Virtually Valid?

On the Importance of Ecological Validity and Virtual Reality for Social Attention Research

Praktischerweise Valide?

Über die Bedeutung von ökologischer Validität und virtueller Realität in der sozialen Aufmerksamkeitsforschung

Doctoral thesis for a doctoral degree
At the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Neuroscience

Submitted by

Jonas David Grossekathöfer

Born in

Lippstadt

Würzburg, 2021

Submitted on:

Members of the Thesis Committee:

Chairperson: Prof. Dr. Keram Pfeiffer, University of Würzburg
Primary Supervisor: Prof. Dr. Matthias Gamer, University of Würzburg
Supervisor (Second): Prof. Dr. Erhard Wischmeyer, University of Bielefeld
Supervisor (Third): Prof. Dr. Lynn Huestegge, University of Würzburg

Date of Public Defence:

Date of Receipt of Certificates:

Acknowledgement

Die vorliegende Dissertation umfasst die Arbeit von mehr als 5 Jahren. In dieser Zeit haben mich viele Personen in meinem Umfeld unterstützt. Diesen gilt an dieser Stelle mein besonderer Dank.

Zuallererst möchte ich mich bei Prof. Dr. Matthias Gamer, meinem Doktorvater, bedanken. Er hat es mir ermöglicht, diese spannende Forschung an der Schnittstelle von sozial-kognitiver Grundlagenforschung und modernster Technik zu verfolgen. Dabei habe ich sehr vom fachlichen und inhaltlichen Austausch sowie seiner Expertise in sozialer Aufmerksamkeit, der virtuellen Realität und dem Eyetracking profitiert. Seine Betreuung hat sich durch anhaltende Unterstützung und eine stets offene Tür (auch im übertragenen Sinne) ausgezeichnet. Ich bin dankbar, dass mir dadurch das akademische Arbeiten auf eine Art nähergebracht wurde, sodass ich nicht nur motiviert bin, sondern mich auch gut gewappnet fühle, eine akademische Karriere anzugehen.

Zudem möchte ich mich auch bei meinem Promotionskomitee für die stets konstruktiven Rückmeldungen und die wertvollen Impulse für meine wissenschaftlichen Arbeiten bedanken.

Von allen Kolleginnen und Kollegen, die mich in Würzburg begleitet haben, möchte ich mich in erster Linie bei Dr. Lea Hildebrandt bedanken. Sie hat als Postdoc zeitgleich mit mir in Würzburg begonnen und hat mir die Welt der modernen Aspekte wissenschaftlicher Arbeit und der neusten quantitativen Methoden eröffnet, wovon meine wissenschaftlichen Arbeiten maßgeblich profitiert haben. Darüber hinaus schätze ich ihren Rat zu allerlei wissenschaftlichen und akademischen Themen.

Des Weiteren gilt mein Dank den aktuellen und bereits weitergezogenen Kolleginnen und Kollegen der Arbeitsgruppe sowie der Kaffeerunde (Prädikat Qualita Rosso) für das freundschaftliche und wohlwollende Miteinander und den hilfsbereiten und bereichernden Austausch. Es war eine gute Zeit.

Die nötige Kraft und Ausdauer für die Doktorarbeit kommen von meiner Familie. Im Endspurt der letzten Wochen hat sich wieder gezeigt, wie sehr ich mich auf meine Frau Laura verlassen kann. Dass sie mich bestärkt, unterstützt, aufbaut und das schon seit vielen Jahren. Vielen Dank dafür, Laura! Und auch die jüngste in unserer kleinen Familie hat einen hervorragenden Beitrag für die vorliegende Arbeit geleistet. Meine Tochter Carlotta gab mir

ihr erstes Lebensjahr über eine dicke Portion Motivation.

Zudem gilt mein Dank meinen Eltern, die mich seit jeher in meinen Vorhaben unterstützen und an mich glauben. Erst dadurch habe ich als erstes Familienmitglied einen Fuß in die akademische Welt gesetzt, dessen jüngste, aber sicherlich nicht letzte Spur, diese Doktorarbeit ist.

Summary

Gazes are of central relevance for people. They are crucial for navigating the world and communicating with others. Nevertheless, research in recent years shows that many findings from experimental research on gaze behavior cannot be transferred from the laboratory to everyday behavior. For example, the frequency with which conspecifics are looked at is considerably higher in experimental contexts than what can be observed in daily behavior. In short: findings from laboratories cannot be generalized into general statements. This thesis is dedicated to this matter.

