavior in the eye-tracking experiment revealed first saccade allocation towards the side of the social feature. First saccade direction offered a direct measure for early visual attention upon the onset of a scene. Those first saccades in the short presentation condition indicate reflexive attentional mechanisms, because a stimulus duration of only 200 ms precludes a detailed visual exploration of the presented stimulus (see Rösler et al., 2017). Thus, initial attention was preferentially directed towards social features covertly as well as overtly. In the covert dot-probe paradigm, responses for the detection task were faster than those for the more cognitively demanding discrimination task. For longer presentation durations in which participants could freely view the scene for 5000 ms, fixations were also preferentially directed to social features, especially heads. Importantly, results of both experiments remained consistent when the data set was reduced to stimuli with higher mean and peak saliency on the non-social side with regard to low-level features. High-level features on the other hand, accounted well for the saliency of social features with respect to early attention processes, but less so for sustained social attention. Interestingly, the congruency effect tested within the dot-probe paradigm was higher for the reduced sets as compared to the full set for both tasks, while first saccades exhibited lower congruency effects for the reduced sets than for the full set. The fixation densities for the head region of the displayed human being were slightly higher in the reduced sets of GBVS and ICF than for the full set. However, reaction time results and eye movements may not be directly comparable as eye movements represent a direct measure of visual processing (Kirchner & Thorpe, 2006; Parkhurst & Niebur, 2003), while response times require additional response initiation (Thorpe, Fize, & Marlot, 1996). Nevertheless, there is a possibility that covert attention is more prone to interference by social attention, at least for early attention processes, as fixation densities over longer viewing durations seem to depict a higher social bias for the reduced stimulus sets of the low-level feature algorithms compared to the full set. Despite this interesting dissociation, the combination of our social dot-probe with eye movement measures (as also suggested by Bantin, Stevens, Gerlach, & Hermann, 2016, for example) and coherent results of both experiments reinforce one another, and allow us to draw conclusions about both overt and covert attention processes. Our findings are also in line with previous studies showing an overt attention bias towards heads of human beings (e.g., Birmingham et al., 2008a, 2009a; End & Gamer, 2017; Fletcher-Watson et al., 2008), but also underline findings of experiments on covert attention towards faces (Bindemann et al., 2007). Earlier studies addressing saccade programming provide further evidence for a tight temporal and spatial coupling between the preparation of saccades and prioritized visual processing, and speak against the ability to attend one location while simultaneously preparing a saccade to another (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995). According to these accounts, attention selection precedes saccade initiation and determines the endpoint of a saccade (see Kowler et al., 1995), independent of whether they were directed by peripheral (Schneider & Deubel, 1995) or central cues (Deubel & Schneider, 1996). Huestegge & Koch (2010) later suggested that attentional disengagement always precedes oculomotor disengagement, a process that applies to both covert and overt orienting, thus further supporting our results. Collectively, these results suggest reflexive covert as well as overt orienting towards social features in naturalistic scenes, even when visually salient non-social information is present.

A question that is frequently raised in the context of attention towards other human beings is that social attention may be a form of top-down process engaged by ascribed meaning. Social information may be selected because it is most relevant to the viewer and may be considered a task goal in monitoring a given situation. However, considering the low interference of top-down instructions with social attention (Birmingham et al., 2008a), even when this social bias is detrimental to solving the task (Flechsenhar & Gamer, 2017) suggests that social stimuli may be effective on their own, independent of bottom-up or top-down mechanisms (see also Kuhn et al., 2009). Alternatively, Mack, Pappas, Silverman, & Gay (2002) conclude that attention is only captured after meaning has been assigned. This sequential processing would impose a very high perceptual load, which is refuted by our finding that social features are selected reflexively within 200 ms in the absence of explicit top-down influences. Even though others found covert and overt orienting to be independent, for reflexive (Hunt & Kingstone, 2003) as well as voluntary shifts of attention (Amelia R. Hunt & Kingstone, 2003), our results speak in favor of a linked relationship, at least in a social context. Although we did not look into neural mechanisms of overt and covert attention, the current findings might nicely correspond with the premotor theory of attention, that is, covert shifts of attention are seen as unexecuted overt shifts that rely on the same mechanisms (De Haan et al., 2008; Rizzolatti et al., 1987).

With respect to reflexive versus sustained attention, our findings are coherent with the results of Cooper & Langton (2006) depicting an attention bias towards most relevant social features for short presentation durations (100 ms). However, they differ for longer presentation times: while our eye-tracking study showed that attention towards social stimuli was dominant for both short (200 ms) and long (5000 ms) presentation durations, the results of Cooper & Langton demonstrated a reversed effect for long durations (500 ms) explained by inhibition of return (Posner et al., 1982; Theeuwes & Van Der Stigchel, 2006). Other dot-probe paradigms using facial expressions as attention capture have shown similar inconsistent results and earned criticism (see Schmukle, 2005). Deviations of our results compared to those of Bindemann et al. (2007) and Cooper and Langton (2006) may originate from using complex scenes instead of isolated features, and therefore complicate a direct comparison. Investigating naturally occurring phenomena before moving into laboratory settings, as stated by the cognitive ethology approach, may allow for more concurrent findings, especially in social contexts (e.g., Kingstone, 2009). For instance, the Posner paradigm had important implications for research of spatial attention by differentiating between exogenous and endogenous cues for allocation of attention (Posner, 1980). By modifying this paradigm, Friesen & Kingstone (1998) demonstrated a discrepancy by revealing that, contrary to results using laboratory-generated cues, naturalistic directional cues with social value such as eye gaze, which endogenously cued a location, triggered reflexive shifts of attention. This would indicate that social attention also relies on a different neuronal network than the ones suggested for bottom-up (ventral network) and top-down (dorsal network) mechanisms (Corbetta & Shulman, 2002). Objects are said to attract attention more efficiently when they are also relevant, which reinforces this assumption (Corbetta & Shulman, 2002). This demonstrates that an important challenge for future research is to reveal the neural underpinnings of social attention (cf. Adolphs & Spezio, 2006; Birmingham, Cerf, & Adolphs, 2011; Gamer & Büchel, 2009) and to define their organization with respect to other attention networks.

A number of mental disorders are particularly characterized by deficits in social functioning (e.g., social phobia or autism spectrum disorders). The dot-probe paradigm was initially conducted to examine vigilance and avoidance behavior in such clinical populations (e.g., MacLeod, Mathews, & Tata, 1986), but eye-tracking studies have also added to the existing literature. For example, patients with social phobia show an initial hypervigilance towards social features (Boll, Bartholomaeus, Peter, Lupke, & Gamer, 2016), sometimes followed by avoidance behavior (Horley, Williams, Gonsalvez, & Gordon, 2003), whereas patients with autism spectrum disorder show general avoidance behavior to social stimuli, especially to the eye regions of faces (Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009; Klin, Jones, Schultz, Volkmar, & Cohen, 2002a). When given the choice between a face and an object, patients with social phobia were found to prioritize the object and thereby show the opposite effect of healthy cohorts who select social information eagerly (e.g., Chen et al., 2002). Similar patterns have been found concerning gaze-following in patients with autism (e.g., Baron-Cohen, Campbell, Karmiloff-Smith, Grant, & Walker, 1995; Baron-Cohen, Wheelwright, & Jolliffe, 1997). Although our study demonstrates that covert and overt shifts of attention towards social information are highly connected in healthy individuals, it would be very interesting to examine their relationship and possible dissociation in patient groups.

CONCLUSIONS

The present study showed an attention bias for social features in complex naturalistic scenes that was evident for covert as well as overt attention. Many dot-probe paradigms have been previously utilized, but we are not aware of an existing *social dot-probe* paradigm like the one designed for this study. Even though we did not use cues, the social feature functioned as a cue by accelerating reaction times for probes presented on the social half of the stimulus. Additionally, we compared this dot-probe task to eye-tracking results in order to investigate covert and overt mechanisms of attention to the same set of stimuli, and revealed a robust effect of preferential attention towards social features, even in competition with highly salient non-social features. This social bias was evident for reflexive as well as sustained attention and highlight the importance of social features that are attended even in complex real-world scenes consisting of an abundance of competing, potentially more salient objects. These findings offer further support for the special role of social attention and set a strong example of social dominance.

ACKNOWLEDGEMENTS

The authors particularly thank Matthias Kümmerer for the provision and kind, helpful assistance regarding the implementation of his ICF and DeepGazeII algorithms. We further thank Eva Baur for her help in stimulus preparation and data collection. This work was supported by the European Research Council (ERC-2013-StG #336305).

AUTHOR CONTRIBUTIONS

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

IMPLICATIONS STUDY 3

After revealing that social features were also selected reflexively, we wanted to unearth if this selection was in fact also voluntary. To accomplish this, we sidelined bottom-up influences through a gaze-contingent display to induce top-down driven gaze orientation in order to explore the stimuli. One could expect an equal distribution of eye movements across such displays with no preference for certain locations given that the contextual information is missing. However, considering the robust social bias found in our previous experiments, participants may deliberately seek out social information. If the latter assumption was validated and the results of the two previously discussed studies (section 2.1 and 2.2) considered, a construct of social processing may be unearthed that deviates from that of the traditional attention dichotomy in support of the notion that social stimuli engage special perceptual processing, potentially through a partial override of these mechanisms in attentional orienting.

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

ALEYA F. FLECHSENHAR, LARA RÖSLER & MATTHIAS GAMER

JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG

Corresponding author: Lara Rösler, Department of Psychology, Julius Maximilian University Wuerzburg, Marcusstrasse 9-11, 97070 Würzburg, Germany.

E-mail address: lara.roesler@uni-wuerzburg.de

Aleya Flechsenhar and Lara Rösler contributed equally to this work.

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

ABSTRACT

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

INTRODUCTION

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

While such saliency approaches particularly emphasize stimulus-driven, bottom-up attentional control, free-viewing entails the engagement of both bottom-up, as well as goal-directed (top-down) attentional processes (Corbetta & Shulman, 2002; Theeuwes, 2010). In our daily lives, however, we do not only freely perceive our surroundings, but often have a certain question in mind – these task-related requirements are known to engage mainly top-down control (Torralba et al., 2006). Within top-down driven models, a different approach by Najemnik & Geisler (2005) has taken locations of maximum information gain into consideration, characterizing the ideal observer model and emphasizing the role of the resolution of the visual system, which is maximal at the point of the fovea and limited in the periphery. Similarly, Foulsham & Underwood (2008) and Tatler & Vincent (2009) have emphasized the importance of systematic tendencies of eye movements in scenes that may predict gaze behavior as well as saliency models, ensuing that eye movements and attention are associated, as they are driven by the same internal mechanism (see "pre-motor theory of attention" of Rizzolatti et al., 1987). Importantly, while stimulus-driven and goal-driven attention are closely intertwined in free-viewing conditions, gaze-contingent viewing offers the possibility to effectively restrict pre-attentively available feature information (Loschky & McConkie, 2002). When only the currently fixated location is revealed to observers, low-level features of the image periphery cannot attract the observers' eyes in a bottom-up fashion as proposed by saliency models of attention. Indeed, search time, saccade length and fixation durations were found to be affected during gaze-contingent viewing, indicating that differential attentional mechanisms are employed during image exploration (Loschky & McConkie, 2002). Previous studies have used gaze-contingent viewing windows to investigate how information is acquired during reading (Rayner, 1998) and which field of view optimizes picture memorization (Shioiri & Ikeda, 1989). Despite different tasks at hand, both studies rested on the assumption that gaze-contingent windows are moved in such a manner that task execution is optimized. Saccades, however, also tend to process information within the current viewing window as vertical windows trigger a higher number of vertical saccades while horizontal shapes yield more horizontal saccades (Foulsham, Teszka, et al., 2011). It could therefore be argued that gaze-contingent viewing reduces bottom-up processing of peripheral information but cannot entirely eliminate bottom-up processing of stimuli presented within the viewing window.

Nonetheless, Kennedy and Adolphs (2011) demonstrated that gaze-contingent viewing can be used to effectively alter the balance between bottom-up processing and top-down control in order to reveal mechanisms of social perception. They first showed that patient S.M., who suffers from a bilateral amygdala lesion, failed to fixate the eyes of faces when allowed to freely explore the stimuli. However, when viewing the same stimuli through a gaze-contingent window, she exhibited regular eye fixations. This result suggests that gaze-contingent viewing meaningfully eliminates competing bottom-up features of social information which drive gaze behavior. To what extent does gaze-contingent viewing alter gaze patterns when viewing complex naturalistic social scenes? Typically, social features are prioritized over competing physically salient objects when viewing complex naturalistic scenes (Birmingham et al., 2008a; End & Gamer, 2017; Flechsenhar & Gamer, 2017; Rösler et al., 2017). Specifically, Rösler, End & Gamer (2017) have shown that attention to social features takes place reflexively as revealed by the direction of first saccades after a very brief stimulus presentation time of only 200 ms. While bottom-up processes thus seem to drive social attention, top-down processes, e.g. attempting to spot a friend in a crowded bar, are likely to additionally impact gaze behavior. Flechsenhar & Gamer (2017) showed that the implementation of tasks that specifically intended to drive attention away from social aspects of the scene still resulted in preferential allocation of attention onto depicted human beings. Collectively, these studies suggest that bottom-up mechanisms are essential in driving social attention. However, the precise role of top-down attentional control is less clear since the vast majority of studies in this domain used free-viewing conditions that do not permit a dissociation between bottom-up and top-down processes.

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

METHODS

PARTICIPANTS

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

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

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

APPARATUS

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

STIMULI

The stimuli used consisted of 160 naturalistic images. Half of these images displayed scenes containing one or more human beings displayed anywhere within the image, which will be referred to as social images in the following. The other 80 images showed scenes containing non-social features, predominantly complex landscapes, including objects and on rare occasions animals. The stimulus set was taken from End and Gamer (2017) and created from various image databases including the Nencki Affective Picture System (Marchewka et al., 2014), EmoPics (Wessa et al., 2010), the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), McGill Calibrated Colour Image Database (Olmos & Kingdom, 2004), Object and Semantic Images and Eye tracking dataset (OISE; Juan Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014) and websites such as Flickr and Google (selected images from databases are specified in section S1 in the Supplement). Contrast and luminance were adjusted manually by visual judgement. Stimuli were presented in a resolution of 1200 x 900 pixels resulting in a visual angle of 35.81° x 27.24° within the current setup. The currently used social and non-social images were already employed in a previous study and were shown to be comparable regarding basic visual properties such as image complexity or clutter as well as affective quality and personal relevance (End & Gamer, 2017).

DESIGN

The experiment consisted of two different types of viewing modalities for the stimuli: (1) free-viewing and (2) gaze-contingent viewing. For each participant, images were randomly associated to these viewing conditions while ensuring for an equal number of social and non-social images in each condition. In the free-viewing condition, the whole image was visible at a time and could be explored freely. The gaze-contingent display enabled the participant to only see the part of the stimulus that was centered at the current fixation location. The online tracking enabled real-time contingency on the display with the movement of the participant's eye. The visible area was defined by a Gaussian transparency mask with full-width half-maximum of 3° of visual angle around the center of the current fixation location (adapted from Kennedy & Adolphs (2011). The stimuli in the gaze-contingent condition were masked with a fixed grid of small dots located 2.2° from one another with a 3-pixel diameter to offer a sense of coordination during stimulus exploration (Figure

3.1). There was no postulated task for either the free-viewing nor the gaze-contingent condition, but participants were instructed that they could explore the stimuli freely if desired. Further, they were informed that the image would be masked in the gaze-contingent condition and that they would be able to uncover image areas by moving their eyes.

PROCEDURE

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

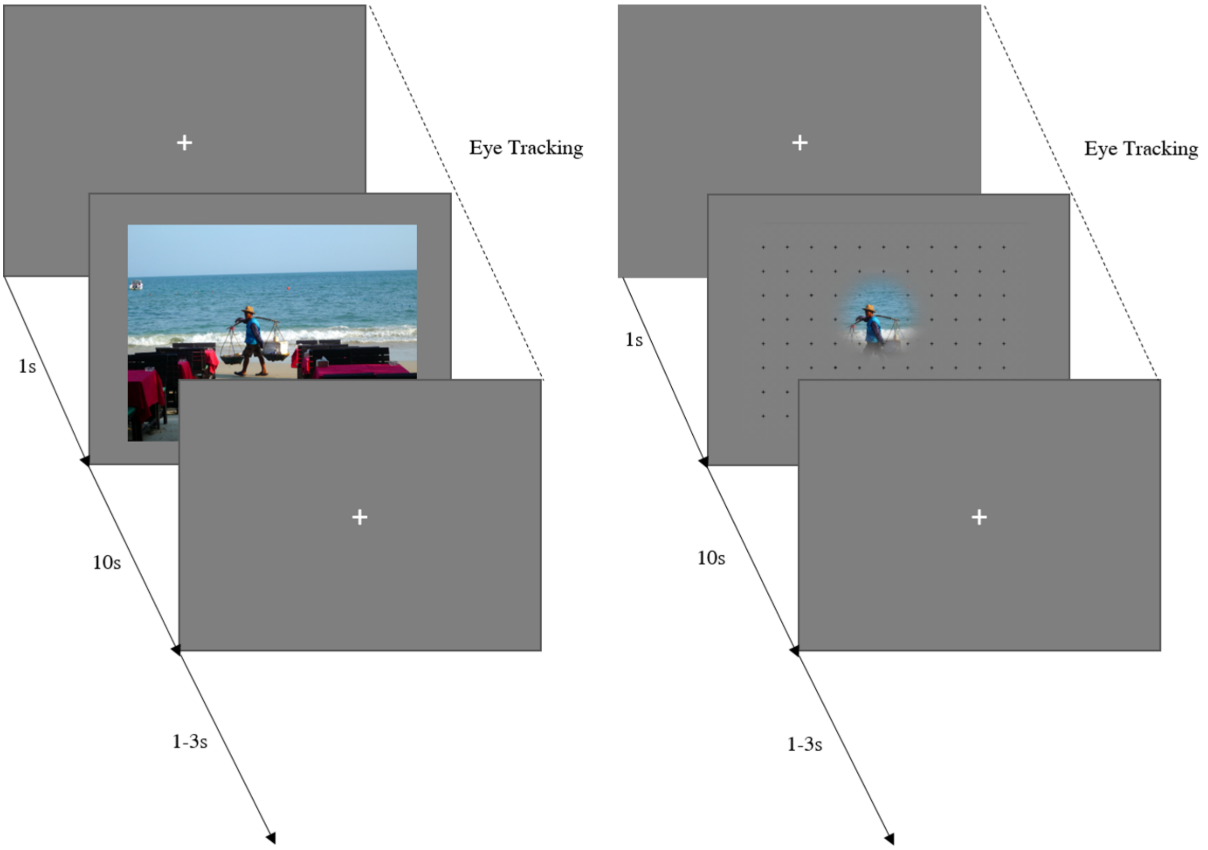

Figure 3.8. Example of an experimental trial for a free-viewing condition (left) and a gaze-contingent condition (right). The presentation time for both conditions was set to 10 s. Image taken with permission from the Nencki Affective Picture System (Marchewka et al., 2014).

DATA ANALYSIS

Data were analyzed using the open-source statistical programming language R (www.r-project.org, version 3.3.3) and MATLAB® R2011b. The R-package *ez* (version 4.3; Lawrence, 2016) was used for all repeated-measures analyses of variance (ANOVAs). An a priori significance level of $\alpha = .05$ was specified for all statistical tests. Generalized η^2 (Bakeman, 2005) and Cohen's *d* are reported as estimates of the effect size for ANOVAs and *t*-tests, respectively. The Huynd-Feldt procedure was used for all repeated-measures ANOVAs containing more than one degree of freedom in the enumerator to account for potential violations of the sphericity assumption.