The dissertation describes and documents the current state of research on social attention through a literature review, including a meta-analysis on the *gaze cueing* paradigm and an empirical study on the robustness of gaze following behavior. In addition, virtual reality was used in one of the first studies in this research field. Virtual reality has the potential to significantly improve the transferability of experimental laboratory studies to everyday behavior. This is because the technology enables a high degree of experimental control in naturalistic research designs. As such, it has the potential to transform empirical research in the same way that the introduction of computers to psychological research did some 50 years ago.

The general literature review on social attention is extended to the classic *gaze cueing* paradigm through a systematic review of publications and a meta-analytic evaluation (Study 1). The cumulative evidence supported the findings of primary studies: Covert spatial attention is directed by faces. However, the experimental factors included do not explain the surprisingly large variance in the published results. Thus, there seem to be further, not well-understood variables influencing these social processes.

Moreover, classic *gaze cueing* studies have limited ecological validity. This is discussed as a central reason for the lack of generalisability. Ecological validity describes the correspondence between experimental factors and realistic situations. A stimulus or an experimental design can have high and low ecological validity on different dimensions and have different influences on behavior. Empirical research on gaze following behavior showed that the *gaze cueing* effect also occurs with contextually embedded stimuli (Study 2). The contextual integration of the directional cue contrasted classical *gaze cueing* studies, which usually show heads in isolation. The research results can thus be transferred *within* laboratory studies to higher ecologically valid research paradigms.

However, research shows that the lack of ecological validity in experimental designs significantly limits the transferability of experimental findings to complex situations *outside* the laboratory. This seems to be particularly the case when social interactions and norms are investigated. However, ecological validity is also often limited in these studies for other factors, such as contextual embedding *of participants*, free exploration behavior (and, thus, attentional control), or multimodality. In a first study, such high ecological validity was achieved for these factors with virtual reality, which could not be achieved in the laboratory so far (Study 3). Notably, the observed fixation patterns showed differences even under *most similar* conditions in the laboratory and natural environments. Interestingly, these were similar to findings also derived from comparisons of eye movement in the laboratory and field investigations. These findings, which previously came from hardly comparable groups, were thus confirmed by the present Study 3 (which did not have this limitation).

Overall, *virtual reality* is a new technical approach to contemporary social attention research that pushes the boundaries of previous experimental research. The traditional trade-off between ecological validity and experimental control thus becomes obsolete, and laboratory studies can closely inherit an excellent approximation of reality. Finally, the present work describes and discusses the possibilities of this technology and its practical implementation. Within this context, the extent to which this development can still guarantee a constructive classification of different laboratory tests in the future is examined.

Zusammenfassung

Blicke sind für Menschen von zentraler Relevanz. Sie sind entscheidend für die Navigation in der Welt und für die Kommunikation mit Mitmenschen. Dennoch zeigt die Forschung der letzten Jahre, dass sich Erkenntnisse aus der experimentellen Forschung zu Blickverhalten vom Labor nicht in alltägliches Verhalten übertragen lassen. So ist beispielsweise die Häufigkeit, mit der Mitmenschen angeschaut werden, erheblich höher in experimentellen Kontexten als das beobachtbare alltägliche Verhalten. Kurz: Erkenntnisse aus Laboren lassen sich nicht zu allgemeinen Aussagen generalisieren. Diesem Spannungsfeld ist die hier vorliegende Arbeit gewidmet.

Diese Doktorarbeit beschreibt und dokumentiert den aktuellen Forschungsstand zur sozialen Aufmerksamkeit anhand einer Literaturübersicht inklusive einer Metaanalyse zum *gaze cueing* Paradigma sowie einer empirischen Untersuchung zur Robustheit des Blickfolgeverhaltens. Zudem wird in einer der ersten Studien in diesem Forschungsfeld virtuelle Realität eingesetzt. Virtuelle Realität hat das Potenzial, die Übertragbarkeit zwischen experimentellen Laboruntersuchungen auf alltägliches Verhalten deutlich zu verbessern. Denn die Technologie ermöglicht eine hohe experimentelle Kontrolle in naturalistischen Forschungdesigns. Damit kann sie die empirische Forschung ebenso stark verändern wie die Einführung des Computers für psychologische Forschung vor rund 50 Jahren.