EYE TRACKING PREPROCESSING

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

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

GENERAL INFLUENCE OF SALIENCY

In order to determine to what degree low-level visual saliency predicts fixations for social and non-social scenes, we compared similarities between fixation density and saliency maps. The latter were calculated for each image using the Graph-Based Visual Saliency (GBVS) algorithm (Harel et al., 2007) that was shown to be capable of predicting visual exploration with considerable accuracy (Borji & Itti, 2013; Judd et al., 2009). Similar to fixation densities, saliency maps were normalized to range from 0 to 1. Both maps were compared using standard metrics (Wilming, Betz, Kietzmann & König, 2011). These comprised the divergence of the distributions of physical saliency and fixation density (Kullback-Leibler divergence, D_{KL} see Itti & Baldi, 2005; Kullback, 1959), the classification of saliency at fixated and non-fixated image locations (area under the receiver-operating characteristic curve; *AUC* see Fawcett, 2006; Tatler, Baddeley, & Gilchrist, 2005a) and the linear dependence between the two variables (Pearson product-moment correlation coefficient *r* see Hwang, Higgins, & Pomplun, 2009; Kootstra, De Boer, & Schomaker, 2011). For AUC, fixation density maps were binarized using the mean fixation density as threshold. All metrics were calculated separately for social and non-social scenes and the two viewing conditions and compared using 2 x 2 repeated-measures ANOVAs with factors viewing condition (free-viewing, gaze-contingent) and stimulus category (social, non-social) on each measure.

REGIONS OF INTEREST

To quantify the fixation density onto physically salient aspects and social features, we introduced regions of interest (ROIs). Similar to our previous studies (End & Gamer, 2017; Flechsenhar & Gamer, 2017), we differentiated between regions of high saliency, low saliency, head and body.

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

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

RECURRENCE QUANTIFICATION ANALYSIS

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

RESULTS

SALIENCY-BASED PREDICTION OF FIXATIONS

As a difference measure between two probability distributions, we analyzed the Kullback-Leibler Divergence (D_{KL}) to examine how well physical saliency predicted the observed eye movements during free and gaze-contingent viewing. Herein, the distributions of saliency and fixations diverged significantly more for social stimuli as compared to non-social ones, as described by a significant main effect of stimulus content ($F_{(1,74)} = 180.05$, $p < .001$, $\eta^2 = .031$). Further, a significant main effect of viewing condition ($F_{(1,74)} = 58.29$, $p < .001$, $\eta^2 = .205$) generally describes lower predictability of fixations by saliency in free-viewing than in gaze-contingency. A significant interaction effect of both factors ($F_{(1,74)} = 30.49$, $p < .001$, $\eta^2 = .005$) refers to smaller differences between stimulus categories within the gaze-contingent condition compared to free-viewing. Coherently, when regarding results for the area under the receiver-operating curve (*AUC*), we found an inverse relationship, namely a significant main effect of viewing condition ($F_{(1,74)} = 10.05$, $p = .002$, $\eta^2 = .041$) with worse saliency-based prediction of fixations for gaze-contingent displays than for free-viewing. A significant interaction between viewing condition and stimulus category ($F_{(1,74)} = 41.82$, $p < .001$, $\eta^2 = .017$) describes the observation that fixation predictions were worse for social stimuli in the free-viewing condition, yet better in the gaze-contingent condition. The main effect of stimulus category, however, was not statistically significant ($F_{(1,74)} = 1.71$, $p = .19$, $\eta^2 < .001$). Results of the Pearson product-moment correlation coefficient (*r*) showed worse saliency-based prediction of fixations for social as compared to non-social stimuli (main effect of stimulus category: $F_{(1,74)} = 38.52, p < .001, \eta^2 = .015$), and a significant difference between viewing conditions (main effect of viewing condition: $F_{(1, 74)} = 45.86, p < .001, \eta^2 = .183$). Similar to the analysis of D_{KL}, the difference in predictability between stimulus categories was higher in the free-viewing than in the gaze-contingent presentation (interaction effect: $F_{(1,74)} = 21.81, p < .001, \eta^2 = .013$) (Figure 3.2).

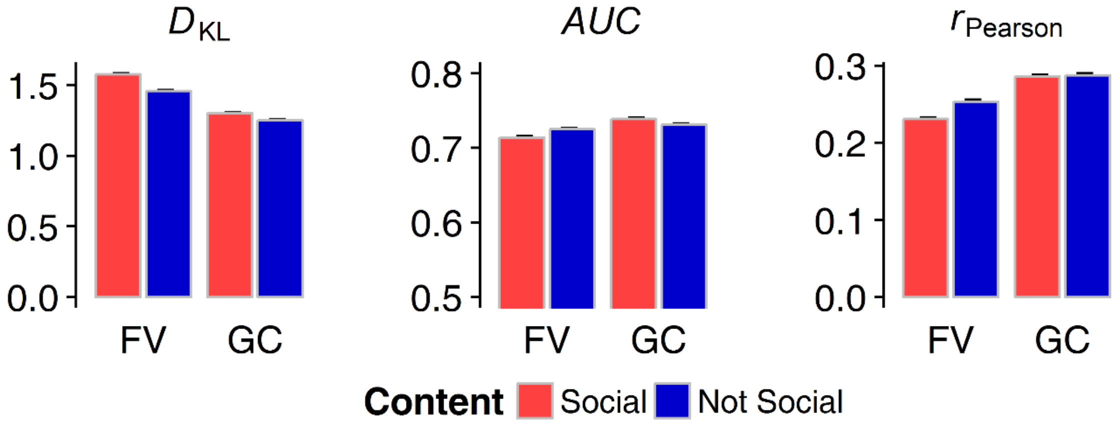

Figure 9.2. Divergence (Kullback-Leibler divergence, D_{KL}) and correspondence (area under the receiver-operating curve, AUC; Pearson product-moment correlation coefficient, r) between saliency and fixation density maps for social and non-social scenes in free-viewing (FV) and gaze-contingent (GC) conditions. Error bars represent standard errors of the mean.

ROI-ANALYSIS

Considering fixation density on pre-defined ROIs, we found a significant main effect of ROI ($F_{(3,183)} = 1291.00, \varepsilon = .44, p < .001, \eta^2 = .899$) with social ROIs gaining most attention, especially heads, compared to all other regions in both conditions. A significant main effect of viewing condition ($F_{(1,61)} = 153.29, p < .001, \eta^2 = .239$) describes higher fixation densities in general for free-viewing. We also observed a significant interaction of viewing condition and ROI ($F_{(3,183)} = 121.52, \varepsilon = .46, p < .001, \eta^2 = .342$) depicting overall lower fixation densities for gaze-contingent displays than for free-viewing, which is especially the case for head and body ROIs. The interaction effect may therefore be driven mainly by the fact that exploration of social ROIs is reduced in gaze-contingent displays compared to the free-viewing condition (Figure 3.3, left panel). To test whether this may arise from the fact that social stimuli could not be immediately attended due to the masking, we reanalyzed the data starting from the time point at which the social aspect was first fixated.

Indeed, when comparing the time points of first fixations on social ROIs for both viewing conditions, we found that participants needed significantly less time until encountering a social ROI in free-viewing (*M* = 894.88 ms, *SD* = 268.55 ms) than in gaze-contingency (*M* = 2153.76 ms, *SD* = 461.66 ms; $t_{(74)}$ = 22.27, *p* < .001, *d* = 3.45). When further analyzing fixation data from the time point of this first social detection until the end of the presentation time, we again obtained significant main effects of ROI ($F_{(3,183)}$ = 1227.15, ε = .42, *p* < .001, η^2 = .874) depicting a fixation bias for social ROIs, and a main effect of viewing condition ($F_{(1,61)}$ = 5.61, *p* = .021, η^2 = .016) implying higher fixation densities for free-viewing as opposed to gaze-contingent viewing. A significant interaction of ROI by viewing condition ($F_{(3,183)}$ = 8.59, ε = .46, *p* < .001, η^2 = .042) emphasizes that fixation densities were different across ROIs and viewing conditions, showing slightly reduced viewing behavior for social ROIs in the gaze-contingent displays. However, compared to the pattern found previously, the difference between free-viewing and gaze-contingency regarding the fixation of social ROIs seems to be slightly smaller (Figure 3.3).

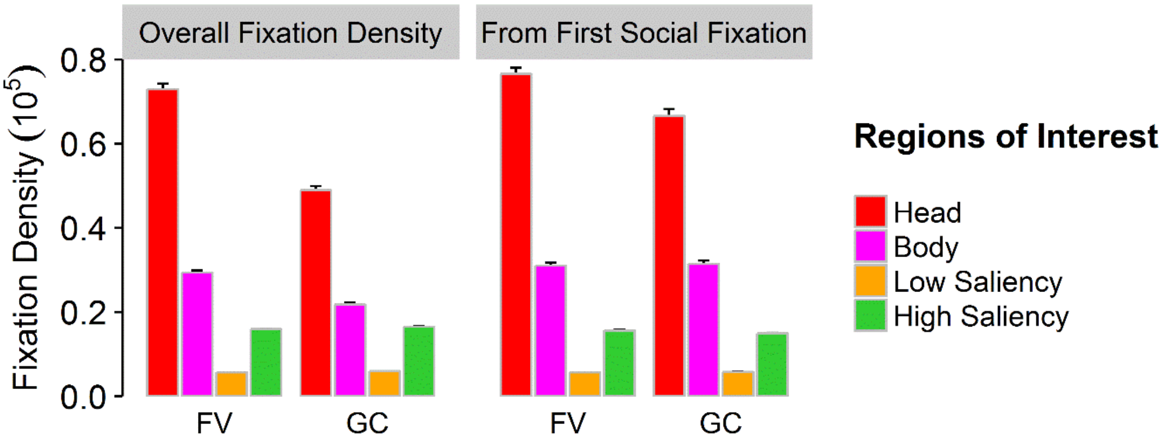

Figure 3.3. Relative area-normed fixation density on regions of interest (ROIs) for free-viewing (FV) and gaze-contingent (GC) viewing. The left panel depicts the overall fixation densities for the presentation duration of 10 s. The right panel shows fixation densities measured from the time point in which the participants first fixated a social feature until the end of the presentation time of 10 s. Error bars represent standard errors of the mean.

RECURRENT QUANTIFICATION ANALYSES

Recurrent quantification analyses (RQA) were suggested to complement analyses of fixation density since they provide additional information on the temporal dynamics of fixations. Figure 3.4 demonstrates that recurrent and deterministic fixations reveal discrepancies to fixation densities that might systematically differ between social and non-social stimulus content.

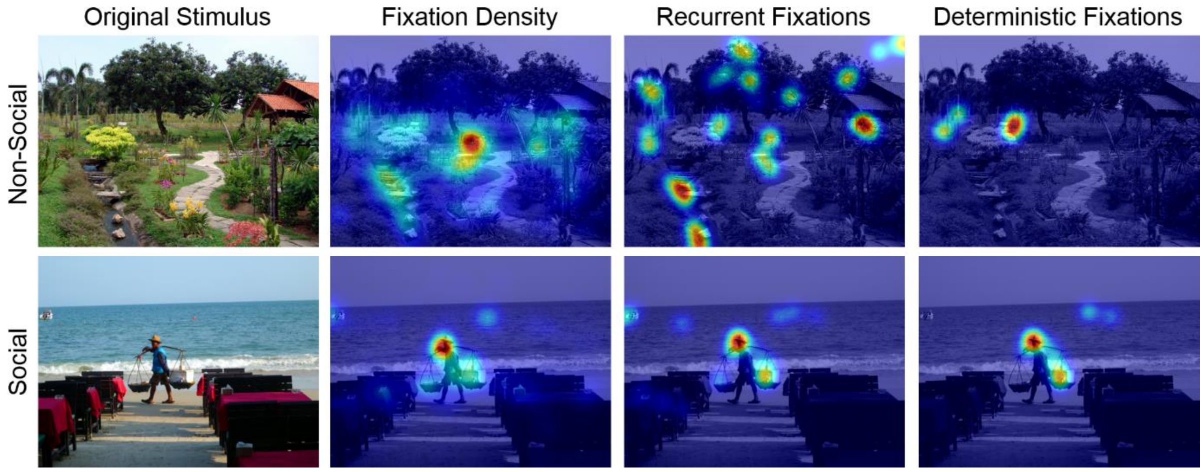

Figure 3.4. Example of a non-social (top) and a social (bottom) stimulus with respective heat maps for fixation densities, recurrent fixations and deterministic fixations of all participants across both viewing conditions. Warm colors represent areas with higher values of the respective measure, whereas cool colors indicate low values. Image taken with permission from the Nencki Affective Picture System (Marchewka et al., 2014).

In order to systematically quantify the influence of viewing conditions and stimulus category, we analyzed four different RQA measures (see Figure 3.5): (1) for the sum of recurrent fixations, we obtained a significant main effect of condition ($F_{(1,74)}$ =31.96, $p < .001$, $\eta^2 = .124$), indicating higher mean recurrence for the free-viewing condition than for the gaze-contingent display. A significant main effect of content ($F_{(1,74)} = 104.40$, $p < .001$, $\eta^2 = .013$) describes higher mean recurrence for social than for non-social stimuli. However, we did not find a statistically significant interaction of condition and content ($F_{(1,74)} = 2.88$, $p = .09$, $\eta^2 < .001$), which signifies that there was no significant difference in the sum of recurring fixations between social and non-social stimuli across viewing conditions. (2) Deterministic fixations displayed a reversed pattern with higher means for gaze-contingent than for free-viewing ($F_{(1,74)} = 97.43$, $p < .001$, $\eta^2 = .275$). Repeated subsequent fixations were also more frequent for social than for non-social stimuli ($F_{(1,74)} = 46.95$, $p < .001$, $\eta^2 = .025$) but this difference between stimulus content was more pronounced for free-viewing as compared to gaze-contingent viewing as indicated by a significant interaction effect ($F_{(1,74)} = 10.47$, $p = .002$, $\eta^2 = .005$). (3) Laminarity is another fixation repetition measure describing the tendency to attend certain locations multiple times (here more than twice). Our results showed a significant main effect of condition ($F_{(1,74)} = 121.25$, $p < .001$, $\eta^2 = .307$) with higher laminarity for free-viewing than gaze-contingent viewing and a significant main effect of content ($F_{(1,74)} = 304.56$, $p < .001$, $\eta^2 = .111$) depicting higher mean values for social stimuli. A significant interaction between condition and content ($F_{(1,74)} = 32.79$, $p < .001$, $\eta^2 = .014$) suggests that in images with social content locations were revisited more often than in images with non-social content in free-viewing, but less so in gaze-contingent viewing. (4) The measure for center of recurrent mass (CORM) enabled us to examine the temporal distribution of recurrent fixations. A significant main effect of condition ($F_{(1,74)} = 270.02$, $p < .001$, $\eta^2 = .474$) describes that recurrent fixations were closer in time for gaze-contingent displays than for free-viewing. A significant main effect of stimulus content further shows that recurrent fixations were closer in time for social than for non-social stimuli ($F_{(1,74)} = 4.16$, $p = .04$, $\eta^2 = .003$). A significant interaction ($F_{(1,74)} = 14.00$, $p < .001$, $\eta^2 = .011$) between viewing condition and content suggests that recurrent fixations occurred closer in time for social stimuli than for non-social ones in free-viewing, but farther in time for the gaze-contingent condition (Figure 3.5).

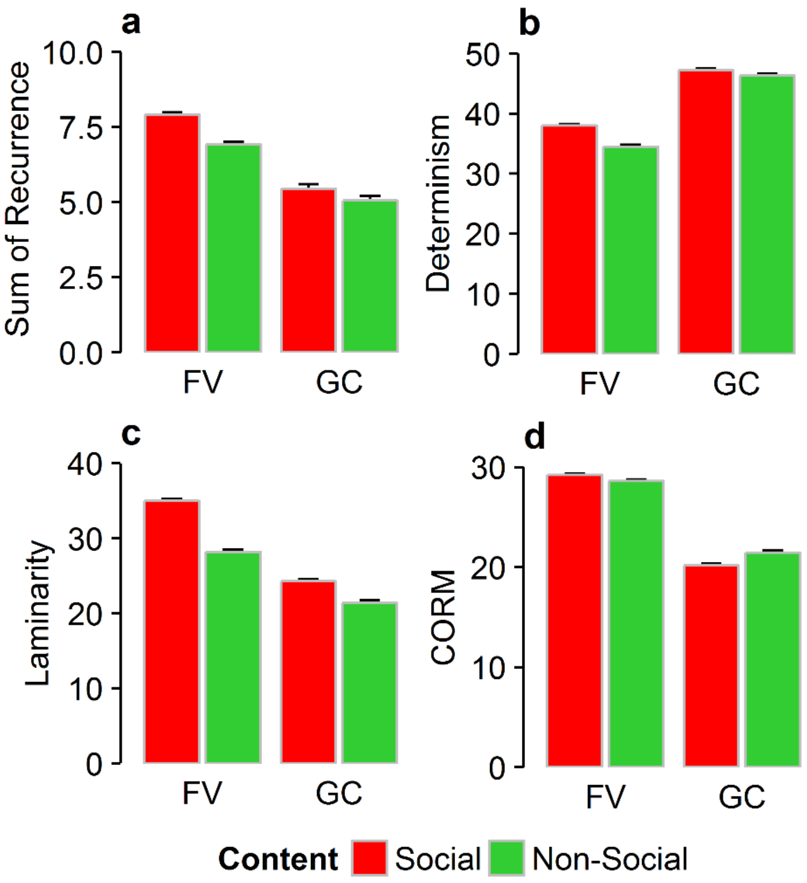

Figure 3.5. Averages of four recurrence quantification analysis measures: a) Recurrence, b) Determinism, c) Laminarity and d) Center of Recurrent Mass across free-viewing (FV) and gaze-contingent (GC) conditions for social and non-social stimulus content. Error bars represent standard errors of the mean.

DISCUSSION

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

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

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

The combination of viewing modalities allowed an additional examination of predominant top-down control (gaze-contingent viewing) and both, bottom-up and top-down influences (free-viewing) on social attention. While bottom-up processing is not completely eliminated in the gaze-contingent viewing condition, fewer low-level salient information is available near the current fixation and no such details are visible in the periphery. Thus, most executed saccades will draw on top-down processes for the determination of saccade endpoints. Our current results therefore suggest that social attention is not merely reflexive but also relies on top-down attentional processes. So how does social attention then fit into the traditional dichotomy of bottom-up and top-down mechanisms? The recurrence quantification analysis used here further implicates that viewing behavior towards social stimuli is different than for non-social stimuli with regard to fixation sequence as well as temporal structure. Foulsham & Kingstone (2010) already showed that gaze patterns can change with image content in a scene, but our data presents explicit differences between social and non-social content, suggesting that social attention is inherently different from general attention mechanisms. This, in turn, raises the question whether a special neuro-cognitive system, distinct from the ventral or dorsal network suggested for bottom-up and top-down attention, mediates social attention and its rapid allocation. The study of Kennedy and Adolphs (2011), who showed that irregular bottom-up processing caused by amygdala lesions can be overcome by using a gaze-contingent paradigm, indicates how important the disentanglement of these processes are. Furthermore, such patient studies can offer insight to underlying mechanisms and further our understanding of brain areas involved in social processing. Future neuroimaging studies investigating potential candidates for a social attention network are necessary to further elucidate this assumption.