Die Literaturübersicht über soziale Aufmerksamkeit wurde durch eine systematische Begutachtung der Publikationen und einer meta-analystische Auswertung zum klassischen *gaze cueing* Paradigma erweitert (Studie 1). Die kumulierte Evidenz unterstützt die Befunde primärer Studien: Verdeckte räumliche Aufmerksamkeit wird durch Gesichter gelenkt. Allerdings zeigte sich eine überraschend große Varianz in den publizierten Ergebnissen, die durch die untersuchten experimentellen Faktoren nicht erklärt werden konnte. Es scheint also noch Weitere, nicht gut verstandene Einflussgrößen auf diesen sozial-kognitiven Prozess zugeben.

Klassische *gaze cueing* Studien besitzen zudem eine eingeschränkte ökologische Validität. Diese wird als ein zentraler Grund für die fehlende Generalisierbarkeit diskutiert. Ökologische Validität beschreibt die Übereinstimmung von experimentellen Faktoren mit realistischen Situationen. Ein Stimulus oder ein experimentelles Design kann auf verschiedenen Dimensionen hohe und niedrige ökologische Validität aufweisen. Dies kann auf verschiedene

Verhaltensbereiche unterschiedlichen Einfluss haben. Die empirischen Untersuchungen zum Blickfolgeverhalten zeigten, dass der *gaze cueing* Effekt auch bei kontextuell eingebundenen Stimuli auftritt (Studie 2). Die kontextuelle Einbindung des Richtungscues stellte dabei einen Kontrast zu klassichen *gaze cueing* Studien dar, die in der Regel Köpfe in Isolation zeigen. Die Forschungsergebnisse lassen sich also *innerhalb* von kontrollierten Laboruntersuchungen auch auf ökologisch validere Forschungsparadigmen übertragen.

Forschungsarbeiten zeigen allerdings, dass die mangelnde ökologische Validität in experimentellen Designs die Übertragbarkeit von experimentellen Befunden auf komplexe Situationen *außerhalb* des Labors erheblich einschränken. Dies scheint insbesondere der Fall zu sein, wenn soziale Interaktionen und Normen untersucht werden. Die ökologische Validität ist in diesen Studien aber auch für weitere Faktoren häufig eingeschränkt, wie beispielsweise die kontextuelle Einbettung *von Versuchspersonen*, freies Explorationsverhalten (und damit Aufmerksamkeitssteuerung) oder Multimodalität. In einer ersten Arbeit in diesem Forschungsfeld wurde für diese Faktoren mit virtueller Realität eine hohe ökologische Validität erreicht, die so bisher nicht im Labor zu erreichen war (Studie 3). Spannenderweise zeigten sich selbst unter *ähnlichsten* Bedingungen im Labor und in der natürlichen Umgebung Unterschiede in den beobachteten Fixationsmustern. Interessanterweise sind diese ähnlich zu Befunden, die ebenfalls aus Vergleichen von Augenbewegung im Labor und Felduntersuchung stammten. Diese Befunde, die bisher auf wenig vergleichbare Untersuchungsgruppen beruhen, wurden durch die vorliegende Studie 3 (die diese Einschränkung nicht besitzt) bestätigt.

Insgesamt steht der heutigen sozialen Aufmerksamkeitsforschung mit *virtueller Realität* ein neuer technischer Ansatz zur Verfügung, der die Grenzen bisheriger experimenteller Forschung verschiebt. Die traditionelle Abwägung zwischen ökologische Validität und experimenteller Kontrolle wird damit hinfällig und in Laboruntersuchungen kann die Realität extrem nah nachgebildet werden. Abschließend werden in dieser Arbeit die Möglichkeiten und die praktische Umsetzung dieser Technologie beschrieben und diskutiert. Dabei wird auch kritisch beleuchtet, inwiefern mit dieser Entwicklung auch in Zukunft noch eine konstruktive Einordnung von verschiedenen Laboruntersuchungen gewährleistet werden kann.