Even though our findings depict robust and successfully replicated results, our study has a few limitations. First, we cannot control for certain influences arising from the use of naturalistic stimuli. For instance, although the distribution of social features within the images was considered, such that they were not always presented centrally, in the foreground or depicted only single individuals, the currently used stimulus set has some variability in the specific scene composition which might reduce the internal validity of the current setup. Furthermore, even though we carefully controlled physical image properties such as feature congestion, subband entropy, edge density, and overall saliency and ensured that these measures did not differ between social and non-social images (End & Gamer, 2017), we could not control for every aspect of scene composition and structure. For example, even though spatial frequencies have repeatedly been reported to affect attentional capture (Gomes, Soares, Silva, & Silva, 2017; Stein, Seymour, Hebart, & Sterzer, 2014), we did not control for a similar distribution of features within specific frequency bands of the current stimulus set. Nevertheless, we chose these complex scenes as they have comparatively high ecological validity and contain contextual information which plays an important role for the orientation in our environment (Torralba et al., 2006). Moreover, we deliberately wanted to defer from isolated or artificial setups, as they include viewing conditions that do not resemble important properties of the input our visual system has to deal with every day (see Kingstone et al., 2002; Smilek et al., 2006). However, it is important to note that the use of photographs of naturalistic scenes, has also been put into question (Kingstone et al., 2002), as these are not equivalent to experiencing the real world and some recent studies have indeed shown conflicting results comparing eye tracking in the laboratory with mobile eye tracking (Foulsham, Walker, & Kingstone, 2011a). Second, our study included animal pictures in the set of non-social scenes and studies have shown that eye movements may be influenced by animacy of depicted features within a complex scene (e.g., Altman, Khislavsky, Coverdale, & Gilger, 2016; New, Cosmides, & Tooby, 2007). However, this theory implies that gaze behavior for our non-social stimuli should be similarly biased (e.g., by enhancing gaze towards animals) as for social stimuli. Even so, our results still show better predictions through saliency measures for non-social stimuli and higher recurrent fixations for social as compared to non-social stimuli within our recurrence quantification analysis. Therefore, social features may still be preferred even over other animate aspects. Future studies should examine this hypothesis using a balanced set of pictures either including humans and animals in the same scene or in different sets of photos. Third, the currently used stimuli might be perceived differentially regarding emotional aspects or personal relevance and these dimensions might in turn also affect exploration patterns. While we refrained from requiring stimulus ratings in the current study due to time constraints, we collected emotional valence, arousal and personal relevance ratings for the currently used stimuli in a previous study and showed that social and non-social scenes were comparable on these dimensions (End & Gamer, 2017). Since the sample of the current study was largely similar to the sample examined before (e.g., regarding age, education and health), we did not expect to find differences in subjective ratings in the current study. However, it might be interesting for future research to directly examine the influence of affective dimensions or perceived personal relevance on viewing patterns. One recent study already demonstrated an influence of emotional valence on the visual exploration of video clips (Rubo & Gamer, 2018) and it is currently unclear whether similar effects can also be obtained for static stimuli.

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

ACKNOWLEDGEMENTS

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

AUTHOR CONTRIBUTIONS

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

3. DISCUSSION

3.1 SUMMARY

This dissertation intended to disentangle mechanisms of social attention, a topic that has been of interest for many decades, yet remains partially unresolved. The experiments conducted here aimed at successive isolation of several basic processes to gain a better understanding of a social bias observed in previous studies (for review see Birmingham & Kingstone, 2009). The empirical approach aimed at determining the reflexive, as well as voluntary nature of social processing, in overt as well as covert attention and further examined if this bias prevailed in the presence of increased top-down control. Results confirmed a general preference to attend social stimuli for free-viewing, reflecting natural gaze behavior influenced by bottom-up as well as top-down control. To examine to what extent social attention would remain superior to top-down influences, tasks were introduced that were constituted to divert attention from the depicted social features, as well as differ in complexity to judge the gradation of interference (section 2.1). Herein, specific tasks mildly influenced social attention, however without ever surpassing it. Further, results for brief presentation times suggest an automatic allocation of attention towards social aspects without top-down influencing factors (section 2.2). This was the case for overt as well as covert attention. A gaze-contingent paradigm reduced bottom-up influences by masking contextual information outside of the current focus of attention and revealed that social aspects are also selected voluntarily (section 2.3). A further analysis of data from this paradigm indicates more recurrent gaze behavior, close in time for social stimuli compared to non-social ones.

In sum these experiments imply that social attention is reflexive as well as voluntary, yet cannot be described by either one: low-level saliency cannot account for preferential gaze behavior of social features and top-down control processes are overridden in the presence of a depicted human being. The observed social bias does not seem to be a purely overt quality, but is also evident for covert attention. Hence, these results paint an interesting and holistic picture regarding the mechanisms of social attention, rendering it a specialized module in attention processing.

These results offer great insight concerning the social attention literature and inquire further integration. Implications concerning past and future research will be discussed in the following, addressing also limiting aspects as well as open questions.

3.2 THE REAL WORLD

One of the critical assumptions driving social neuroscience is that the knowledge gained about the social brain using static or dynamic stimuli will generalize to the richer scenarios associated with everyday social cognition (Risko, Laidlaw, et al., 2012). Although the utilization of naturalistic stimuli already offers a more complex and realistic setup to investigate social attention in the laboratory, it may still differ from behavior of every day social life. Foulsham, Walker and Kingstone (2011) compared gaze patterns in a true social scene in which participants were asked to buy coffee wearing a mobile eye tracking device, with a video condition in which participants viewed the recorded video of another subject's trip with mobile eye tracker. Herein, people in the scene who were far away from the observer were looked at similarly in both conditions (live and video), but when people approached the observer, participants kept their gaze on this person only in the video condition, while gaze was averted in the live condition. Laidlaw, Foulsham, Kuhn and Kingstone (2011) found similar results and Gallup, Chong and Couzin (2012) confirmed the influence of reactive eye movements and approach behavior with a real-world gaze-cueing paradigm, indicating that the bias to attend other peoples' faces or eyes may be influenced by social norms, unwanted social interaction, or unsuitable for the nature of the situation known as *civil inattention* (Zuckerman, Miserandino, & Bernieri, 1983). A recent study found the same effect with sub-clinically socially anxious participants (Rubo, Huestegge, & Gamer, under review). Therefore, a live real-world situation may alter social attention compared to static or dynamic images in which a potential consequence of interaction is absent. Johnson (2005) also suggests differential activations of the subcortical route for stimuli presented at close range, as opposed to faces viewed at a greater distance or the periphery, in which low spatial frequency information about the face changes. However, social norms and potential consequences of breaching these through staring or avoiding an unwanted social interaction does not necessarily mean that the social bias found with static or dynamic stimuli is unnatural, instead it may just be a restraint of a desideratum. A person walking around with sunglasses might depict similar gaze patterns as in a laboratory setting when individuals approach them, simply because their eyes are occluded and therefore elusive to others in the surroundings. In fact, when people know that their gaze is evident to others or being tracked, their viewing behavior changes significantly (Risko & Kingstone, 2011). Furthermore, gaze distributions may also depend on the setting of the situation. A study by Freeth, Foulsham and Kingstone (2013) investigated viewing behavior within the context of an interview and found that eye contact was less present in the video than in the live situation. This leads to the conclusion that there is a *dual function of gaze*, as stated already in 1976 by Argyle and Cook, differentiating between eyes as a channel for gathering information, as well as a signal for communicating information

(Risko, Richardson, & Kingstone, 2016). However, this would only apply to overt attention, where eye movements are evident. Covert shifts on the other hand, would allow allocation of attention while circumventing signal detection, thereby guiding appropriate social looking behavior. As such, covert attention may have evolved to monitor others, enabling them to process their actions while limiting overt looking responses (Laidlaw, Rothwell, & Kingstone, 2016), supporting our results, which show a social bias for both attention modalities in a laboratory setting.

3.3 IMPLICATIONS

Ricciardelli and colleagues (2013) differentiate between passive and active goal-directed viewing and suggest that when engaged in an active task (goal-directed saccade target paradigm), an individual will be able to suppress the bottom-up orienting with the help of top-down filtering. Study 1 (section 2.1) does not support this notion and instead highlights the sustainability of attention towards social cues in spite of increasing top-down demands. While individuals are able to resist the automaticity of gaze following, this does not seem to apply for attending a person potentially initiating a gaze cue in a complex scene. Other studies using laboratory based tasks have also depicted a partially non-volitional attention bias towards social stimuli (Devue et al., 2012; Laidlaw, Badiudeen, Zhu, & Kingstone, 2015; Laidlaw, Risko, & Kingstone, 2012). A social distractor (e.g., face) presented with a non-social saccade target changes saccade trajectories, indicating that distractor relevance can influence saccade planning beyond the task itself (Laidlaw et al., 2015). This further implies that social relevance of a face is influential in oculomotor planning and execution (Laidlaw et al., 2015). It becomes evident that although most studies have used facial stimuli as central attentional cues, it is important to investigate social cues as a distractor and associated attention modification herein (e.g., Hermens & Walker, 2010; Nummenmaa & Hietanen, 2006; but see Laidlaw et al., 2015), as we did in study 1 (section 2.1).

Although task instructions have largely revealed an alteration of gaze behavior (e.g., Yarbus, 1967), our results do not support this notion when the tasks focus on non-social elements in a social setting. Greene, Liu, & Wolfe (2012) were able to use a pattern classifier to identify the image and the observer through specific scan paths, but not to predict a viewer's task. This also reflects a contradiction to Yarbus' conclusion that an observer's task could be predicted from his scan patterns. Our study adds to this opposition, however, we did experience some alterations when comparing time windows (during and after a task), so there seems to be an influence on eye movements through task demands to a certain extent. For the classification of patterns, however, this may still be at chance-level. This aspect could be eluded in future eye tracking studies.

One could also argue that a focus on faces requires processing at a higher, cortical level, where emotional and semantic aspects are represented (Oliva, Torralba, Castelhano, & Henderson, 2003). However, study 2 (section 2.2) disagrees with this notion, as attention towards depicted human beings in the scene was reflexive. Early gaze behavior is said to be most influenced by salient objects in a scene (e.g., Anderson et al., 2015; Mackay, Cerf, & Koch, 2012), yet we controlled for saliency distributions within our analyses and still found an initial prioritization of social aspects. Face-selective magnetoencephalographic (MEG) responses in the occipitotemporal cortex have been found as early as 100 ms after stimulus onset during free-viewing of faces and control stimuli (Liu et al., 2002) and our data show that complex social stimuli may also be processed at a very early stage.

While attention and eye position can be clearly dissociated, knowledge about the precise interaction or cooperation of overt and covert attention mechanisms remains inconclusive. Gaze allocation and covert attention may be overlapping, but attention disengagement may also be a product of a saccade initiation (Rayner, 1998). The premotor theory of attention states that eye movement programming and spatial shifts of attention share a common functional module with a distinct neural basis, which has been confirmed by several studies (e.g., Beauchamp, Petit, Ellmore, Ingeholm, & Haxby, 2001; Corbetta et al., 1998; De Haan, Morgan, & Rorden, 2008; but see Ignashchenkova, Dicke, Haarmeier, & Thier, 2004 and Huestegge & Koch, 2010), especially with regard to the superior colliculus, relevant for the control of saccades (Ignashchenkova et al., 2004). Further, both types of attention deployment trigger similar event-related potentials (ERPs; Eimer, Velzen, Gherri, & Press, 2007), as well as similar perceptual gap effects, thereby indicating a functional, but also temporal coupling between the two domains (Huestegge & Koch, 2010). This speaks against the opposing *decoupling hypothesis*, as it seems impossible for human beings to direct their attention to a certain location, yet initiate saccades to a different one (Deubel & Schneider, 1996). However, the premotor theory has to be adjusted, as shifting and maintenance of covert attention can be made without intention to execute a saccade (Belopolsky & Theeuwes, 2012). Still some semantics of a scene can be accessed outside the focus of attention (Li, VanRullen, Koch, & Perona, 2002), or preattentively (Biederman, 1972; Wolfe, 1998). Although covert attention has previously merely been associated as a byproduct of oculomotor planning (e.g., Rizzolatti et al., 1987), its social function renders it behaviorally important, i.e. for monitoring others (Belopolsky & Theeuwes, 2012) to initiate appropriative overt behavior (Laidlaw et al., 2016), or hiding own intentions (Klein, Shepherd, & Platt, 2009). Nevertheless, covert attention requires a certain perceptual span for scene perception, which is suggested to be around 2.6° (Nelson & Loftus, 1980) to 4° (Henderson, Williams, Castelhano, & Falk, 2003) of visual angle away from the current fixation to still be able to recognize an object.

This is also relevant for gaze-contingent displays, which have been especially beneficial in reading research and found applicability in scene perception (Rayner, 2009). While the gist of a scene can be extracted after merely 40 ms (Castelhano & Henderson, 2008), masking reduces contextual information and thereby also influences the attention selection process. Restricted viewing induces considerable overlap of information across fixations (Saida & Ikeda, 1979), confirming the results of recurrent viewing behavior found in our third study (section 2.3). Furthermore, masking the center of a current fixation tends to increase fixation durations, while peripheral masking decreases saccade length and increases the number of fixations, indicating that the visual system chooses information that is currently available to initiate subsequent fixations instead of dwelling on the current location (van Diepen & D'Ydewalle, 2003b).

3.3.1 RETHINKING SALIENCY COMPUTATIONS

There is a lasting debate about the socio-biological value of faces, claiming that the visual system may be tuned to a set of special low-level features on the one hand, and a visual system that extracts their meaning very rapidly on the other (Devue et al., 2012). This suggests that the human visual system is not only tuned to low-level features that make up a face, but also its meaning. While faces are detected quickly in a visual search task among other distractors and prolong search times for other target objects when used as a distractor, this effect vanishes when faces are inverted even though the low-level saliency distribution remains the same (Langton et al., 2008). Our first study (section 2.1) depicts similar results when a social aspect was unrelated to the goal of the experimental task. On the contrary, the task-relevance of a face can potentiate behavior, even in the presence of other animate features (Langton et al., 2008).

While research on the impact of low-level features on visual attention have been extensive, high-level factors have been considered less. Admittedly, high-level factors are more complex and more difficult to subsume. The consideration of relevance and top-down influence when predicting gaze behavior is a feature that is usually not incorporated in standard saliency algorithms, which may explain the low predictability of such computations for social stimuli (see End & Gamer, 2017). Intermediate models, such as the one of Xu and colleagues (2014), successfully identified semantics (e.g., faces) that contribute to saliency and gaze behavior. Saliency in Context (SALICON) is an ongoing effort to understand and predict visual attention by narrowing the semantic gap (Huang, Shen, Boix, & Zhao, 2015). Pre-trained (e.g., Jiang, Huang, Duan, & Zhao, 2015) or deep neural models (e.g. VGG-19; Simonyan & Zisserman, 2014) may also be feasible alternatives for future prediction maps (see section 2.2), especially since the "learning" can be adjusted and other data samples with different domains leveraged. This highlights the complexity of human social behavior, as we need behavioral input to predict behavioral output, instead of mere stimulus feature analysis.

3.3.2 THE ASPECT OF RELEVANCE

The selection of information within the dual function of gaze depends on its relevance to the observer. Mackworth & Morandi (1967) were the first to observe the tendency that fixation density correlated with the regions rated as most informative. Furthermore, it seems that important information can be selected very early during scene viewing, as the first fixation mostly allocates onto a rated informative region rather than a rated uninformative one (Antes, 1974, but see Henderson & Hollingworth, 1999). Corbetta and Shulman (2002) state that objects attract attention more efficiently when they are also relevant. Just as emotional information is generally prioritized over neutral stimuli and receives privileged access to attention and awareness (Vuilleumier, 2005), social stimuli per se may be of such high relevance that they capture attention innately. Different studies have investigated meaning maps (e.g., Henderson & Hayes, 2017), interestingness maps (e.g., Einhäuser et al., 2008; Onat, Açik, Schumann, & König, 2014; Rensink, O'Regan, & Clark, 1997), emotional saliency (e.g., Niu, Todd, Kyan, & Anderson, 2012) or low-level saliency judgements (e.g., Borji, Sihite, & Itti, 2013). While some studies found saliency to be a good predictor of selected objects that the participant deemed interesting (Elazary & Itti, 2008; Koehler, Guo, Zhang, & Eckstein, 2014; Masciocchi, Mihalas, Parkhurst, & Niebur, 2009), others showed that saliency only indirectly affects attention: objects that were determined to be interesting to the viewer predicted fixations in images better than physical saliency and irrespective of task requirements (Einhäuser, Spain, et al., 2008). Xu, Jiang, Wang, Kankanhalli and Zhao (2014) introduced an attribute-based framework incorporating object and semantic information to investigate how these contribute to saliency. When participants are asked to select the most informative regions, eye tracking data reveals that these areas are fixated longer than other areas (Mackworth & Morandi, 1967). This is also the case for emotional content (Pilarczyk & Kuniecki, 2014) and emotional salience (Niu et al., 2012). As we deem social features to be of great relevance, interest and emotional salience, these results further support our findings of a social bias. An unpublished study of our group investigated fixation distributions according to priorities of image regions individually determined by the viewer for social and non-social stimuli. Participants could choose three different regions, which they deemed most relevant for understanding the depicted scene and subsequently, these areas were compared with the eye tracking data of different participants. Results revealed that those locations given the first priority corresponded to highly fixated regions within a 10 s free-viewing presentation. Importantly, the selected priorities were not associated with high salience, but rather with the presence of a human being. Conversely, when participants were asked to choose one of two different title choices for a given scene, they were more likely to choose the title inferring a social meaning over a non-social one (Kehrlein & Gamer, 2016). A recent study by Lo Gerfo and colleagues (2018) compared gaze direction with monetary reward in a gaze cueing task and found that the motivational valence of reward did not alter orienting of attention mediated gaze. This indicates that social cues have a greater relevance than rewards. In contrast, studies using non-social stimuli found monetary reward to be associated with gaze direction (Chelazzi, Perlato, Santandrea, & Della Libera, 2013; Pessoa & Engelmann, 2010) and reward to automatically capture attention (Sali, Anderson, & Yantis, 2014). This supports the notion that social stimuli embody a highly relevant stimulus, as the selection of informative or relevant aspects draws attention to features that are semantically important.

3.3.3 INDIVIDUAL DIFFERENCES

Personal relevance coincides with individual interests and intentions and eye movements evidently vary across participants in free-viewing conditions (e.g., Andrews & Coppola, 1999; Castelhano & Henderson, 2008). However, scan patterns are very consistent within an individual across different scenes (Underwood, Foulsham, & Humphrey, 2009), indicating stimulus-independent individual differences in eye movement behavior (Risko, Anderson, et al., 2012). Apart from cultural differences (e.g., Chua, Boland, & Nisbett, 2005; Masuda et al., 2008), personality traits can also affect scene perception, but also as a function of social development, such as social familiarity (section 3.3.4; for a review, see Hannula, 2010), or lack thereof in association with some clinical characteristics (section 3.5). Hence, individual differences of social understanding may influence how people perceive their surroundings and attend to social stimuli (Wu, Bischof, Anderson, Jakobsen, & Kingstone, 2014).

Where a person looks must be ultimately linked to cognitive, emotional, and social processing (e.g., Adolphs et al., 2005; Bush, Pantelis, Morin Duchesne, Kagemann, & Kennedy, 2015). While some aspects are generally fixated more than others across the population, differences between individuals are evident. Gaze behavior within an individual is highly consistent and systematic in terms of semantic dimensions, yet varies across observers (de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2018). Although saliency maps try to model saliency as a stimulus-given feature, these inter-individual differences speak for individual saliency biases that are a stable trait of the observer. One that is likely heritable, as investigated in a large sample of monozygotic and dizygotic twins (Kennedy et al., 2017), but also among patients (Constantino et al., 2017). Results demonstrated a genetic influence on visual exploration of scene content and spatiotemporal fixation patterns during free viewing of complex scenes. These effects were so precise that twins could be matched by their gaze patterns. Similarly, pattern classifiers including features from individual scan paths were able to identify the image and the observer at above-chance level (Greene et al., 2012). Moreover, these idiosyncratic eye movement patterns are highly stable across individual and time, as each individual employs the same strategy for visual recognition after 18 months as they do when they first view a stimulus (Mehoudar, Arizpe, Baker, & Yovel, 2014). This is the case for long (Mehoudar et al., 2014), as well as brief image presentations (Peterson & Eckstein, 2013). Therefore, two observers viewing the same image, may perceive (Henderson et al., 2003) and interpret (Bush et al., 2015) the stimulus very differently from their counterpart. Hence, there is no "typical observer", but a seemingly genuine interest in other human beings as a salient semantic category across individuals.

The aforementioned genetic disposition further eludes gene-environment interactions, as eye movements are a voluntary selection of experiences, which are influenced by genes, creating an intra-individual micro-level environment (see review of Kendler & Baker, 2007). A trait-congruency model was proposed, describing attention selection as a function of personality traits (Eizenman et al., 2003), implying a potential influence on the oculomotor system activity that is linked to personality (Canli, 2006). The study of Rauthmann, Seubert, Sachse and Furtner (2012) even concluded that the traits surveyed within the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1989) are able to predict gaze behavior. These personality questionnaires have a biological basis (see e.g., DeYoung & Gray, 2009) and assess five different character traits: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Extraversion and agreeableness relate to positive affect, linked to empathy and prosocial behavior (Graziano, Habashi, Sheese, & Tobin, 2007; Yik & Russell, 2001) and have been related positively to the amount of attention committed to the eyes of others (Wu et al., 2014). While openness to experience also relates to positive characteristics, it is also linked to imagination and the tendency to give complex narratives (McAdams et al., 2004; but see Wu et al., 2014). Neuroticism and conscientiousness are associated with internal aspects of personality and effortful control (Jensen-Campbell et al., 2002), and have been positively correlated with the amount of time spent looking at the eyes of fearful faces (Perlman et al., 2009), but showed no relation to increased fixations towards the eye region in complex scenes containing human beings (Wu et al., 2014). Other traits, such as curiosity, were also found to correlate with the amount of exploration of a scene (Risko, Anderson, et al., 2012).

To follow-up on these findings, we conducted post hoc analyses on the data of all three studies presented within this dissertation. In each study, we included several questionnaires that were completed at the end of each experiment. Common questionnaires of all studies included the Autism Spectrum Quotient (AQ-k; Freitag et al., 2007) and measures of anxiety, namely the State-Trait-Anxiety Inventory (STAI-T; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998). Study 1 (section 2.1) included the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1989), while study 2 (section 2.2) and study 3 (section 2.3) used the Big Five Inventory (BFI-10; Rammstedt, Kemper, Beierlein, & Kovaleva, 2012) to assess personality traits. To apprehend the influence of certain character traits on gaze behavior we conducted correlation analyses between the time spent fixating faces using the free-viewing data of all experiments and personality attributes collected via the NEO-FFI and the BFI-10 (see Table 3.1).

Table 3.1. Pearson product-moment correlations (*r*) for 5 different personality traits, as surveyed by the Big Five Inventory (BFI-10; Rammstedt, Kemper, Beierlein, & Kovaleva, 2012) for eye tracking data of studies 2 and 3, and NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1989) for the eye tracking data of study 1 of this dissertation. The eye movement data is restricted to long viewing durations (≥ 5000 ms) directed towards the head of the depicted human being(s) in the social scenes of the used set of stimuli for each experiment. Measures reported include the mean (and standard deviation) of the listed personality trait and the correlation measures Pearson's *r*, as well as the significance level *p*.

Questionnaire	**Trait**	**Mean (*SD*)**	**Pearson's *r***	***p***
NEO-FFI	Neuroticism	17.93 (6.81)	.071	.665
	Extraversion	30.05 (6.32)	-.041	.801
	Openness	28.80 (7.47)	.214	.185
	Agreeableness	32.08 (5.64)	.173	.287
	Conscientiousness	33.85 (7.19)	.113	.487

Questionnaire	**Trait**	**Mean (*SD*)**	**Pearson's *r***	***p***
BFI-10	Neuroticism	5.31 (1.86)	.081	.531
	Extraversion	7.21 (1.78)	-.173	.180
	Openness	7.68 (1.93)	.282	.027
	Agreeableness	6.82 (1.81)	.208	.104
	Conscientiousness	7.19 (1.76)	-.016	.899

Note: The number of participants for the NEO-FFI questionnaire amounted to 40, and the combined studies using the BFI-10 questionnaire consisted of 62 subjects.

While results from the study of Wu and colleagues (2014) revealed that extraversion and agreeableness were related to greater gaze selection towards the eye region and openness to new experiences was associated with diminished gaze. Depicting three traits that are uniquely related to social attention, our analyses yielded only one positive correlation between attention towards human heads and openness to new experiences. This trait is associated with longer fixation durations (Rauthmann et al., 2012), as well as an increased eye fixation points (Matsumoto, Shibata, Seiji, Mori, & Shioe, 2010), reflecting action to seek information, personal relevance, and deeper processing. It has also been related to the quality of a social interaction (Berry & Hansen, 2000). Yet many studies lack correlational evidence between gaze behavior and personality traits within a healthy population (see e.g., Kennedy et al., 2017). However, it seems unlikely that the way we view the world should not be associated with differences in cognitive abilities (Haldemann, Stauffer, Troche, & Rammsayer, 2011), personality (Wu et al., 2014) or social behavior (Constantino et al., 2017; Wang et al., 2015) or vice versa. Nevertheless, the link between personality and gaze behavior offers an objective and practically applicable measure for social attention and interaction.

3.3.4 THE INFLUENCE OF FAMILIARITY

Gaze patterns are not only idiosyncratic, but also differ within an individual under certain circumstances, such as familiarity (Miellet, Caldara, & Schyns, 2011). Beginning in childhood, social interactions have significant impact on an individual's development (Ladd, 1999). The presence of caretakers and peers not only encourages socialization, but also increases the importance and familiarizes one with the presence of other human beings. This constant long-term exposure may drive social bias in contrast to other aspects within our surroundings. As such, McGugin, McKeeff, Tong, & Gauthier (2011) tested a specific group of car experts, which showed conflicting behavior when a search display simultaneously presented social images and the object of their expertise.

Specifically, reaction times for the detection of a human face was worse, when the distractor was a car compared to other distractor types. The authors concluded that the processing of faces may be affected in the presence of another highly familiar object for which participants inhabit a high proficiency. Electrophysiological evidence is similar, yielding a change (enhanced early negative component N170) when experts categorized words with corresponding images depicting objects in their domain of expertise (birds or dogs) relative to when the categorizations were outside of their domain (Tanaka & Curran, 2001). This early component has previously been found to be enhanced when participants view face stimuli (Eimer, 2000). These studies indicate that well-learned categories differ in their neurological signature from more unfamiliar categories at an early processing stage. However, this also suggests that human beings are experts in recognizing faces (Tanaka & Gauthier, 1997).

The aspect of familiarity was also investigated within a social context across faces. Initially, familiar faces draw attention more than novel ones and are difficult to avoid, even when instructed to do so (Ryan, Hannula, & Cohen, 2007). The latter aspect supports the findings of our first study, confirming that top-down control does not interfere with preferential attention for social cues. However, familiarity cannot account for the general social bias that we observe, as it implies memory-based effects requiring previous exposure (Althoff & Cohen, 1999), which was not the case in our studies. However, a potential explanation may revert back to the factor of social relevance and a general regard of fellow human beings as a familiar aspect. This would also be in line with the central debate around the fusiform face area, of which some say that it is involved in processing visual stimuli in domains of perceptual expertise (Gauthier, Tarr, & Anderson, 1999; but see Johnson, 2005). However, others state that this region is involved in processing domain-specific computational properties for selective activation for faces (Farah, Rabinowitz, Quinn, & Liu, 2000). According to Johnson (2005), there may also be an intermediate course, such that parts of the fusiform cortex become specialized for processing faces through cortical projections of the subcortical route, which initially causes the orientation towards faces. Through the cortical visual pathways, the subcortical route may enhance the face-sensitive activation of specific areas, such as the fusiform cortex, which then become "experts" at detecting these features.

The other way to interpret familiarity is by known identity. For one, this resumes the face inversion effect, as an unfamiliar order of features decreases recognition (J. W. Tanaka & Farah, 1993). Extremely familiar faces, like those of famous individuals (Jackson & Raymond, 2006, 2008), as well as the own face (Buttle & Raymond, 2003), seems to require fewer attentional resources than unknown faces. With fewer resources needed, one would expect less attention orienting towards familiar faces. However, studies have shown an increased orienting response to familiar as opposed to unfamiliar faces (Lancry-Dayan, Nahari, Ben-Shakhar, & Pertzov, 2018) and further highlighted the obligatory effects of memory on eye movements (Ryan et al., 2007).

A frequent remark made within the context of social attention is also the element of animacy and social constructs. Animacy refers to e.g. animals that are ascribed a social component, due to their vitality, but also with reference to their connection to human beings, such as pets. Furthermore, interactions between two animals or an animal and a human being could also be categorized as social attention. Social constructs refer to man-made objects, or items that are related to humans (e.g., signs, glasses, houses). Although all these aspects can be seen as social components, they elicit different behaviors from that of social attention defined by the mere interest in other human beings. Indeed, animals do attract attention and changes for animate objects are detected more rapidly than for inanimate objects in accordance with the *animate-monitoring hypothesis* based on evolutionary importance of detection (New et al., 2007). However, when animals and humans are presented simultaneously, saccades to humans are faster than those directed to animals (Crouzet et al., 2010; Rousselet, Macé, & Fabre-Thorpe, 2003). Hershler & Hochstein (2005) also describe a *pop out* effect of human faces, while animal faces are searched serially. This further indicates that the efficiency of finding animal faces among an increasing number of distractors would decrease, while participants would be just as effective in finding human faces independent of distractor variables. Conversely, the tendency to prioritize human faces was also shown to act as a reliable distraction, increasing search times for non-social targets (Simpson, Buchin, Werner, Worrell, & Jakobsen, 2014).

With regard to social constructs, text on signs represents an important daily aspect, which is also high in contrast, resulting in a trained prioritized orienting. Furthermore, faces and text have been defined as stimuli competing for cortical territory (Dehaene et al., 2010). The concept of *automaticity in reading* is well-known and the involuntary processing and lack of capacity requirements become evident in the widely-used Stroop Task (Stroop, 1935). Nevertheless, in direct comparison, human faces are detected more quickly than characters (Cerf et al., 2008), with face categorization occurring as early as 100 ms, whereas words require at least 200ms (Pegna et al., 2004). We found the same effect in our research group with a similar setup as study 2 (section 2.2; Jordan, 2018). Specifically, initial saccades were primarily directed towards the human side and fixation densities on heads also surpassed those on text, even when the text was more salient than the face in terms of low-level features.

3.4 NEURAL CORRELATES OF SOCIAL ATTENTION

Attention priority maps have been identified in several brain regions, such as the frontal eye fields (Serences & Yantis, 2007), precentral sulcus (Jerde, Merriam, Riggall, Hedges, & Curtis, 2012), lateral interparietal cortex (Bisley, 2003; Bisley & Goldberg, 2010; Gottlieb, Kusunoki, & Goldberg, 1998), and V4 (Mazer & Gallant, 2003), but also in the striate cortex (V1; Sprague & Serences, 2013; but see Mo, He, & Fang, 2017). Recently, attention priority maps have been generated for upright and inverted faces and could be predicted from the reconstructed topographic representations in V1-V3, indicating an involvement of these structures in face processing (Mo et al., 2017). Particularly, cortical representations of upright faces were more enhanced in V2/3 than for inverted ones. This indicates a contingency between stimulus configuration and an increase in functional coupling of neural activity and behavior along the visual pathway. It further highlights the importance of context and suggests an extension to the classical view of attention priority map theories consisting only of physical saliency and task goals (Fecteau & Munoz, 2006; Serences & Yantis, 2006; but see Mo et al., 2017).

It remains to be seen whether a specialized processing for social stimuli is generally reflected functionally in the brain, diverting from purely ventral and dorsal networks. Previous literature is discordant concerning social attention, finding activations of the ventral and lateral occipito-temporal cortex (medial superior temporal area (MT), posterior middle temporal gyrus (pMTG), right TPJ, precuneus, middle frontal gyrus (MFG), and superior frontal sulcus (STS)) in the presence of a social stimulus (Nardo et al., 2011). Others find activation in subcortical areas (superior colliculus, posterior thalamus, pulvinar and amygdala), especially in response to faces (for review see Johnson, 2005). A third opinion differentiates between the analysis of faces represented in a core system (occipito-temporal regions in extrastriate visual cortex), the meaning of faces through an extended system, and social relevance of information gleaned from faces through the amygdala (Haxby, Hoffman, & Gobbini, 2002). The TPJ is generally deemed a prominent candidate for social processing, due to its suggested involvement in theory of mind (see e.g., Saxe & Kanwisher, 2003). Furthermore, the TPJ is adjacent to the FFA (Adolphs, 2009), which implies an association to faces. However, TPJ activations seem to occur only after 300 ms (Menon et al., 1997), while attention towards social parts of a complex scene are already evident prior to 200 ms (Flechsenhar, Larson, End & Gamer, 2018; Fletcher-Watson et al., 2008; Rösler et al., 2017). Furthermore, fast eye movements would require TPJ to quickly detect social information in a scene within these short latencies, speaking against this assumption. A subcortical route would fulfil these requirement, as it receives direct projections from the retina (Tamietto et al., 2012) and prioritizes personally relevant information (Pessoa & Adolphs, 2010) within as early as 70 ms (McFadyen, Mermillod, Mattingley, Halász, & Garrido, 2017).

The amygdala is largely associated with the processing of fear (Morris, 1998; Whalen et al., 2001), even when these emotional faces are not consciously perceived (Whalen et al., 1998). Furthermore, it has been found to drive saccades towards fearful eyes (Gamer & Büchel, 2009). Prior to this emotional response, the amygdala is also sensitive to environmental contingencies and responds to unpredictable events (Herry et al., 2007), and is also relevant for associative learning (Whalen, 2007). As mentioned in the introduction, a much disputed function of the amygdala is its role in detecting and processing social stimuli. The so-called "low road" of a short and direct colliculus-pulvinar pathway to the amygdala is believed to transmit coarse visual information, while the "high road" is associated with fine-grained details via the visual cortex (LeDoux, 1998). Indeed, the amygdala responds quickly to fearful faces of low spatial frequency (LSF), while those presented in high spatial frequency elicited activation in extrastriate areas (Vuilleumier, Armony, Driver, & Dolan, 2003). However, McFadyen and colleagues (2017) found a subcortical amygdala connection regardless of spatial frequency or emotional expression. This renders the amygdala an evolved system for relevance detection (Sander, Grafman, & Zalla, 2003) that can detect a face (LSF), but also identify it (HSF) for a specific purpose. Hence, there are several arguments speaking in favor of a subcortical processing of faces compared to a cortical one: a temporal advantage over cortical processing (Tamietto & Gelder, 2010), a broad transmission of spatial frequencies, and a generalized role not specific to certain emotional facial expressions (McFadyen et al., 2017).

Pessoa & Adolphs (2010) argue against a subcortical route in social processing and suggest a cortical approach. While it is true that cortical regions, such as the FFA show strong and rapid activation during face perception, activity diminishes when faces are not directly attended, whereas amygdala activation is maintained (Vuilleumier, Armony, Driver, & Dolan, 2001). This indicates that cortical processes require resources, while amygdala coding is both mandatory and resource-independent (e.g., Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Williams, McGlone, Abbott, & Mattingley, 2005; but see Palermo & Rhodes, 2007). Therefore, the amygdala may tag biologically relevant stimuli, subsequently recruiting attentional resources and cortical processing to categorize the stimulus (Anderson et al., 2003), in turn potentially enhancing circuit breaking capacities of the ventral attention circuit (Corbetta & Shulman, 2002; Taylor & Fragopanagos, 2005). Through its extensive connectivity to the visual cortex, the amygdala may directly be able to regulate cortical perceptual processing and select the input received (Davis & Whalen, 2001; LeDoux, 2000).

Social cognition, as a higher-order function of basic social attention mechanisms, has also been associated with amygdala, orbito-frontal cortex (OFC) and superior temporal gyrus (STG) in Brothers' early work entitled "The Social Brain" (1990). She indicates two theories describing the core process of social cognition. First, emotions are intimately tied to the representations of other human beings. Second, human behavior evolved as a result of a variety of internal signals aiming at a correct response within a context. The amygdala influences both drive-related behavior and related emotions through connections with the septal area, the hypothalamus, and the PFC (Baron-Cohen et al., 2000). Evidence from clinical studies allows Brothers to conclude that social cognition appears to operate separately from all other domains of knowledge with a discrete neural system.

3.5 CLINICAL SIGNIFICANCE

Studies with patients suffering from specific psychiatric disorders or lesions provide a valuable source to investigate the functionality of specific brain regions. In fact, for many diseases, such as autism (e.g., Baron-cohen, Wheelwright, & Jolliffe, 1997), or social phobia (e.g., Boll, Bartholomaeus, Peter, Lupke, & Gamer, 2016), as well as psychopathy (e.g., Boll & Gamer, 2016), a social dysfunction is key factor as part of the diagnostic criteria (Adolphs, 2010). Specifically, individuals with autism exhibit abnormal viewing behavior towards the eyes (Fletcher-Watson et al., 2009; Klin et al., 2002a) and patients with social phobia elicit differences in attentional exploration of social cues, show a differential time course of eye gaze processing (Boll et al., 2016) and will prefer an object cue when given the choice in a dot-probe paradigm (Chen et al., 2002; MacLeod et al., 1986).