Contents

Preface **i**

Acknowledgement i

Summary iii

Zusammenfassung v

List of Figures xiii

List of Tables xvi

I Theoretical background **1**

1 Social attention and gaze cueing **5**

1.1 Basic concepts of attention 5

1.2 Covert and overt attention 6

1.3 Social attention and cognition 8

1.4 Social attention and gaze cueing 11

2 Social attention and ecological validity **21**

2.1 Basic concepts of validity 21

2.2 Ecological validity in gaze cueing paradigms 23

2.3 Ecological validity and VR 27

3 The present work **31**

II Empirical investigation 33

4 Meta-analysis of the gaze cueing effect 35
  4.1 Brief summary 35
  4.2 Methods 36
    4.2.1 Literature search and coding procedure 36
    4.2.2 Moderator variables 38
    4.2.3 Statistical analysis 39
  4.3 Results 41
    4.3.1 Publication bias 42
    4.3.2 Analysis of the gaze cueing effect 42
    4.3.3 Moderator analysis 43
    4.3.4 (pseudo) R^2 46
  4.4 Discussion 48

5 Gaze cueing in naturalistic scenes 53
  5.1 Brief summary 53
  5.2 Methods 54
    5.2.1 Participants 54
    5.2.2 Stimuli and apparatus 55
    5.2.3 Design and procedure 55
    5.2.4 Data analysis 57
  5.3 Results 59
    5.3.1 Gaze following 59
    5.3.2 Memory for objects 61
    5.3.3 Social prioritization 62
  5.4 Discussion 65
  5.5 Acknowledgements 69

6 Reality in a sphere 71
  6.1 Brief summary 71
  6.2 Methods 72

6.2.1 Participants 72
6.2.2 Stimuli and apparatus 72
6.2.3 Procedure 74
6.2.4 Questionnaires 77
6.2.5 Image data processing 77
6.2.6 Data analysis 78
6.3 Results 80
6.3.1 Comparison between real and virtual environment 80
6.3.2 Consistency of viewing behavior within and across environments . . . 81
6.3.3 Influence of personality 84
6.3.4 Relevance of the number of persons present in a scene 87
6.4 Discussion 89
6.5 Acknowledgement 95

III General discussion **97**

7 Rethinking ecological validity **101**

8 Practical considerations for VR **105**

9 The study of gaze cueing in VR **109**

10 Outlook and concluding remarks **115**

References **119**

Appendix **143**

A Supplement: Gaze cueing in scenes **143**
A.1 Regions of Interest (ROIs) 143
A.2 Follow-up ANOVAs on social attention 147
A.3 ANOVA including all ROIs 150

B Supplement: Reality in a sphere 155

B.1 Assumption check & robust LMM 155

B.2 Convergence issues in Model 4 157

C Publications list 159

D Curriculum vitæ 161

E Affidavit / Eidesstattliche Erklärung 165

List of Figures

1.1 Histogram of Records in the Web of Science of the term *social cognition.* . . 14

1.2 Schematic sequence of two classic gaze cueing trials. (A) A valid trial with a typical target used for detection or localisation tasks, depending on the instructions given to the participants. (B) An invalid trial of an identification task, e.g., typically between two or more letters or symbols. Depicted faces are modified line drawings by Barry Langdon-Lassagne, distributed under a CC-BY 3.0 license. 15

4.1 PRISMA-style flowchart showing the study selection for meta-analysis on the gaze cueing effect 36

4.2 Funnel plot of the relationship between expected standard error and observed RMD. Note that visual interpretation of the funnel plot is impeded by clusters of points that indicate the multi-level structure of the data. However, such patterns also serve as indications of publication biases. 42

4.3 Interaction plot for the meta-regression model showing estimated marginal means as a function of stimulus onset asynchrony and cue-target overlap. Error regions around regression slopes depict the 95% confidence intervals at the predicted values. 47

5.1 Example photographs of a single scene. Gaze direction and objects were balanced over participants. In total 104 photographs of 26 scenes were used. Please note that since we did not obtain permission for publishing the original stimuli, this image shows an example that was not used in the experiment but taken post-hoc in order to illustrate the generation of the stimulus set. . . . 56

5.2 Bar plots of the different prioritization measures for the attentional orienting towards the cued and uncued objects as a function of instruction group. Note that data were aggregated across time bins for fixation duration and numbers. Error bars represent standard errors of the mean. 59

5.3 Time course of fixation durations and numbers as a function of the region of interest (ROI). Error bars represent standard errors of the mean. 61

5.4 Bar plot of the memory performance for the cued and uncued objects as a function for instruction group. Error bars represent standard errors of the mean. 62