Individuals with autism depict an impairment in eye contact and response to gaze (Dawson et al., 2002; Lord et al., 2000), which is already evident in infancy (2-6 months of age; see Jones & Klin, 2013). This is in line with a focus on non-feature areas of the face and a deficit to recognize emotions (Pelphrey et al., 2002), or familiar faces (e.g., Klin et al., 1999). Instead, patients with autism elicit a general increase in pixel-level saliency at the expense of semantic-level saliency (Wang et al., 2015), which extends to similar attentional effects for animate objects (New et al., 2010). Furthermore, these individuals show less of an inversion effect for faces in combination with better object recognition compared to healthy subjects (e.g., Boucher & Lewis, 1992; Hauck, Fein, Maltby, Waterhouse, & Feinstein, 1998), indicating that they may adopt a segmental approach to face processing, rather than a holistic one (Pelphrey et al., 2002). Neuroimaging methods have shown a reduced activation of the right fusiform gyrus and greater involvement of the right inferior temporal gyri during face processing in autistic compared to non-autistic individuals (Klin, Jones, Schultz, Volkmar, & Cohen, 2002b). The abnormal fixation patterns for faces found in individuals with autism are also associated with amygdala dysfunction (Baron-Cohen et al., 2000; Dalton et al., 2005). Specifically, there seems to be a correlation between the activation in the fusiform gyrus and the amygdala with the time spent fixating the eyes of another individual. The resulting diminished activation of the fusiform gyrus with an augmented activation in the amygdala imply an association of face processing in affective central circuitry in autism that is not specific to the emotional content of a face, but a response to faces in general (Dalton et al., 2005). Therefore, face-processing deficits in autism may also be interpreted as an avoidance of overarousal caused by social stimuli, leading to the diminished focus on the eyes and a hyperactivation of the amygdala combined with atypical activation of the fusiform gyrus (Dalton et al., 2005). In line with this are studies testing individuals with generalized social phobia, which confirmed a higher reactivity of the amygdala for disorder-salient stimuli (e.g., Stein, Goldin, Sareen, Zorrilla, & Brown, 2017). Even healthy individuals depict differences in amygdala reactivity to emotional stimuli, which reflects a modulatory influence of higher cortical regions and indicates a lack of adequate cortical override to counter inordinate responsiveness (Hariri, Bookheimer, & Mazziotta, 2000). In addition to these functional differences in patients, structural distinctions have also been reported with regard to reduced amygdala volume in patients with autism (Abell et al., 1999). These volume reductions may account for morphological abnormalities at the sources of the amygdala inputs based on the assumption that the amygdala tags inputs with value or saliency (see Friston, Tononi, Reeke, Sporns, & Edelman, 1994) and may be associated with deficits of integrating emotional and social learning (Abell et al., 1999).

Some autistic traits descend onto relatives (e.g., Briskman, Frith, & Happé, 2001; Piven, 2001). However, the effect of atypical development of social skills seems to be bidirectional, as individuals with Williams syndrome show abnormally high interest in socially relevant information (Riby & Hancock, 2009). Depicting a hypersociability, persons with Williams syndrome also show relative strengths in processing facial identity (Bellugi, Lichtenberger, Jones, Lai, & St. George, 2000).

Similar to autism, patients with schizophrenia also elicit deficits of social cognition that may be related to amygdala dysfunction (Sasson et al., 2007), as both groups depict impairments in emotion processing and neurocognition (Eack et al., 2013). However, patients with Schizophrenia are faster in orienting towards faces than patients with autism, indicating that both groups share an abnormality in assessing facial information in social scenes, but differ in selecting relevant social information in complex stimuli (Sasson et al., 2007).

Patients with social phobia have reported an increased vigilance to social threat cues, as well as a tendency to excessively scan faces (Bradley, Mogg, & Millar, 2000), but also display an avoidance of socially relevant stimuli such as eye gaze (e.g., Heinrichs & Hofmann, 2001). Boll and colleagues (2016) investigated attentional mechanisms relevant for perceiving social cues by means of abnormalities in scanning of facial features in patients with social phobia. Their results not only reinforced the theory of hypervigilance for the eye region, but found that this elicited preference could be observed at very early stages (first saccade after 150 ms of stimulus presentation), speaking in favor of reflexive attentional orienting. Furthermore, patients with social phobia elicit higher amygdala activations during general emotion processing (Stein, Simmons, Feinstein, & Paulus, 2007). The other extreme of hypovigilance is found in individuals with psychopathic traits (e.g., Ceballos & Bauer, 2004) and deficits in emotional reactivity and recognition, showing correlations between character traits and reduced face exploration (Boll & Gamer, 2016). Reduced attention towards the eyes in combination with high psychopathic traits also results in problems of recognizing fear, which again is consistent with amygdala dysfunction failing to promote attention to emotionally salient aspects of facial expressions (Dadds, El Masry, Wimalaweera, & Guastella, 2008).

The studies of this dissertation were only conducted with healthy individuals, but several questionnaires characterizing the tested populations have been collected and analyzed (see section 3.3.3). As all experiments contained free-viewing data and a common set of questionnaires, we were able to calculate Pearson product-moment correlations concerning the eye tracking data and the characterizations for each participant across all projects (see Table 3.2). Correlations refer to sustained viewing behavior in free-viewing conditions targeting the head ROIs of depicted human beings. As discussed previously, individual differences in gaze allocation can correlate with certain character traits. Similarly, we investigated whether individuals with short fixation durations on faces would elicit certain traits related to clinical disorders, such as autism and (social) anxiety.

Table 3.2. Pearson product-moment correlations for 5 different questionnaires as surveyed by the Autism Spectrum Quotient (AQ-k; Freitag et al., 2007), State-Trait-Anxiety Inventory (STAI-T; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and the Social Interaction Anxiety Scale

(SIAS; Mattick & Clarke, 1998) for eye tracking data of all projects of this dissertation. The eye movement data is restricted to long viewing durations ($\geq$ 5000 ms) directed towards the head of the depicted human being(s) in the social scenes of the used set of stimuli across all experiments. Measures reported include the mean (and standard deviation) of the listed personality trait across all participants and the correlation measures Pearson's *r*, and the significance level *p*.

Questionnaire		***M* (*SD*)**	**Pearson's *r***	***p***
AQ-k				
	Social Interaction	1.37 (1.50)	-.063	.531
	Imagination	3.05 (2.27)	-.099	.319
	Communication	2.75 (1.48)	-.211	.033
	Sum	7.17 (3.85)	-.165	.098
SIAS		17.38 (8.51)	.029	.730
STAI		37.35 (9.65)	.043	.606

Note: The number of participants amounted to *n* = 102 for autism and *n* = 148 for anxiety measures.

Results were not significant with regard to anxiety measures, but a trend of a negative correlation was evident for the Autism Spectrum Quotient and the tendency to attend faces. This implies an avoidance to fixate human faces in line with increasing autistic traits, especially concerning decreased communicability. This is in line with previous studies, which found decreased gaze behavior towards the face with increasing autistic traits in healthy subjects (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Chen & Yoon, 2011; Freeth et al. 2013). Similar to findings of Chen and Yoon (2011), who investigated gaze patterns toward directed and averted eyes, a relationship between traits and eye movements did not generalize to a social anxiety measure in the general population. This indicates that the hypervigilance found in patients with social phobia, may not be evident at a subclinical level.

Other disorders depicting social dysfunction include those associated with actual brain damage, e.g. patients with visual spatial neglect. These individuals can only perceive one side of their visual field, depending on the affected brain regions. Vuilleumier (2000) discovered that such patients extinguished faces (real, as well as schematic ones) presented in their impaired hemifields less often than other categories, such as names, shapes or even scrambled faces. This indicates that extinction is modulated by the relevance of a stimulus and supports the idea that categorization occurs in the visual system before input from the other hemifield is included in attentive vision (Vuilleumier, 2000; Vuilleumier & Sagiv, 2001).

Patients with prosopagnosia have a specific disability to recognize faces while retaining an intact ability to perceive objects (De Renzi, 1986). This impairment is evident, when only the right hemisphere is damaged, indicating its important role in face processing (De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994), which is in line with neuroimaging studies showing greater activation in the right hemisphere during face processing tasks (e.g., Kanwisher et al., 1997; but see Pelphrey et al., 2002). The right hemisphere is primarily associated with relational encoding, while the left is related to feature based encoding (Farah, 2004), which may explain these differences. Though these patients have clear deficits in overt face matching, some cases are still able to match faces covertly, as depicted by similar ERP latencies between patients and healthy controls (Bobes et al., 2003). Covert face recognition may therefore be the consequence of partial and incomplete activation of person semantics (Burton, Bruce, & Hancock, 1990; Farah, O'reilly, & Vecera, 1993; but see Sperber & Spinnler, 2003). The suggested involvement of the subcortical network in face processing relates to this covert recognition (Nagai, 2007).

Patients with lesions to the amygdala have also yielded valuable insight concerning its relevance in biologically-relevant events. One patient (S.M.) with Urbach-Wiethe disease resulting in bilateral amygdala damage further highlighted the role of the amygdala in recognizing facial expressions (Adolphs, Tranel, Damasio, & Damasio, 1994). The patient showed impaired judgement of emotions from facial expressions, arising from an inability to attend informative features (Adolphs et al., 2005). This defect was established from a lack of spontaneous fixations towards the eyes during free viewing, as well as real social interactions (Spezio, Huang, Castelli, & Adolphs, 2007), due to the impairment of the amygdala to drive this orientation (Adolphs et al., 2005). Hence, the amygdala not only feeds back to the visual cortex, already modulating early visual information processing (Anderson & Phelps, 2001; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004), but might also influence the visual information that our eyes seek in the first place. This is in line with the aforementioned findings of Gamer & Büchel (2009). In a later publication, Adolphs (2010) concludes that the amygdala is indispensable to recognize multiple emotions in a single facial expression, but is not required to recognize personal identity from faces, providing further evidence for a double dissociation between processing of facial identity and of facial affect and therefore also separate neural systems. Interestingly, when explicitly instructed to use the eye region for emotion classification, S.M. was able to correctly identify facial expressions (Adolphs et al., 2005). This may indicate that a dysfunctional amygdala may induce an interference with normal viewing behavior due to a lack of top-down control, allowing low-level visual cues to guide over attention (Peters et al., 2005; but see Spezio et al., 2007). Indeed, S.M.'s deficit was imperceptible in a gaze-contingent presentation, where only a small region of the stimulus became visible at the center of the current gaze position monitored through online eye tracking (Kennedy & Adolphs, 2011). The authors concluded that the normalization of S.M.'s fixations to the eye region in this paradigm was due to a purely top-down driven gaze allocation. This finding substantiates the notion that the amygdala responds to help guide fixations to the most socially salient parts of the face (Adolphs, 2010; Gamer & Büchel, 2009; Whalen, 2007) and further questions the voluntary selection of social elements in healthy individuals (Flechsenhar, Rösler, & Gamer, 2018).

Patients with hemispatial neglect or blindsight have provided evidence for residual ability to detect faces, even though these disorders result in visual extinction to stimuli in the affected field (Morris, 2001; Vuilleumier, 2000; Vuilleumier & Sagiv, 2001). The condition of patient G.Y. who suffered damage to the human striate cortex resulting in blindness in the corresponding visual field (Holmes, 1918) supports the notion of a subcortical route for faces. Interestingly, this patient retained some residual ability to discriminate emotional facial expressions, supported by differential activity in the amygdala of fearful faces (Morris, 2001). This suggests that, while the striate cortex is crucial for conscious visual perception, it is not required for implicit visual processes, which seem to engage subcortical structures (Morris, 2001). Consequently, there may be separate routes for face detection and face identification (Johnson, 2005). As such, face detection may require a fast subcortical pathway (around 100 ms; see Liu, Harris, & Kanwisher, 2002), whereas face identification includes structural encoding of high spatial frequencies with the need for cortical processing. Naturally, these areas are not segregated, but indulge in communication within and across networks, i.e. PFC seems to be able to regulate the response of the amygdala to negatively as opposed to positively cued surprised faces (Kim et al., 2004), whereas the amygdala can store perceptual information and thereby influence memory, attention, decision making and other cognitive functions, even enhance cortical processing, like the lateral occipital, fusiform and orbitofrontal cortex for such biologically relevant stimuli (Johnson, 2005). Consequently, the amygdala may be involved in both initial, rapid perception of these stimuli, but also in later evaluation within a given context (Adolphs, 2003).

3.6 OUTLOOK

This converging evidence indicates a special type of processing for faces via a subcortical route, however, the question remains whether this is also the case for social stimuli in general embedded in a natural environment in competition with other potentially more salient aspects. The amygdala may be specialized for "tagging" social features in the visual field and subsequently alters activity in the early visual cortex to facilitate the processing of this information and program saccades towards social stimulus elements (Jazayeri, Lindbloom-Brown, & Horwitz, 2012). Attention orienting towards social features thus may be related to spatially specific activity and an increase in functional coupling between subcortical regions (superior colliculus, pulvinar amygdala) and visual cortex. Our research group is currently examining neural activity to briefly shown (200 ms) social images, as well as dynamic social scenes using combined eye tracking and fMRI. While one setup investigates attention selection of four different quadrants, in which one displays a human being, the other displays an attention-grabbing event occurring within social videos to compare neural activity for social as opposed to non-social fixations in direct competition and without the disappearance of the social aspect (see Nardo et al., 2011). If the results of our studies translate to neuroimaging, we would expect differential activation when fixations are directed towards social features with potential involvement of said subcortical regions. Neural correlates of non-social fixations are expected to adhere to the traditional dichotomy of ventral and dorsal pathway activations (Corbetta & Shulman, 2002). Although risky due to high complexity, this experiment may substantiate the previously mentioned subcortical route and extend it to social naturalistic dynamic scenes in direct comparison to non-social events, hence, allowing a fixation-dependent segregation of brain activity, specific for social attention. Future studies should further aim to identify the role of the amygdala and interactions with higher-level brain areas, and also take into account different amygdalar subnuclei to define the involvement in triggering attentional effects towards emotional and social aspects (Vuilleumier, 2005)

Even though eye movements have a very good temporal resolution, it is important to validate results with the use of other techniques, such as electroencephalography (EEG) or virtual environments (VR) to investigate social attention correlates. Herein, EEG may elicit distinct ERPs and elucidate the timing of social information processing in visual cortices. Specifically, induced alterations by social features concerning the laterality and amplitude of early visual evoked potentials that are considered a reflection of information processing in V1 (Jeffreys & Axford, 1972). Virtual reality closes the gap between laboratory based experiments and the real world, by using controlled, but realistic environments, incorporating important head and body movements to investigate interactions with avatars or changes in posture as further indicator of approach and avoidance behavior.

3.7 SUMMARY

Paying attention to fellow human beings seems like the most natural behavior. Through it, we communicate intentions, mood, emotionality, warnings, or intended interaction, which make up our every-day life. The underlying processes of these self-evident and ubiquitous actions prove to be highly complex. As stated by Compton (2003), social features, such as the face of another person, have such high emotional significance, that it is not only processed pre-attentively, but is also given priority in the competition for selective attention. The fact that social mechanisms are the foundation of every other higher-order capability and ultimately human behavior, makes the examination of its basic processes relevant for other disciplines. This dissertation has highlighted multiple components of basic social attention, aiming to disentangle singular processes and reinforce as well as extend previous literature. Specifically, we have demonstrated the influence of different attention components (bottom-up, top-down) and physical saliency. We have manipulated viewing conditions (task demands, overt attention, covert attention, gaze-contingent display) to put to test the uncovered social bias and discussed certain aspects that may influence social attention, such as familiarity, personality traits or clinical characteristics. We have argued neural implications, speaking in favor of a subcortical route for processing social stimuli, extending and substantiating this line of reasoning through clinical manifestations. Although, unequivocal neuroimaging evidence of the suggested subcortical route has yet to be concluded, the behavioral measures within this thesis have been profitable, suggesting strong and largely independent processing of social features. Finally, we have tried to highlight the importance to identify factors that influence social attention to better understand human social behavior in general (Freeth et al., 2013). Furthermore, these findings suggest that attentional mechanisms may generally be governed by multiple processes that select and organize sensory inputs for access to awareness (Driver, 2001; Vuilleumier, 2005). Profound knowledge of basic social attention functioning will allow for a comprehensive understanding of higher social cognition, reaching from theory of mind, judgement of threat or trustworthiness, to prosociality and motives guiding such behavior.

4. SUPPLEMENTARY MATERIAL

S1 SELECTED IMAGES OF DATABASES USED IN STUDY 2 (SECTION 2.2)

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

Database	**Reference**	**Content**
Emotional Picture Set	9.jpg	social
Emotional Picture Set	119.jpg	social
Emotional Picture Set	131.jpg	social
Emotional Picture Set	133.jpg	social
Emotional Picture Set	138.jpg	social
Emotional Picture Set	191.jpg	social
Emotional Picture Set	196.jpg	social
Emotional Picture Set	197.jpg	social
Emotional Picture Set	205.jpg	social
Emotional Picture Set	267.jpg	non-social
Emotional Picture Set	280.jpg	non-social
International Affective Picture System	5199.jpg	social
International Affective Picture System	9150.jpg	social
International Affective Picture System	9186.jpg	non-social
International Affective Picture System	9422.jpg	non-social
McGill Calibrated Colour Image Database	Merry_0005_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_0014_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_0060_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_0064_Lasalle.jpg	non-social
McGill Calibrated Colour Image Database	Merry_florida0011.jpg	social
McGill Calibrated Colour Image Database	Merry_florida0017.jpg	non-social
McGill Calibrated Colour Image Database	Merry_mexico0072.jpg	social
McGill Calibrated Colour Image Database	Merry_mexico0143.jpg	social
McGill Calibrated Colour Image Database	Merry_0081.jpg	non-social
McGill Calibrated Colour Image Database	Pippin_city6.jpg	social
McGill Calibrated Colour Image Database	Pippin_city66.jpg	social
Nencki Affective Picture System	Animals_025.jpg	non-social
Nencki Affective Picture System	Animals_048_h.jpg	non-social
Nencki Affective Picture System	Animals_074_h.jpg	non-social
Nencki Affective Picture System	Animals_102_h.jpg	non-social
Nencki Affective Picture System	Animals_128_h.jpg	non-social
Nencki Affective Picture System	Animals_194_h.jpg	non-social
Nencki Affective Picture System	Animals_195_h.jpg	non-social
Nencki Affective Picture System	Animals_201_h.jpg	non-social
Nencki Affective Picture System	Animals_218_h.jpg	non-social
Nencki Affective Picture System	Faces_023_h.jpg	social

Nencki Affective Picture System	Faces_265_h.jpg	social
Nencki Affective Picture System	Faces_290_h.jpg	social
Nencki Affective Picture System	Faces_302_h.jpg	social
Nencki Affective Picture System	Landscapes_016_h.jpg	non-social
Nencki Affective Picture System	Landscapes_025_h.jpg	non-social
Nencki Affective Picture System	Landscapes_040_h.jpg	non-social
Nencki Affective Picture System	Landscapes_043_h.jpg	non-social
Nencki Affective Picture System	Landscapes_064_h.jpg	non-social
Nencki Affective Picture System	Landscapes_071_h.jpg	non-social
Nencki Affective Picture System	Landscapes_085_h.jpg	non-social
Nencki Affective Picture System	Landscapes_178_h.jpg	non-social
Nencki Affective Picture System	Objects_002_h.jpg	non-social
Nencki Affective Picture System	Objects_013_h.jpg	non-social
Nencki Affective Picture System	Objects_058_h.jpg	non-social
Nencki Affective Picture System	Objects_183_h.jpg	non-social
Nencki Affective Picture System	Objects_202_h.jpg	non-social
Nencki Affective Picture System	Objects_214_h.jpg	non-social
Nencki Affective Picture System	People_009_h.jpg	social
Nencki Affective Picture System	People_015_h.jpg	social
Nencki Affective Picture System	People_022_h.jpg	social
Nencki Affective Picture System	People_054_h.jpg	social
Nencki Affective Picture System	People_058_h.jpg	social
Nencki Affective Picture System	People_109_h.jpg	social
Nencki Affective Picture System	People_116_h.jpg	social
Nencki Affective Picture System	People_131_h.jpg	social
Nencki Affective Picture System	People_157_h.jpg	social
Nencki Affective Picture System	People_158_h.jpg	social
Nencki Affective Picture System	People_167_h.jpg	social
Nencki Affective Picture System	People_182_h.jpg	social
Nencki Affective Picture System	People_195_h.jpg	social
Object and Semantic Images and Eyetracking dataset	118.jpg	non-social

5. REFERENCES

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K. J., Frackowiak, R., … Frith, U. (1999). The neuroanatomy of autism. *NeuroReport*, *10*(8), 1647–1651. https://doi.org/10.1097/00001756-199906030-00005

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*(2), 231–9.

Adolphs, R. (2003). Cognitive neuroscience: Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*(3), 165–178.

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. *Annual Review of Psychology*, *60*(1), 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Adolphs, R. (2010). Conceptual Challenges and Directions for Social Neuroscience. *Neuron*, *65*(6), 752–767. https://doi.org/10.1016/j.neuron.2010.03.006

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

Adolphs, R., & Spezio, M. (2006). Role of the amygdala in processing visual social stimuli. In Anders, Ende, Junghöfer, Kissler, & Wildgruber (Eds.), *Progress in Brain Research* (Vol. 156, pp. 363–378). Elsevier B.V. https://doi.org/10.1016/S0079-6123(06)56020-0

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

Allport, D. A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. *Perspectives on Perception and Action*, *15*, 395–419.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997–1010. https://doi.org/10.1037/0278-7393.25.4.997

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

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. E. (2003). Neural Correlates of the Automatic Processing of Threat Facial Signals. *The Journal of Neuroscience*, *23*(13), 5627–5633. https://doi.org/10.1523/JNEUROSCI.23-13-05627.2003

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*(6835), 305–309. https://doi.org/10.1038/35077083

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

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it:

Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, *15*(5), 9. https://doi.org/10.1167/15.5.9

Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, *39*(17), 2947–2953. https://doi.org/10.1016/S0042-6989(99)00019-X

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103*(1), 62–70. https://doi.org/10.1037/h0036799

Argyle, M., Cook, M., & Cramer, D. (1994). Gaze and Mutual Gaze. *British Journal of Psychiatry*, *165*(06), 848–850. https://doi.org/10.1017/S0007125000073980

Asmundson, G. J., & Stein, M. B. (1994). Selective Processing of Social Threat in Patients With Generalized Social Phobia : Evaluation Using a Dot-Probe Paradigm. *Journal of Anxiety Disorders*, *8*(2), 107–117. https://doi.org/10.1016/0887-6185(94)90009-4

Baddeley, A., Rubak, E., & Turner, R. (2015). Spatial Point Patterns: Methodology and Applications with R. London: Chapman and Hall/CRC Press. Retrieved from http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/

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

Bantin, T., Stevens, S., Gerlach, A. L., & Hermann, C. (2016). What does the facial dot-probe task tell us about attentional processes in social anxiety? A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry*, (50), 40–51. https://doi.org/10.1016/j.jbtep.2015.04.009

Bar-Haim, Y., Lamy, D., 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., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, *13*(4), 379–398. https://doi.org/10.1111/j.2044-835X.1995.tb00687.x

Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: evidence for the role of perception in the development of a theory of mind. *Mind & Language*, *17*(1–2), 172–186.

Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., & Williams, S. C. R. (2000). The amygdala theory of autism. *Neuroscience & Biobehavioral Reviews*, *24*(3), 355–364. https://doi.org/10.1016/S0149-7634(00)00011-7

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is There a “Language of the Eyes”? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. *Visual Cognition*, *4*(3), 311–331. https://doi.org/10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient : Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., & Haxby, J. V. (2001). A Parametric fMRI

Study of Overt and Covert Shifts of Visuospatial Attention. *NeuroImage*, *14*(2), 310–321. https://doi.org/10.1006/nimg.2001.0788

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). The Neurocognitive Profile of Williams Syndrome: A Complex Pattern of Strengths and Weaknesses. *Journal of Cognitive Neuroscience*, *12*(supplement 1), 7–29. https://doi.org/10.1162/089892900561959

Belopolsky, A. V., & Theeuwes, J. (2012). Updating the premotor theory: The allocation of attention is not always accompanied by saccade preparation. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 902–914. https://doi.org/10.1037/a0028662

Berry, D. S., & Hansen, J. S. (2000). Personality, Nonverbal Behavior, and Interaction Quality in Female Dyads. *Personality and Social Psychology Bulletin*, *26*(3), 278–292. https://doi.org/10.1177/0146167200265002

Biederman, I. (1972). Perceiving Real-World Scenes. *Science*, *177*(4043), 77–80. https://doi.org/10.1126/science.177.4043.77

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

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

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008a). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2–3), 341–355. https://doi.org/10.1080/13506280701434532

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). Social Attention and Real-World Scenes: The Roles of Action, Competition and Social Content. *Quarterly Journal of Experimental Psychology*, *61*(7), 986–998. https://doi.org/10.1080/17470210701410375

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

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

Birmingham, E., Cerf, M., & Adolphs, R. (2011). Comparing social attention in autism and amygdala lesions: Effects of stimulus and task condition. *Social Neuroscience*, *6*(5–6), 420–435. https://doi.org/10.1080/17470919.2011.561547

Birmingham, E., & Kingstone, A. (2009). Human Social Attention. A New Look at Past , Present , and Future Investigations. *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. (2003). Neuronal Activity in the Lateral Intraparietal Area and Spatial Attention. *Science*, *299*(5603), 81–86. https://doi.org/10.1126/science.1077395

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

Bobes, M. A., Lopera, F., Garcia, M., Díaz-Comas, L., Galan, L., & Valdes-Sosa, M. (2003). Covert matching of unfamiliar faces in a case of prosopagnosia: An ERP study. *Cortex*, *39*(1), 41–56. https://doi.org/10.1016/S0010-9452(08)70073-X

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Observing shared attention modulates gaze following. *Cognition*, *120*(2), 292–298. https://doi.org/10.1016/j.cognition.2011.05.002

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

Boll, S., & Gamer, M. (2016). Psychopathic traits affect the visual exploration of facial expressions. *Biological Psychology*, *117*, 194–201. https://doi.org/10.1016/j.biopsycho.2016.03.010

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

Borji, A., Sihite, D. N., & Itti, L. (2013a). Objects do not predict fixations better than early saliency: A re-analysis of Einhauser et al.'s data. *Journal of Vision*, *13*(10), 18–18. https://doi.org/10.1167/13.10.18

Borji, A., Sihite, D. N., & Itti, L. (2013b). What stands out in a scene? A study of human explicit saliency judgment. *Vision Research*, *91*, 62–77. https://doi.org/10.1016/j.visres.2013.07.016

Boucher, J., & Lewis, V. (1992). Unfamiliar Face Recognition in Relatively Able Autistic Children. *Journal of Child Psychology and Psychiatry*, *33*(5), 843–859. https://doi.org/10.1111/j.1469-7610.1992.tb01960.x

Bradley, B. P., Mogg, K., Falla, S. J., & Hamilton, L. R. (1998). Attentional Bias for Threatening Facial Expressions in Anxiety: Manipulation of Stimulus Duration. *Cognition & Emotion*, *12*(6), 737–753. https://doi.org/10.1080/026999398379411

Bradley, B. P., Mogg, K., & Millar, N. H. (2000). Covert and overt orienting of attention to emotional faces in anxiety. *Cognition & Emotion*, *14*(6), 789–808. https://doi.org/10.1080/02699930050156636

Bradley, B. P., Mogg, K., White, J., Groom, C., & Bono, J. (1999). Attentional bias for emotional faces in generalized anxiety disorder. *British Journal of Clinical Psychology*, *38*(3), 267–278. https://doi.org/10.1348/014466599162845

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

Briskman, J., Frith, U., & Happé, F. (2001). Exploring the Cognitive Phenotype of Autism: Weak "Central Coherence" in Parents and Siblings of Children with Autism: II. Real-life Skills and Preferences. *Journal of Child Psychology and Psychiatry*, *42*(3), S0021963001006904. https://doi.org/10.1017/S0021963001006904

Broadbent, D. E. (1958). Perception and Communication. London: Pergamon Press.

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.*, *1*, 27–51.

Burton, A. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, *23*(1), 1–31. https://doi.org/10.1016/S0364-0213(99)80050-0

Bush, J. C., Pantelis, P. C., Morin Duchesne, X., Kagemann, S. A., & Kennedy, D. P. (2015). Viewing Complex, Dynamic Scenes "Through the Eyes" of Another Person: The Gaze-Replay Paradigm.

PLOS ONE, *10*(8), e0134347. https://doi.org/10.1371/journal.pone.0134347

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Perception & Psychophysics*, *65*(8), 1296–1306. https://doi.org/10.3758/BF03194853

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., & Oliva, A. (2016). MIT Saliency Benchmark. Retrieved June 29, 2018, from http://saliency.mit.edu/

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., & Henson, R. N. A. (2007). Separate Coding of Different Gaze Directions in the Superior Temporal Sulcus and Inferior Parietal Lobule. *Current Biology*, *17*(1), 20–25. https://doi.org/10.1016/j.cub.2006.10.052

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

Canli, T. (2006). *Biology of personality and individual differences.* Guilford Press.

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

Castelhano, M. S., & Henderson, J. M. (2008). The Influence of Color on the Perception of Scene Gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 660–675. https://doi.org/10.1037/0096-1523.34.3.660

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 L. Paletta & E. Rome (Eds.), *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint* (pp. 251–262). Berlin: Springer. https://doi.org/10.1007/978-3-540-77343-6_16

Ceballos, N. A., & Bauer, L. O. (2004). Effects of Antisocial Personality, Cocaine and Opioid Dependence, and Gender on Eye Movement Control. *Psychological Reports*, *95*(2), 551–563. https://doi.org/10.2466/pr0.95.2.551-563

Cerf, M., Frady, E. P., & Koch, C. (2008). Using semantic content as cues for better scanpath prediction. In *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08* (p. 143). New York, New York, USA: ACM Press. https://doi.org/10.1145/1344471.1344508

Chapman, A., Devue, C., & Grimshaw, G. M. (2017a). Fleeting reliability in the dot-probe task. *Psychological Research*, 241–248. https://doi.org/10.1007/s00426-017-0947-6

Chapman, A., Devue, C., & Grimshaw, G. M. (2017b). Fleeting reliability in the dot-probe task. *Psychological Research*, 1–13. https://doi.org/10.1007/s00426-017-0947-6

Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Research*, *85*, 58–72. https://doi.org/10.1016/j.visres.2012.12.005

Chen, F. S., & Yoon, J. M. D. (2011). Brief Report: Broader Autism Phenotype Predicts Spontaneous Reciprocity of Direct Gaze. *Journal of Autism and Developmental Disorders*, *41*(8), 1131–1134. https://doi.org/10.1007/s10803-010-1136-2

Chen, Y. P., Ehlers, A., Clark, D. M., & Mansell, W. (2002). Patients with generalized social phobia direct their attention away from faces. *Behaviour Research and Therapy*, *40*(6), 677–687.

https://doi.org/10.1016/S0005-7967(01)00086-9

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12629–12633. https://doi.org/10.1073/pnas.0506162102

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A Taxonomy of External and Internal Attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Cohen, J. (1988). *Statistical Power Analysis For The Behavioral Sciences. Lawrence Erlbaum Associates.*

Compton, R. J. (2003). The Interface Between Emotion and Attention: A Review of Evidence from Psychology and Neuroscience. *Behavioral and Cognitive Neuroscience Reviews*, *2*(2), 115–129. https://doi.org/10.1177/1534582303002002003

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

Cooper, R. M., & Langton, S. R. H. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour Research and Therapy*, *44*(9), 1321–1329. https://doi.org/10.1016/j.brat.2005.10.004

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., … Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761–773. https://doi.org/10.1016/S0896-6273(00)80593-0

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.

Corbetta, M., & Shulman, G. L. (2002). Control of Goal-Directed and Stimulus-Driven Attention in the Brain. *Nature Reviews Neuroscience*, *3*(3), 215–229. https://doi.org/10.1038/nrn755

Costa, P. T., & McCrae, R. R. (1989). NEO five-factor inventory (NEO-FFI). Odessa, FL: Psychological Assessment Resources.

Coull, J. . (1998). Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. *Progress in Neurobiology*, *55*(4), 343–361. https://doi.org/10.1016/S0301-0082(98)00011-2

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of Vision*, *10*(4), 16.1-17. https://doi.org/10.1167/10.4.16

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*(5), 736–760.

Dadds, M. R., El Masry, Y., Wimalaweera, S., & Guastella, A. J. (2008). Reduced Eye Gaze Explains "Fear Blindness" in Childhood Psychopathic Traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(4), 455–463. https://doi.org/10.1097/CHI.0b013e31816407f1

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., … Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature*

Neuroscience, *8*(4), 519–526. https://doi.org/10.1038/nn1421

Davis, M., & Whalen, P. J. (2001). The amygdala: vigilance and emotion. *Molecular Psychiatry*, *6*(1), 13–34. https://doi.org/10.1038/sj.mp.4000812

Dawson, G., Webb, S., Schellenberg, G. D., Dager, S., Friedman, S., Aylward, E., & Richards, T. (2002). *Defining the broader phenotype of autism: Genetic, brain, and behavioral perspectives. Development and Psychopathology* (Vol. 14). https://doi.org/10.1017/S0954579402003103

De Haan, B., Morgan, P. S., & Rorden, C. (2008). Covert orienting of attention and overt eye movements activate identical brain regions. *Brain Research*, (1204), 102–111. https://doi.org/10.1016/j.brainres.2008.01.105

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

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

De Renzi, E. (1986). *Current issues on prosopagnosia. In Aspects of face processing.* Netherlands: Springer.

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere—An MRI and PET study and a review of the literature. *Neuropsychologia*, *32*(8), 893–902. https://doi.org/10.1016/0028-3932(94)90041-8

Deaner, R. O., & Platt, M. L. (2003). Reflexive Social Attention in Monkeys and Humans. *Current Biology*, *13*(18), 1609–1613. https://doi.org/10.1016/j.cub.2003.08.025

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

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

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.ne.18.030195.001205

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

Devue, C., Belopolsky, A. V, & Theeuwes, J. (2012). Oculomotor Guidance and Capture by Irrelevant Faces. *PLoS ONE*, *7*(4), e34598. https://doi.org/10.1371/journal.pone.0034598

DeYoung, C. G., & Gray, J. R. (2009). Personality neuroscience: explaining individual differences in affect, behaviour and cognition. In P. J. Corr & G. Matthews (Eds.), *The Cambridge Handbook of Personality Psychology* (pp. 323–346). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511596544.023

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

Domes, G., Sibold, M., Schulze, L., Lischke, A., Herpertz, S. C., & Heinrichs, M. (2013). Intranasal oxytocin increases covert attention to positive social cues. *Psychological Medicine*, *43*(08), 1747–1753. https://doi.org/10.1017/S0033291712002565

Donk, M., & Soesman, L. (2010). Salience is only briefly represented: Evidence from probe-detection performance. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 286–302. https://doi.org/10.1037/a0017605

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

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*(1), 53–78. https://doi.org/10.1348/000712601162103

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

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, and Computers*, *34*(4), 455–470. https://doi.org/10.3758/BF03195475

Duchowski, A. T. (2017). *Eye Tracking Methodology* (Third Edit). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57883-5

Eack, S. M., Bahorik, A. L., McKnight, S. A. F., Hogarty, S. S., Greenwald, D. P., Newhill, C. E., … Minshew, N. J. (2013). Commonalities in social and non-social cognitive impairments in adults with autism spectrum disorder and schizophrenia. *Schizophrenia Research*, *148*(1–3), 24–28. https://doi.org/10.1016/j.schres.2013.05.013

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & Psychophysics*, *63*(6), 1004–1013. https://doi.org/10.3758/BF03194519

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology*, *111*(4), 694–705. https://doi.org/10.1016/S1388-2457(99)00285-0

Eimer, M., Velzen, J. Van, Gherri, E., & Press, C. (2007). ERP correlates of shared control mechanisms involved in saccade preparation and in covert attention. *Brain Research*, *1135*(1), 154–166. https://doi.org/10.1016/j.brainres.2006.12.007

Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*(5), 1089–1097.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 2.1-19. https://doi.org/10.1167/8.2.2

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18.1-26. https://doi.org/10.1167/13.10.18

Eizenman, M., Yu, L. H., Grupp, L., Eizenman, E., Ellenbogen, M., Gemar, M., & Levitan, R. D. (2003). A naturalistic visual scanning approach to assess selective attention in major depressive disorder.

Psychiatry Research, *118*(2), 117–128. https://doi.org/10.1016/S0165-1781(03)00068-4

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3), 3. https://doi.org/10.1167/8.3.3

End, A., & Gamer, M. (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, *8*, 418. https://doi.org/10.3389/fpsyg.2017.00418

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Epstein, R., Stanley, D., Harris, A., & Kanwisher, N. (2000). The parahippocampal place area: perception, encoding, or memory retrieval. *Neuron*, *23*(1), 115–125.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*(2), 201–204. https://doi.org/10.3758/BF03212870

Fairhall, S. L., Indovina, I., Driver, J., & Macaluso, E. (2009). The Brain Network Underlying Serial Visual Search: Comparing Overt and Covert Spatial Orienting, for Activations and for Effective Connectivity. *Cerebral Cortex*, *19*(12), 2946–2958. https://doi.org/10.1093/cercor/bhp064

Farah, M. J. (2004). *Visual Agnosia*. MIT Press.

Farah, M. J., O'Reilly, R. C., & Vecera, S. P. (1993). Dissociated overt and covert recognition as an emergent property of a lesioned neural network. *Psychological Review*, *100*(4), 571–588. https://doi.org/10.1037/0033-295X.100.4.571

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). EARLY COMMITMENT OF NEURAL SUBSTRATES FOR FACE RECOGNITION. *Cognitive Neuropsychology*, *17*(1–3), 117–123. https://doi.org/10.1080/026432900380526

Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of Eye Gaze Perception during Infancy. *Journal of Cognitive Neuroscience*, *16*(8), 1320–1326.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–91.

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

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

Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524793.001.0001

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *NeuroImage*, *30*(4), 1441–1448. https://doi.org/10.1016/j.neuroimage.2005.11.003

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

Flechsenhar, A., Larson, O., End, A., & Gamer, M. (2018). Investigating overt and covert shifts of attention within social naturalistic scenes. *Journal of Vision*, *18*(12), 11. https://doi.org/10.1167/18.12.11

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

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid Detection of Person Information in a Naturalistic Scene. *Perception*, *37*(4), 571–583. https://doi.org/10.1068/p5705

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia*, *47*(1), 248–257. https://doi.org/10.1016/j.neuropsychologia.2008.07.016

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030. https://doi.org/10.1037/0096-1523.18.4.1030

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

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

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

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

Foulsham, T., Walker, E., & Kingstone, A. (2011b). 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

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

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., … Von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - Die kurzversion AQ-k. *Zeitschrift Fur Klinische Psychologie Und Psychotherapie*, *36*(4), 280–289.

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

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual Search for Faces With Emotional Expressions. *Psychological Bulletin*, *134*(5), 662–676. https://doi.org/10.1037/0033-2909.134.5.662

Friston, K. J., Tononi, G., Reeke, G. N., Sporns, O., & Edelman, G. M. (1994). Value-dependent selection in the brain: Simulation in a synthetic neural model. *Neuroscience*, *59*(2), 229–243. https://doi.org/10.1016/0306-4522(94)90592-4

Frith, C. D., & Frith, U. (2008). Perspective Implicit and Explicit Processes in Social Cognition. *Neuron*, *60*(3), 503–510. https://doi.org/10.1016/j.neuron.2008.10.032

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

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., 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

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573. https://doi.org/10.1038/9224

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2013). Learning and memory. In J. Durbin (Ed.), *Cognitive neuroscience: The biology of the mind.* (IV Edition, pp. 312–363). New York: W.W. Norton & Company.