5.5 Bar plot of the different prioritization measures for attentional orienting towards and visual exploration of the depicted person's head and body as a function of instruction group. Note that data were aggregated across time bins for fixation duration and numbers. Error bars represent standard errors of the mean. 64

6.1 Google Maps route the experimenter and participants walked to the five measurement locations (highlighted with yellow dots and corresponding descriptive labels) in the real environment. Spherical videos were recorded at the identical locations for the virtual environment. 73

6.2 Interaction plots of all linear mixed models showing estimated marginal means as a function of the included predictors. Error bars depict 95% confidence intervals of predicted values. (A) The preregistered Model 1 included environment (real environment, RE vs. virtual environment, VE) and ROI (person vs. object) as predictors. Models 2 (B) and 3 (C) additionally included standardized scores of the Social Interaction Anxiety Scale (SIAS, depicted range from 12 (-1 *SD*) to 29 (+1 *SD*)), and the Autism-Spectrum-Quotient short version (AQ-k, depicted range from 3 (-1 *SD*) to 10 (+1 *SD*)), respectively. (D) Compared to Model 1, Model 4 additionally included the standardized number of pedestrians in the environment at each location (depicted range from 3 (-1 *SD*) to 15 (+1 *SD*)). 82

6.3 Heat map of pairwise Pearson correlation coefficients r for gaze on persons (A) and objects (B) for all five locations in both environments. Framed coefficients highlight correlations at identical locations between both environments. . . . 83
6.4 Smoothed density maps illustrating the distribution of gaze coordinate relative to the field of view in the real environment (RE) and the virtual environment (VE). Densities are depicted for a central area of 46° vertically and 60° horizontally (normalized to a range of 0 to 1). 85
6.5 Dotwhisker plot of estimated model coefficients for Model 1 to 4 (dots) with 95% confidence intervals shown as whiskers. Environment (RE): Environment reference *Real environments*, ROI (object): ROI reference *Objects*, SIAS: standardized sum score of the Social Interaction Anxiety Scale, AQ-k: standardized sum score of the Autism-Spectrum-Quotient short version. Pedestrians: standardized number of pedestrians. 88

A.1 Fixation latencies as a function of the region of interest and group. Error bars represent 95% confidence interval. 151
A.2 Fixation duration as a function of the region of interest and group. Error bars represent 95% confidence interval. 152
A.3 Fixation numbers as a function of the region of interest and group. Error bars represent 95% confidence interval. 153

B.1 QQ plot for the residuals of the originally preregistered model. 156

List of Tables

4.1 Estimated coefficients for the meta-regression model for task type. 44
4.2 Estimated coefficients for the meta-regression model for stimulus type. . . . 44
4.3 Estimated coefficients for the meta-regression model for cue type. 45
4.4 Estimated coefficients for the meta-regression model for emotions. 45
4.5 Estimated coefficients for the meta-regression model for stimulus onset asynchrony. 46
4.6 Estimated coefficients for the meta-regression model for stimulus onset asynchrony times cue-target overlap. 47

5.1 Parameters and model selection criteria of general linear mixed models predicting object recall from group, number/duration of fixation and object role. 63

6.1 Estimated coefficients for the preregistered Model 1 with environment and ROI as fixed and participant ID as random effects for the prediction of gaze proportions. 81
6.2 Estimated coefficients for the Model 2 with environment, ROI and SIAS as fixed and participant ID as random effects for the prediction of gaze proportions. 86
6.3 Estimated coefficients for the Model 3 with environment, ROI and AQ-k as fixed and participant ID as random effects for the prediction of gaze proportions. 86
6.4 Model comparison of the preregistered model with the models including personality traits. 87
6.5 Estimated coefficients for the Model 4 with environment, ROI and number of pedestrians as fixed and participant ID and location as random effects for the prediction of gaze proportions. 89

A.2 Sizes of the regions of interest in pixels. 145
A.3 Width and height of regions of interest in pixels. 145
A.4 Width and height of regions of interest in centimeters on the computer screen. 146
A.5 Width and height of regions of interest in degrees of visual angle for the current viewing distance. 146
A.6 Positions (x- and y-coordinates) of the different regions of interest on the image in pixels relative to the top left corner. 147
A.7 Distance between objects (cued & uncued) and the head of the depicted person in pixels, centimeters and degree of visual angle. 147
A.8 time point 1: 0 - 2 seconds 148
A.9 time point 2: 2 - 4 seconds 148
A.10 time point 3: 4 - 6 seconds 149
A.11 time point 4: 6 - 8 seconds 149
A.12 time point 5: 8 - 10 seconds 150
A.13 Fixation latency 151
A.14 Fixation duration 152
A.15 Fixation number 153