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

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

Gilchrist, I. D. (2011). *Saccades. The Oxford handbook of eye movements*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0005

Gomes, N., Soares, S. C., Silva, S., & Silva, C. F. (2017). Mind the Snake: Fear Detection Relies on Low Spatial Frequencies. *Emotion*. https://doi.org/10.1037/emo0000391

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

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person × situation perspective. *Journal of Personality and Social Psychology*, *93*(4), 583–599. https://doi.org/10.1037/0022-3514.93.4.583

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers’ task from eye movement patterns. *Vision Research*, *62*, 1–8. https://doi.org/10.1016/j.visres.2012.03.019

Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Phil. Trans. R. Soc. Lond. B*, *335*(1273), 3–10. https://doi.org/10.1098/rstb.1992.0001

Haldemann, J., Stauffer, C., Troche, S., & Rammsayer, T. (2011). Processing Visual Temporal Information and Its Relationship to Psychometric Intelligence. *Journal of Individual Differences*, *32*(4), 181–188. https://doi.org/10.1027/1614-0001/a000050

Hannula, D. E. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*(October), 1–16. https://doi.org/10.3389/fnhum.2010.00166

Harel, J., Koch, C., & Perona, P. (2007). Graph-Based Visual Saliency. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.), *Advances in neural information processing systems* (pp. 545–552). Cambridge: MIT Press.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neurocortical network on the limbic system. Neuroreport: For Rapid Communication of. *Neuroscience Research*, *11*(1), 43–48.

Hauck, M., Fein, D., Maltby, N., Waterhouse, L., & Feinstein, C. (1998). Memory for Faces in Children with Autism. *Child Neuropsychology (Neuropsychology, Development and Cognition: Section C)*, *4*(3), 187–198. https://doi.org/10.1076/chin.4.3.187.3174

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

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol Psychiatry*, *51*(1), 59–67. https://doi.org/10.1016/S0006-3223(01)01330-0

Hayhoe, M. M. (2004). Advances in Relating Eye Movements and Cognition. *Infancy*, *6*(2), 267–274.

Heinrichs, N., & Hofmann, S. G. (2001). Information processing in social phobia: a critical review. *Clin Psychol Rev*, *21*(5), 751–770. https://doi.org/10.1016/S0272-7358(00)00067-2

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye Movements* (p. 537–III). Elsevier. https://doi.org/10.1016/B978-008044980-7/50027-6

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

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

Henderson, J. M., Williams, C. C., Castelhano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception & Psychophysics*, *65*(5), 725–734. https://doi.org/10.3758/BF03194809

Hermens, F., & Walker, R. (2010). Gaze and Arrow Distractors Influence Saccade Trajectories Similarly. *Quarterly Journal of Experimental Psychology*, *63*(11), 2120–2140.

https://doi.org/10.1080/17470211003718721

Herry, C., Bach, D., Esposito, F., Di Salle, F., Perrig, W., Scheffler, K., … Seifritz, E. (2007). Processing of Temporal Unpredictability in Human and Animal Amygdala. *Journal of Neuroscience*, *27*(22), 5958–5966. https://doi.org/10.1523/JNEUROSCI.5218-06.2007

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

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795. https://doi.org/10.3758/BF03206794

Holmes, G. (1918). Disturbances of Visual Orientation. *British Journal of Ophthalmology*, *2*(10), 506–516. https://doi.org/10.1136/bjo.2.10.506

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

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not see eye to eye: a visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders*, *17*(1), 33–44.

Huang, X., Shen, C., Boix, X., & Zhao, Q. (2015). SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks. In *2015 IEEE International Conference on Computer Vision (ICCV)* (Vol. 64, pp. 262–270). IEEE. https://doi.org/10.1109/ICCV.2015.38

Huestegge, L., & Koch, I. (2010). Fixation disengagement enhances peripheral perceptual processing: Evidence for a perceptual gap effect. *Experimental Brain Research*, *201*(4), 631–640. https://doi.org/10.1007/s00221-009-2080-2

Humphrey, K., & Underwood, G. (2009). Domain knowledge moderates the influence of visual saliency in scene recognition. *British Journal of Psychology*, *100*(2), 377–398. https://doi.org/10.1348/000712608X344780

Hunt, A. R., & Kingstone, A. (2003). Covert and overt voluntary attention: linked or independent? *Cognitive Brain Research*, *18*(1), 102–105. https://doi.org/10.1016/j.cogbrainres.2003.08.006

Hunt, A. R., & Kingstone, A. (2003). Inhibition of Return : Dissociating Attentional and Oculomotor Components. *Journal of Experimental Psychology - Human Perception and Performance*, *29*(5), 1068–1074. https://doi.org/10.1037/0096-1523.29.5.1068

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

Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nature Neuroscience*, *7*(1), 56–64. https://doi.org/10.1038/nn1169

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed Neural Systems for the Generation of Visual Images. *Neuron*, *28*(3), 979–990. https://doi.org/10.1016/S0896-6273(00)00168-9

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

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

Itti, L., & Koch, C. (2001). Computational Modelling of Visual Attention. *Nature Reviews. Neuroscience*, *2*(March), 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. https://doi.org/10.1109/34.730558

Jackson, M. C., & Raymond, J. E. (2006). The role of attention and familiarity in face identification. *Perception & Psychophysics*, *68*(4), 543–557. https://doi.org/10.3758/BF03208757

Jackson, M. C., & Raymond, J. E. (2008). Familiarity Enhances Visual Working Memory for Faces. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 556–568. https://doi.org/10.1037/0096-1523.34.3.556

James, W. (1890). *The Principles of Psychology*. New York: Holt and Company.

Jazayeri, M., Lindbloom-Brown, Z., & Horwitz, G. D. (2012). Saccadic eye movements evoked by optogenetic activation of primate V1. *Nature Neuroscience*, *15*(10), 1368–1370. https://doi.org/10.1038/nn.3210

Jeffreys, D. A., & Axford, J. G. (1972). Source locations of pattern-specific components of human visual evoked potentials. I. Component of striate cortical origin. *Experimental Brain Research*, *16*(1), 1–21. https://doi.org/10.1007/BF00233371

Jensen-Campbell, L. A., Rosselli, M., Workman, K. A., Santisi, M., Rios, J. D., & Bojan, D. (2002). Agreeableness, conscientiousness, and effortful control processes. *Journal of Research in Personality*, *36*(5), 476–489. https://doi.org/10.1016/S0092-6566(02)00004-1

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

Jiang, M., Huang, S., Duan, J., & Zhao, Q. (2015). SALICON : Saliency in Context. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1072–1080). https://doi.org/10.1109/CVPR.2015.7298710

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

Jones, W., & Klin, A. (2013). NIH Public Access. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715.Attention

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. *Attention and Performance*, *9*, 187–203.

Jonides, J. (1983). Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society*, *21*(4), 247–250.

Jordan, A.-L. (2018). *Soziale Aufmerksamkeit vs. Leseautomatismus: die Konkurrenz von sozialen Stimuli und Textinhalten beim Betrachten natürlicher Szenen*. Julius-Maximilians-Unversität Würzburg.

Joseph, J. S., & Optican, L. M. (1996). Involuntary attentional shifts due to orientation differences.

Perception and Psychophysics, *58*(5), 651–665. https://doi.org/10.3758/BF03213098

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *2009 IEEE 12th International Conference on Computer Vision* (Vol. 1, pp. 2106–2113). Cambridge, MA: IEEE. https://doi.org/10.1109/ICCV.2009.5459462

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

Kaspar, K., & König, P. (2012). Emotions and personality traits as high-level factors in visual attention: a review. *Frontiers in Human Neuroscience*, *6*(November), 1–14. https://doi.org/10.3389/fnhum.2012.00321

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

Kehrlein, E., & Gamer, M. (2016). *Zusammenhang zwischen sozialer Aufmerksamkeit und Interpretation bei Betrachtung komplexer naturalistischer Szenen*. Julius-Maximilians-University Würzburg.

Kelly, S. P., Lalor, E., Finucane, C., & Reilly, R. B. (2004). A comparison of covert and overt attention as a control option in a steady-state visual evoked potential-based brain computer interface. In H. J. Trussell (Ed.), *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 2, pp. 4725–4728). San Francisco, USA: IEEE. https://doi.org/10.1109/IEMBS.2004.1404308

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychological Medicine*, *37*(05), 615. https://doi.org/10.1017/S0033291706009524

Kennedy, D. P., & Adolphs, R. (2011). Reprint of: Impaired fixation to eyes following amygdala damage arises from abnormal bottom-up attention. *Neuropsychologia*, *49*(4), 589–595. https://doi.org/10.1016/j.neuropsychologia.2011.02.026

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

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual Modulation of Amygdala Responsivity to Surprised Faces. *Journal of Cognitive Neuroscience*, *16*(10), 1730–1745. https://doi.org/10.1162/0898929042947865

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, *112*(1), 24–38. https://doi.org/10.1037/0033-2909.112.1.24

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology*, *19*(1), 52–56. https://doi.org/10.1016/j.conb.2009.05.004

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2002). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science*, *12*(5), 1–17. https://doi.org/10.1111/1467-8721.01255

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Klein, J. T., Shepherd, S. V., & Platt, M. L. (2009). Social Attention and the Brain. *Current Biology*, *19*(20), R958–R962. https://doi.org/10.1016/j.cub.2009.08.010

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

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

Klin, A., Jones, W., Schultz, R. T., Volkmar, F. R., & Cohen, D. (2002a). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism. *Archives of General Psychiatry*, *59*(9), 809. https://doi.org/10.1001/archpsyc.59.9.809

Klin, A., Jones, W., Schultz, R., Volkmar, F. R., & Cohen, D. (2002b). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism. *Archives of General Psychiatry*, *59*(9), 809. https://doi.org/10.1001/archpsyc.59.9.809

Klin, A., Sparrow, S. S., De Bildt, A., Cicchetti, D. V., Cohen, D. J., & Volkmar, F. R. (1999). A normed study of face recognition in autism and related disorders. *Journal of Autism and Developmental Disorders*, *29*(6), 499–508. https://doi.org/10.1023/A:1022299920240

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.

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

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision*, *14*(3), 14–14. https://doi.org/10.1167/14.3.14

Kootstra, G., De Boer, B., & Schomaker, L. (2011). Predicting eye fixations on complex visual stimuli using local symmetry. *Cognit Comput*, *3*(1), 223–240. https://doi.org/10.1007/s12559-010-9089-5

Koval, M. J., Thomas, B. S., & Everling, S. (2005). Task-dependent effects of social attention on saccadic reaction times. *Experimental Brain Research*, *167*(3), 475–480. https://doi.org/10.1007/s00221-005-0206-8

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916. https://doi.org/10.1016/0042-6989(94)00279-U

Kramer, S. H. A. F. (1998). Further Evidence for the Division of Attention Among Non-contiguous Locations. *Visual Cognition*, *5*(1–2), 217–256. https://doi.org/10.1080/713756781

Kroeze, S., & van den Hout, M. A. (2000). Selective attention for cardiac information in panic patients. *Behaviour Research and Therapy*, *38*(1), 63–72. https://doi.org/10.1016/S0005-7967(99)00023-6

Krolak-Salmon, P., Fischer, C., Vighetto, A., & Mauguière, F. (2001). Processing of facial emotional expression: spatio-temporal data as assessed by scalp event-related potentials. *European Journal of Neuroscience*, *13*(5), 987–994. https://doi.org/10.1046/j.0953-816x.2001.01454.x

Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics*, *71*(2), 314–327. https://doi.org/10.3758/APP

Kuhn, G., & Land, M. F. (2006). There's more to magic than meets the eye. *Current Biology*, *16*(22), R950–R951. https://doi.org/10.1016/j.cub.2006.10.012

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition*, *17*(6–7), 925–944. https://doi.org/10.1080/13506280902826775

Kuhn, G., & Teszka, R. (2017). Don't get misdirected! Differences in overt and covert attentional inhibition between children and adults. *Quarterly Journal of Experimental Psychology*, 1–9. https://doi.org/10.1080/17470218.2016.1277770

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

Kullback, S. (1959). *Information Theory and Statistics.* New York: Wiley.

Kümmerer, M., Theis, L., & Bethge, M. (2014a). Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet. *ArXiv Preprint ArXiv*, *1411*(1610), 01563. Retrieved from http://arxiv.org/abs/1411.1045

Kümmerer, M., Theis, L., & Bethge, M. (2014b). Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet. *2015 Conference on Learning Representations - Workshop Track (ICLR).* Retrieved from https://arvix.org/abs/1411.1045

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

Kümmerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding Low- and High-Level Contributions to Fixation Prediction. In *2017 IEEE International Conference on Computer Vision (ICCV)* (Vol. 2, pp. 4799–4808). IEEE. https://doi.org/10.1109/ICCV.2017.513

Ladd, G. W. (1999). Peer Relationships and Social Competence During Early and Middle Childhood. *Annual Review of Psychology*, *50*(1), 333–359. https://doi.org/10.1146/annurev.psych.50.1.333

Laidlaw, K. E. W., Badiudeen, T. A., Zhu, M. J. H., & Kingstone, A. (2015). A fresh look at saccadic trajectories and task irrelevant stimuli: Social relevance matters. *Vision Research*, *111*(Part A), 82–90. https://doi.org/10.1016/j.visres.2015.03.024

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Social attention to a live person is critically different than looking at a videotaped person. *Proc. Natl. Acad. Sci., USA*(108), 5548–5553.

Laidlaw, K. E. W., Risko, E. F., & Kingstone, A. (2012). A new look at social attention: Orienting to the eyes is not (entirely) under volitional control. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1132–1143. https://doi.org/10.1037/a0027075

Laidlaw, K. E. W., 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

Lancry-Dayan, O. C., Nahari, T., Ben-Shakhar, G., & Pertzov, Y. (2018). Do You Know Him? Gaze Dynamics Toward Familiar Faces on a Concealed Information Test. *Journal of Applied Research in Memory and Cognition*, *7*(2), 291–302. https://doi.org/10.1016/j.jarmac.2018.01.011

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical report. In *Technical Report* (Vol. A-8). Gainesville, FL: University of Florida.

Langton, S. R. H., & Bruce, V. (1999). Reflexive Visual Orienting in Response to the Social Attention of Others Attention of Others. *Visual Cognition*, *6*(5), 541–567. https://doi.org/10.1167/7.10.15

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

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

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

LeDoux, J. E. (1998). *The emotional brain: The mysterious underpinnings of emotional life.* (S. Schuster, Ed.).

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience*, *23*(1), 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience*, *5*(9), 910–6. https://doi.org/10.1038/nn909

Liu, L., Ioannides, A., & Streit, M. (1999). Single Trial Analysis of Neurophysiological Correlates of the Recognition of Complex Objects and Facial Expressions of Emotion. *Brain Tomography*, *11*(4), 291–303.

Lo Gerfo, E., De Angelis, J., Vergallito, A., Bossi, F., Romero Lauro, L. J., & Ricciardelli, P. (2018). Can Monetary Reward Modulate Social Attention? *Frontiers in Psychology*, *9*(November), 1–12. https://doi.org/10.3389/fpsyg.2018.02213

Lord, C., Risi, S., Lambrecht, L., Cook, E. H. J., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Schedule – Generic: A standard measures of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. https://doi.org/10.1023/A:1005592401947

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

Mack, A., Pappas, Z., Silverman, M., & Gay, R. (2002). What we see: Inattention and the capture of attention by meaning. *Consciousness and Cognition*, *11*(4), 488–506. https://doi.org/10.1016/S1053-8100(02)00028-4

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

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects information details within picutres. *Perception and Psychophysics*, *2*(11), 547–552.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*(1), 15–20. https://doi.org/10.1037/0021-843X.95.1.15

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

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., & Kurths, J. (2002). Recurrence Plot Based Measures of Complexity and its Application to Heart Rate Variability Data. *Physical Review E*, *66*(2), 1–16. https://doi.org/10.1103/PhysRevE.66.026702

Masciocchi, C. M., Mihalas, S., Parkhurst, D. J., & Niebur, E. (2009). Everyone knows what is interesting: Salient locations which should be fixated. *Journal of Vision*, *9*(11), 25–25. https://doi.org/10.1167/9.11.25

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the Face in Context: Cultural Differences in the Perception of Facial Emotion. *Journal of Personality and Social Psychology*, *94*(3), 365–381. https://doi.org/10.1037/0022-3514.94.3.365

Matsumoto, K., Shibata, S., Seiji, S., Mori, C., & Shioe, K. (2010). Factors influencing the processing of visual information from non-verbal communications. *Psychiatry and Clinical Neurosciences*, *64*(3), 299–308. https://doi.org/10.1111/j.1440-1819.2010.02077.x

Mattick, R., & Clarke, C. (1998). Development and Validation of Measure of Social Phobia Scrutiny Fear and Social Interaction Anxiety. *Behavior 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: Evidence for a ventral stream visual salience map. *Neuron*, *40*(6), 1241–1250.

McAdams, D. P., Anyidoho, N. A., Brown, C., Huang, Y. T., Kaplan, B., & Machado, M. A. (2004). Traits and Stories: Links Between Dispositional and Narrative Features of Personality. *Journal of Personality*, *72*(4), 761–784. https://doi.org/10.1111/j.0022-3506.2004.00279.x

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

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

McGugin, R. W., McKeeff, T. J., Tong, F., & Gauthier, I. (2011). Irrelevant objects of expertise compete with faces during visual search. *Attention, Perception, and Psychophysics*, *73*(2), 309–317. https://doi.org/10.3758/s13414-010-0006-5

McMains, S. A., & Kastner, S. (2009). Visual Attention. In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of Neuroscience* (pp. 4296–4302). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_6344

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

Menon, V., Ford, J. M., Lim, K. O., Glover, G. H., & Pfefferbaum, A. (1997). Combined event-related fMRI and EEG evidence for temporal—parietal cortex activation during target detection. *NeuroReport*, *8*(14), 3029–3037. https://doi.org/10.1097/00001756-199709290-00007

Miellet, S., Caldara, R., & Schyns, P. G. (2011). Local Jekyll and Global Hyde. *Psychological Science*, *22*(12), 1518–1526. https://doi.org/10.1177/0956797611424290

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*, *99*(23), 15238–15243. https://doi.org/10.1073/pnas.232395699

Mo, C., He, D., & Fang, F. (2017). Attention priority map of face images in human early visual cortex. *The Journal of Neuroscience*, *38*(1), 1206–17. https://doi.org/10.1523/JNEUROSCI.1206-17.2017

Mogg, K., Bradley, B. P., Dixon, C., Fisher, S., Twelftree, H., & McWilliams, A. (2000). Trait anxiety, defensiveness and selective processing of threat: an investigation using two measures of attentional bias. *Personality and Individual Differences*, *28*(6), 1063–1077. https://doi.org/10.1016/S0191-8869(99)00157-9

Mogg, K., Mathews, A., & Eysenck, M. (1992). Attentional bias to threat in clinical anxiety states. *Cognition & Emotion*, *6*(2), 149–159. https://doi.org/10.1080/02699939208411064

Morris, J. P. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain*, *121*(1), 47–57. https://doi.org/10.1093/brain/121.1.47

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

Müller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and Voluntary Orienting of Visual Attention: Time Course of Activation and Resistance to Interruption. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 315–330. https://doi.org/10.1037/0096-1523.15.2.315

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*(6946), 309.

Nagai, C. (2007). Neural mechanisms of facial recognition. *Brain and Nerve*, *59*(1), 45–51.

Nardo, D., Santangelo, V., & Macaluso, E. (2011). Stimulus-Driven Orienting of Visuo-Spatial Attention in Complex Dynamic Environments. *Neuron*, *69*(5), 1015–1028.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Cognitive Brain Research*, *12*(2), 225–231. https://doi.org/10.1016/S0926-6410(01)00053-2

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 391–399. https://doi.org/10.1037/0278-7393.6.4.391

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

New, J. J., Schultz, R. T., Wolf, J., Niehaus, J. L., Klin, A., German, T. C., & Scholl, B. J. (2010). The scope of social attention deficits in autism: Prioritized orienting to people and animals in static natural scenes. *Neuropsychologia*, *48*(1), 51–59. https://doi.org/10.1016/j.neuropsychologia.2009.08.008

Niu, Y., Todd, R. M., Kyan, M., & Anderson, A. K. (2012). Visual and emotional salience influence eye movements. *ACM Transactions on Applied Perception*, *9*(3), 1–18. https://doi.org/10.1145/2325722.2325726

Nummenmaa, L., & Hietanen, J. K. (2006). Gaze distractors influence saccadic curvature: Evidence for the role of the oculomotor system in gaze-cued orienting. *Vision Research*, *46*(21), 3674–3680. https://doi.org/10.1016/j.visres.2006.06.004

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

Oliva, A., & Torralba, A. (2006). Chapter 2 Building the gist of a scene: the role of global image features in recognition. In *Progress in brain research* (Vol. 155, pp. 23–36). Elsevier B.V. https://doi.org/10.1016/S0079-6123(06)55002-2

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)* (Vol. 1, p. I-253-6). IEEE. https://doi.org/10.1109/ICIP.2003.1246946

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

Onat, S., Açık, A., Schumann, F., & König, P. (2014). The Contributions of Image Content and Behavioral Relevancy to Overt Attention. *PLoS ONE*, *9*(4), e93254. https://doi.org/10.1371/journal.pone.0093254

Ordikhani-Seyedlar, M., Sorensen, H. B. D., Kjaer, T. W., Siebner, H. R., & Puthusserypady, S. (2014). SSVEP-modulation by covert and overt attention: Novel features for BCI in attention neuro-rehabilitation. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology*

Society (pp. 5462–5465). IEEE. https://doi.org/10.1109/EMBC.2014.6944862

Ousdal, O. T., Specht, K., Server, A., Andreassen, O. A., Dolan, R. J., & Jensen, J. (2014). The human amygdala encodes value and space during decision making. *NeuroImage*, *101*, 712–719. https://doi.org/10.1016/j.neuroimage.2014.07.055

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, *45*(1), 75–92. https://doi.org/10.1016/j.neuropsychologia.2006.04.025

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

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*(2), 125–154. https://doi.org/10.1163/15685680360511645

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and Performance. *Annual Review of Psychology*, *52*(1), 629–651. https://doi.org/10.1146/annurev.psych.52.1.629

Pegna, A. J., Khateb, A., Michel, C. M., & Landis, T. (2004). Visual recognition of faces, objects, and words using degraded stimuli: Where and when it occurs. *Human Brain Mapping*, *22*(4), 300–311. https://doi.org/10.1002/hbm.20039

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

Pelphrey, K. A., Sasson, N., Reznick, J., Paul, G., Goldman, B., & Piven, J. (2002). Visual scanning of faces in autism. *Journal Of Autism And Developmental Disorders*, *32*(4), 249–261.

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia*, *41*(2), 156–170. https://doi.org/10.1016/S0028-3932(02)00146-X

Perlman, S. B., Morris, J. P., Vander Wyk, B. C., Green, S. R., Doyle, J. L., & Pelphrey, K. A. (2009). Individual Differences in Personality Predict How People Look at Faces. *PLoS ONE*, *4*(6), e5952. https://doi.org/10.1371/journal.pone.0005952

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

Pessoa, L., & Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Frontiers in Neuroscience*, *4*, 17. https://doi.org/10.3389/fnins.2010.00017

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

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

Pilarczyk, J., & Kuniecki, M. (2014). Emotional content of an image attracts attention more than visually salient features in various signal-to-noise ratio conditions. *Journal of Vision*, *14*(12), 4–4.

https://doi.org/10.1167/14.12.4

Piven, J. (2001). The broad autism phenotype: A complementary strategy for molecular genetic studies of autism. *American Journal of Medical Genetics*, *105*(1), 34–35. https://doi.org/10.1002/1096-8628(20010108)105:1<34::AID-AJMG1052>3.0.CO;2-D

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

Posner, M. I., Cohen, Y., & Rafal, R. D. (1982). Neural Systems Control of Spatial Orienting. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *298*(1089), 187–198. https://doi.org/10.1098/rstb.1982.0081

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. *Modes of Perceiving and Processing Information*, *137*(158), 2.

Posner, M. I., & Petersen, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience*, *13*(1), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rammstedt, B., Kemper, C. J., Beierlein, C., & Kovaleva, A. (2012). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10). *GESIS Working Paper*, *23*(2), 1–32. https://doi.org/10.12758/mda.2013.013

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). *Eye movements and attention in reading, scene perception, and visual search*. *Quarterly Journal of Experimental Psychology* (Vol. 62). https://doi.org/10.1080/17470210902816461

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, *8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Riby, D., & Hancock, P. J. B. (2009). Looking at movies and cartoons: Eye-tracking evidence from Williams syndrome and autism. *Journal of Intellectual Disability Research*, *53*(2), 169–181. https://doi.org/10.1111/j.1365-2788.2008.01142.x

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking : Exploring the tendency to imitate another individual ' s gaze. *NeuroReport*, *13*(17), 2259–2264. https://doi.org/10.1097/01.wnr.0000044227

Ricciardelli, P., Carcagno, S., Vallar, G., & Bricolo, E. (2013). Is gaze following purely reflexive or goal-directed instead? Revisiting the automaticity of orienting attention by gaze cues. *Experimental Brain*

Research, *224*(1), 93–106. https://doi.org/10.1007/s00221-012-3291-5

Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition*, *122*(1), 86–90. https://doi.org/10.1016/j.cognition.2011.08.014

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

Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6*(May), 1–11. https://doi.org/10.3389/fnhum.2012.00143

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

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

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

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

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*(6), 5. https://doi.org/10.1167/3.6.5

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

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

Ryan, J. D., Hannula, D. E., & Cohen, N. J. (2007). The obligatory effects of memory on eye movements. *Memory*, *15*(5), 508–525. https://doi.org/10.1080/09658210701391022

Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics*, *25*(2), 119–125. https://doi.org/10.3758/BF03198797

Sali, A. W., Anderson, B. A., & Yantis, S. (2014). The role of reward prediction in the control of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1654–1664. https://doi.org/10.1037/a0037267

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

Sasson, N., Tsuchiya, N., Hurley, R., Couture, S. M., Penn, D. L., Adolphs, R., & Piven, J. (2007). Orienting to social stimuli differentiates social cognitive impairment in autism and schizophrenia.

Neuropsychologia, *45*(11), 2580–2588. https://doi.org/10.1016/j.neuropsychologia.2007.03.009

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in “theory of mind.” *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. https://doi.org/10.1038/253600a0

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

Schmukle, S. C. (2005). Unreliability of the Dot Probe Task. *European Journal of Personality*, *19*, 595–605. https://doi.org/10.1002/per.554

Schneider, W. X., & Deubel, H. (1995). Visual Attention and Saccadic Eye Movements: Evidence for Obligatory and Selective Spatial Coupling. In *Studies in visual information processing* (6th ed., pp. 317–324). North-Holland. https://doi.org/10.1016/S0926-907X(05)80027-3

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

Serences, J. T., & Yantis, S. (2007). Spatially Selective Representations of Voluntary and Stimulus-Driven Attentional Priority in Human Occipital, Parietal, and Frontal Cortex. *Cerebral Cortex*, *17*(2), 284–293. https://doi.org/10.1093/cercor/bhj146

Sharma, P. (2015). Evaluating Visual Saliency Algorithms : Past , Present and Future. *Journal of Imaging Science and Technology*, *59*(5), 1–17. https://doi.org/10.2352/J.ImagingSci.Technol.2015.59.5.050501

Shioiri, S., & Ikeda, M. (1989). Useful Resolution for Picture Perception as a Function of Eccentricity. *Perception*, *18*(3), 347–361. https://doi.org/10.1068/p180347

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv*, (1409), 1556.

Simpson, E. A., Buchin, Z., Werner, K., Worrell, R., & Jakobsen, K. V. (2014). Finding faces among faces: human faces are located more quickly and accurately than other primate and mammal faces. *Attention, Perception, and Psychophysics*, *76*(8), 2175–2183. https://doi.org/10.3758/s13414-014-0744-x

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

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

Sperber, S., & Spinnler, H. (2003). Covert person recognition: Its fadeout in a case of temporal lobe degeneration. *Cortex*, *39*(1), 57–67. https://doi.org/10.1016/S0010-9452(08)70074-1

Spezio, M. L., Huang, P.-Y. S., Castelli, F., & Adolphs, R. (2007). Amygdala Damage Impairs Eye Contact During Conversations with Real People. *Journal of Neuroscience*, *27*(15), 3994–3997.

https://doi.org/10.1523/JNEUROSCI.3789-06.2007

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety scale. Consulting Psychologists*.

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

Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science Quarterly*, *51*(3), 339–350.

Stein, M. B., Simmons, A. N., Feinstein, J. S., & Paulus, M. P. (2007). Increased Amygdala and Insula Activation During Emotion Processing in Anxiety-Prone Subjects. *American Journal of Psychiatry*, *164*(2), 318–327. https://doi.org/10.1176/ajp.2007.164.2.318

Stein, M., Goldin, P., Sareen, J., Zorrilla, L., & Brown, G. (2017). Increased Amygdala Activation to Angry and Contemptuous Faces in Generalized Social Phobia. *Archives of General Psychiatry*, *59*(11), 1027–1034.

Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2014). Rapid Fear Detection Relies on High Spatial Frequencies. *Psychological Science*, *25*(2), 566–574. https://doi.org/10.1177/0956797613512509

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Tamietto, M., & Gelder, B. De. (2010). Neural bases of the non-conscious perception of emotional signals, *11*. https://doi.org/10.1038/nrn2889

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

Tanaka, J., & Gauthier, I. (1997). Expertise in Object and Face Recognition. In *Psychology of Learning and Motivation* (pp. 83–125). https://doi.org/10.1016/S0079-7421(08)60282-0

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

Tanaka, J. W., & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225–245. https://doi.org/10.1080/14640749308401045

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

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. https://doi.org/10.1167/7.14.4

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

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

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6–7), 1029–1054. https://doi.org/10.1080/13506280902764539

Taylor, J. G., & Fragopanagos, N. F. (2005). The interaction of attention and emotion. *Neural Networks*, *18*(4), 353–369. https://doi.org/10.1016/j.neunet.2005.03.005

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

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

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520. https://doi.org/10.1038/381520a0

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

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

Treder, M. S., & Blankertz, B. (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and Brain Functions*, *6*(1), 28. https://doi.org/10.1186/1744-9081-6-28

Treisman, A. M., & Gelade, G. (1980). A Feature-Integration of Attention. *Cognitive Psychology*, *12*(1), 97–136.

Ullman, S. (1996). *High-Level Vision: Object Recognition and Visual Cognition.* Cambridge, MA: MIT Press.

Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, *17*(6–7), 812–834. https://doi.org/10.1080/13506280902771278

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*(4), 471–491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

van Diepen, P., & D'Ydewalle, G. (2003a). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition*, *10*(1), 79–100. https://doi.org/10.1080/713756668

van Diepen, P., & D'Ydewalle, G. (2003b). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition*, *10*(1), 79–100. https://doi.org/10.1080/713756668

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

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Vol. 1, p. I-511-I-518). IEEE Comput. Soc. https://doi.org/10.1109/CVPR.2001.990517

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

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*, *30*(3), 829–841.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2002). Effects of Attention and Emotion on Face Processing in the Human Brain: An Event-Related fMRI Study. *Neuron*, *30*(3), 829–841.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*(6), 624–631. https://doi.org/10.1038/nn1057

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*(11), 1271–1278. https://doi.org/10.1038/nn1341

Vuilleumier, P., & Sagiv, N. (2001). Two eyes make a pair: facial organization and perceptual learning reduce visual extinction. *Neuropsychologia*, *39*(11), 1144–1149. https://doi.org/10.1016/S0028-3932(01)00048-3

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

Webber, C. L., & Zbilut, J. P. (2005). Recurrence Quantification Analysis of Nonlinear Dynamical Systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, 26–94. Retrieved from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.pdf

Wessa, M., Kanske, P., Neumeister, P., Bode, K., Heissler, J., & Schönfeld, S. (2010). EmoPics: subjective and psychophysiological evaluation of new imagery for clinical biopsychological research. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *1*(11), 77.

Whalen, P. J. (2007). The uncertainty of it all. *Trends in Cognitive Sciences*, *11*(12), 499–500. https://doi.org/10.1016/j.tics.2007.08.016

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

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion*, *1*(1), 70–83. https://doi.org/10.1037/1528-3542.1.1.70

Wilming, N., Betz, T., Kietzmann, T. C., & König, P. (2011). Measures and Limits of Models of Fixation Selection. *PLoS ONE*, *6*(9), e24038. https://doi.org/10.1371/journal.pone.0024038

Wilming, N., Onat, S., Ossandón, J. P., Açık, A., Kietzmann, T. C., Kaspar, K., … König, P. (2017). An extensive dataset of eye movements during viewing of complex images. *Scientific Data*, *4*(4), 160126. https://doi.org/10.1038/sdata.2016.126

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology*, *8*(9), R303–R304. https://doi.org/10.1016/S0960-9822(98)70192-7

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

Wundt, W. M. Das Institut für experimentelle Psychologie. (1909). Leipzig.

Wundt, W. M. (1912). *An introduction to psychology*. Macmillan.

Wurtz, R., & Kandel, E. (2000). Central Visual Pathways. In *Principles of neural science* (pp. 523–545).

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

Yarbus, A. L. (1967). Eye movements and vision. In *Plenum press*. New York. https://doi.org/10.1007/978-1-4899-5379-7_8

Yik, M. S. M., & Russell, J. A. (2001). Predicting the Big Two of Affect from the Big Five of Personality. *Journal of Research in Personality*, *35*(3), 247–277. https://doi.org/10.1006/jrpe.2001.2322

Zuckerman, M., Miserandino, M., & Bernieri, F. (1983). Civil inattention exists—in elevators. *Personality and Social Psychology Bulletin*, *9*(4), 578–586.

CURRICULUM VITAE

“Dissertation Based on Several Published Manuscripts“

STATEMENT OF INDIVIDUAL AUTHOR CONTRIBUTIONS AND OF LEGAL SECOND PUBLICATION RIGHTS

Publication (complete reference): Flechsenhar, A. F., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PloS one*, *12*(8), e0183799. http://dx.doi.org/10.1371/journal.pone.0183799.

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	M.G.	M.G.			
Methods Development	A.F.				
Data Collection	A.F.	M.G.			
Data Analysis and Interpretation	A.F.	M.G.			
Manuscript Writing					
Writing of Introduction	A.F.	M.G.			
Writing of Materials & Methods	A.F.	M.G.			
Writing of Discussion	A.F.	M.G.			
Writing of First Draft	A.F.	M.G.			
	A.F.	M.G.			

A. Flechsenhar: conceptualization, data curation, formal analysis, visualization, writing original draft as well as writing review and editing.

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

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

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	M.G.				
Methods Development	A.F. & L.R.	M.G.			
Data Collection	A.F. & L.R.	M.G.			
Data Analysis and Interpretation	A.F. & L.R.	M.G.			

Manuscript Writing					
Writing of Introduction	A.F. & L.R.	M.G.			
Writing of Materials & Methods	A.F. & L.R.	M.G.			
Writing of Discussion	A.F. & L.R.	M.G.			
Writing of First Draft	A.F. & L.R.	M.G.			
	A.F. & L.R.	M.G.			

A. Flechsenhar and L. Rösler share authorship of this publication.

A. Flechsenhar & L. Rösler: conceptualization, data curation, formal analysis, visualization, writing original draft as well as writing review and editing.

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

Publication (complete reference): Flechsenhar, A.F., Larson, O., End, A. & Gamer, M. (2018). Investigating Overt and Covert Shifts of Attention within Social Naturalistic Scenes. *Journal of Vision*, 18(12); 11. http://dx.doi.org/10.1167/18.12.11.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	M.G.	A.E.			
Methods Development	A.F.	M.G.			
Data Collection	O.L.	A.F.	M.G.		
Data Analysis and Interpretation	A.F. & O.L.	M.G.	A.E.		
Manuscript Writing					
Writing of Introduction	A.F.	O.L.	A.E.	M.G.	
Writing of Materials & Methods	A.F.	O.L.	A.E.	M.G.	
Writing of Discussion	A.F.	O.L.	A.E.	M.G.	
Writing of First Draft	A.F.	O.L.	A.E.	M.G.	
	A.F.	O.L.	A.E.	M.G.	

A. Flechsenhar: conceptualization, formal analysis, visualization, writing original draft as well as writing review and editing.

O. Larson: data curation, partial writing of original draft, review and editing.

A. End: conceptualization, review and editing.

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

The doctoral researcher confirms that she/he has obtained permission from both the publishers and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.

Aleya F. Flechsenhar		Würzburg	
Doctoral Researcher’s Name	Date	Place	Signature

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

“Dissertation Based on Several Published Manuscripts“

STATEMENT OF INDIVIDUAL AUTHOR CONTRIBUTIONS TO FIGURES/TABLES/CHAPTERS INCLUDED IN THE MANUSCRIPTS

Publication (complete reference):
Flechsenhar, A. F., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PloS one*, *12*(8), e0183799. http://dx.doi.org/10.1371/journal.pone.0183799

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	A.F.	M.G.			
2	A.F.	M.G.			
3	A.F.	M.G.			
4	A.F.	M.G			
5	A.F.	M.G.			
6	A.F.	M.G.			

A. Flechsenhar: conceptualization, data curation, formal analysis, visualization, writing original draft as well as writing review and editing.

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

Publication (complete reference):

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

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	A.F. & L.R.	M.G.			
2	A.F. & L.R.	M.G.			
3	A.F. & L.R.	M.G.			
4	A.F. & L.R.	M.G.			
5	A.F. & L.R.	M.G.			

A. Flechsenhar and L. Rösler share authorship of this publication.

A. Flechsenhar & L. Rösler: conceptualization, data curation, formal analysis, visualization, writing original draft as well as writing review and editing.

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

Publication (complete reference):

Flechsenhar, A.F., Larson, O., End, A. & Gamer, M. (2018). Investigating Overt and Covert Shifts of Attention within Social Naturalistic Scenes. *Journal of Vision*, 18(12); 11. http://dx.doi.org/10.1167/18.12.11

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	M.G.	A.F.	O.L. & A.E.		
2	A.F.	M.G.	O.L. & A.E.		
3	M.G.	A.F.	O.L. & A.E.		
4	A.F.	M.G.	O.L. & A.E.		
5	A.F.	M.G.	O.L. & A.E.		

A. Flechsenhar: conceptualization, formal analysis, visualization, writing original draft as well as writing review and editing.

O. Larson: data curation, partial writing of original draft, review and editing.

A. End: conceptualization, review and editing.

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

I also confirm my primary supervisor's acceptance.

Aleya F. Flechsenhar Würzburg

Doctoral Researcher's Name Date Place Signature

AFFIDAVIT

I hereby confirm that my thesis entitled "The Ubiquity of Social Attention – a Detailed Investigation of the Underlying Mechanisms" 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 "Die Allgegenwärtigkeit Sozialer Aufmerksamkeit – eine Detaillierte Erforschung zugrundeliegender Mechanismen" eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum

Unterschrift