B.1 Estimated coefficients for a robust linear mixed model (Koller, 2016), specified exactly like the preregistered model 1 (see main article) with environment and ROI as fixed and participant ID as random effects for the prediction of gaze proportions. 156
B.2 Estimated standard deviations, and correlations between the random-effects. 157

Part I

Theoretical background

> No matter how one may try, one cannot not communicate.
>
> – (Watzlawick, 1967)

Humans inevitably begin to communicate through the mere presence of another individual. The conspecific does not notice us? She looks in a completely different direction? Focuses her smartphone? It doesn't matter. Her behavior transports enough information for some potentially very relevant information: We might be able to read whether she is bored or waiting for someone. We might be able to tell in what kind of mood she is. She communicates by her mere presence, all without a sole word being spoken and no intention whatsoever from her. In fact, she cannot not communicate.

This little interaction is of utter importance to humans. It is the foundation of synchronized behavior such as cooperating and talking. The small introductory passage describes a daily situation where the most vital communication source lies in the face: Where do we look? What do we look at? How do we look? These faces are very rarely, if ever, available in isolation, unlike in psychological research. This dissertation is dedicated to this matter.

The first chapter provides an overview of the vast field of attentional research. First, attention as a competitive selection process with its cognitive components is introduced. The two modes of attention, namely overt and covert attention, are then described before research with social stimuli is discussed. Such research shows that social attention findings often defy many traditional views of attention. Mainly regarding bottom-up and reflexive processing of rather complex stimuli. Afterward, the section describes the first approach to fill the gap between cognitive and social psychology: The gaze cueing paradigm.

In the second chapter, a critical assessment focusing on the ecological validity of the social attention research is provided. First, an introduction to validity as a scientific cornerstone as well as ecological validity is given. It is followed by an overview showing that focusing on ecological validity can improve studies on social attention research, which apparently often does not generalize to natural behavior. The chapter ends with an outlook on virtual reality (VR). Virtual reality is discussed as a potential solution increasing ecological validity without giving up experimental control and as a promising technical solution for these long-standing problems of experimental paradigms used in social attention research.

Finally, the motivation for the present work is outlined.

Chapter 1

Attention, social attention, and gaze cueing

1.1 Basic concepts of attention

Attention is the process by which the information human spend their cognitive resources on gets selected. This filtering is a foundational requirement for the free will and subjective experience in humans because at any waking moment *everything* else could have been attended. This central problem occupied research in experimental psychology from its very origins on (Helmholtz, 1896; James, 1890; Wundt, 1912) and is still under active investigation today (Carrasco, 2011; Knudsen, 2007; Moore and Zirnsak, 2017; Posner and Rothbart, 2007).

Attention is a fundamental of cognitive function and a central module in the complex architecture of the human information processing system. In general, only attended information can be consciously accessible. Current models see attention primarily as a mechanism for gate keeping the information processed by the brain (Knudsen, 2007). In fact, information becomes only conscious when it enters working memory. Holding and manipulating it lies out of the scope of attention, but it selectively regulates the information stream. Without this selection, the working memory capacities are easily overflown by the total amount of information. This includes information from the external world (i.e., via the senses) and from the internal monitoring process (i.e., homeostatic processes, like being hungry).

Additionally, the working memory can only store (and manipulate) information of a single domain at any given point in time and only for a few seconds (Baddeley, 2012). Thus, attention plays a central role in maintaining cognitive processes and switching tasks. Every information that is to be integrated with internal states, evaluated for goals, or manipulated and processed for memories have to have won a competitive attentional selection process. It is the only entry point into working memory. Only then can information be consciously used for planning goal-driven and complex behavior in working memory (Genovesio et al., 2006).

The competitive selection process can be won following two routes: One route to win the selection process is voluntarily guided attention. This attentional control is described as top-down selection (Knudsen, 2007), exogenous (Posner, 1980) or non-automatic control (Jonides, 1981). During top-down selection, attention is guided towards arbitrary objects and locations that align with the organism's goals or motivational states. Attention onto these objects is often, but by no means only, aligned with directional movements to increase the information uptake.

In contrast to goal-driven control of attention, allocation of attention can also be stimulus-driven. In this bottom-up process, any stimuli with appropriate features can enter the working memory, i.e., win the selection process (Itti and Koch, 2001, or see below for a discussion on *appropriate stimulus features*). This kind of attentional control is also known as exogenous (Posner, 1980), reflexive (Müller and Rabbitt, 1989), or automatic (Jonides, 1981).

In sum, attention is formed by four distinct processes, working memory, top-down selection, bottom-up filtering, and competitive selection. Its function is to manage and limit the information stream entering the working memory, thus preventing flooding of consciousness with irrelevant information (Knudsen, 2007).

1.2 Covert and overt attention

Attention can also be described in two separate modes: Covert and overt attention. These modes facilitate the perception of attended or to-be attended locations in different ways (Posner, 1980). Overt attention occurs when the spatial resolution is improved with behavioral adaptation. For example, by aligning body, head and eye position, an object can be moved into foveal focus to be better processed. Covert attention is defined as an attentional shift in the absence of behavioral adaptation (Posner, 1980). For human vision, covert attention plays a critical role in the navigation of eye movements. It is central for processing visual information in the periphery (Kowler, 2011). The planning of saccades and which locations are next to be fixated is the product of a complex interplay between peripheral and foveal vision (Stewart et al., 2020). As such, covert attention often precedes overt attention (Carrasco, 2011; Kowler, 2011). Furthermore, covert attention is assumed to be deployed to more than one location *in parallel* (Carrasco, 2011). This is in contrast to overt attention, which is tightly bound to eye movements and thus only allows serial processing.

Covert attention increases performance at the attended location by increasing the contrast sensitivity of cognitive processes (Carrasco, 2011) and thus does not require the alignment of sensory organs for attentional focus. In general, is the spatial resolution at covertly attended locations in the visual field increased (Yeshurun and Carrasco, 1998,9). This signal improvement could be shown to be feature (Desimone and Duncan, 1995) and timing specific (Khayat, 2006). According to the perceptual template model, covert attention facilitates perception by increasing contrast sensitivity (Carrasco, 2011). According to the model, covert attention operates via three mechanisms that alter the signal-to-noise ratio. Thereby perception is facilitated without an orienting of the sensory organs. In general, facilitation is achieved by adjusting perceptual templates. First, attention can increase the gain of a template and amplify the signal stimulus. Second, by focusing on the perceptual template for the spatial region or stimulus feature, external noise can be filtered and its influence reduced. Third, an internal noise reduction is also capable of improving signal processing. Stimulus size and the attentional spatial spread determine the effect of covert attention on these mechanisms, as shown in cueing tasks (Herrmann et al., 2010). Depending on those properties, attention modulates the neuronal responses in the visual cortex via response gain or contrast gain to facilitate performance at attended regions. An alternative to the perceptual template model is the spatial resolution hypothesis (Carrasco, 2011; Yeshurun and Carrasco, 1998). This hypothesis assumes a signal enhancement at covertly attended locations. According to this hypothesis, spatial resolution in the visual field can be increased to facilitate performance (in the absence of eye movements). From the anatomy of the eye, the spatial resolution in the visual field of humans is highest at the central area of the retina, i.e., the fovea centralis, and is gradually decreasing towards the periphery of the visual field (e.g., Intriligator and Cavanagh, 2001). However, in contrast to altering the signal-to-noise ratio, covert attention can also improve the spatial resolution at a covertly attended region by increasing the sensitivity of relevant neurons at the receptive field. Thus, the overall sensitivity shifts towards higher spatial frequencies at attended locations (Yeshurun and Carrasco, 1998).

The guidance of attention is a daily routine, and often this is achieved by allocating the focus of our gaze. Nonetheless, we use covert attention in as many daily routines. For example, for search tasks, for a task that requires monitoring of the environment (e.g., when driving), and, as mentioned before, it also guides subsequent eye movements to salient or relevant aspects in the visual field. Thus it plays a major role in early visual processing. Additionally, it can also be applicable in social interactions. Overt as well as covert attention can be guided voluntarily. Eye movements quickly reveal what lies in the focus of interest of each other. However, sometimes humans want or need to conceal where they are attending. Covert attention enables humans to attend to objects of interest in the visual scene without allocating their gaze to it. Consequentially, one's intention might be concealed.

One strategy to separate overt from covert attentional processes is to observe behavioral improvements without eye or head movements. One paradigm that achieves this is the Posner cueing paradigm (Posner and Cohen, 1984) and correspondingly the gaze cueing paradigm for social situations (e.g., Driver et al., 1999; Friesen and Kingstone, 1998, see Section 1.3).

1.3 Social attention and cognition

Humans live in an environment with a wealth of social information. In everyday life, countless encounters of people occur, some unpredictable and random. Encounters with people we know, people we fear, people we like, people who can't see us, people who want to tell us something, people on billboards, people on screens … The fast extraction and correct classification of social information is necessary in such environments. Social cognition is the general term that describes the entire cognitive process from detecting, processing, storing to manipulating that is required in social situations. These cognitive mechanisms are thought to build the foundation for making inferences about another one's state of mind (Baron-Cohen, 1991; Emery, 2000). Such social understanding is assumed to be the highest level of cognition (Neuberg and Schaller, 2015; Perrett and Emery, 1994) and a capability which is argued to have developed only in humans (Barresi and Moore, 1996).

The current work focuses on social attention. Social attention describes the rapid and reflexive visual detection of social information (Fletcher-Watson et al., 2008; Rösler et al., 2017). This broad definition covers a wide variety of phenomena. It includes cognitive processes for detecting social information in complex scenes and the definition also covers the cognitive foundations for human social interactions, such as gaze contact, gaze following, and joint attention.

Eye movements are a prosperous source for uncovering cognitive processes in social attention. Researchers investigated the influence of a scenes' social aspects on eye movements for decades. They found that humans spend a considerable amount of fixations on other humans (e.g., Yarbus, 1967). The preferential processing of social aspects in visual scenes was since then supported by several lines of evidence (Birmingham et al., 2008b; Cerf et al., 2009; End and Gamer, 2017; Flechsenhar et al., 2018; Flechsenhar and Gamer, 2017; Kawakami et al., 2014; Rösler et al., 2017; Zwickel and Võ, 2010). Such studies show unanimously very early and dense fixation on humans in general, and particularly on their eye and head region. The strong prioritization of eyes seems to be only partially under voluntary control. For example Laidlaw et al. (2012) measured participants eye movements under several explicit instructions. Participants were asked to watch upright and inverted faces under three conditions: free viewing, restricted (i.e., avoiding to look at the eyes), and control (i.e., avoiding to look at the mouth). Participants avoided the requested region for inverted faces in the *restricted* condition as they were asked to do. For upright faces, however, participants failed to fully restrain their fixations towards the face's eyes. Thus, the authors argue, focus on eyes as social cues cannot be voluntarily suppressed. It is even assumed that the morphology of the human eye might have evolved especially for this task (Kobayashi and Kohshima, 1997, 2001). Specifically, even compared to other primates, the unique white sclera around the dark iris and black pupil in humans makes gaze detection a lot easier. And in fact, Ricciardelli et al. (2000) showed that gaze perception is severely impaired with the reversed polarity of the eye. However, the orienting towards persons also seems to occur reflexively. Rösler et al. (2017) showed that the very first saccades moved towards locations with social information (and this was even more predictive than physical saliency, see discussion below). All in all, these findings highlight a rapid and reflexive prioritization of social features by cognitive processes.

As mentioned above, attention is about winning a selection process from an overwhelming input of information. Further evidence for the unique role of social attention in human cognition comes from computational approaches simulating such selection processes. Before a gaze can be followed, an attentional process must select faces and eyes. The preattentive selection mechanism relies on covert attention and bottom-up processes. This preattentive selection can be successfully predicted by computational models, for example, by utilizing so-called saliency maps (Borji and Itti, 2013; Cerf et al., 2008; Itti and Koch, 2001). In short, the most successful models work with a saliency map of a visual scene composed from feature maps (Borji and Itti, 2013; Itti and Koch, 2000, 2001). The feature maps are computed for physical saliencies of contrasts in color, intensity, or orientation for any given scene (inheriting the central ideas of the feature-integration theory of attention, Treisman and Gelade, 1980). The weights these algorithms provide for a scene closely map real human fixations Borji and Itti (2013); Itti and Koch (2001). These algorithms, however, have a significant limitation. They perform considerably worse as soon as social aspects are introduced into the scene (Birmingham et al., 2009; End and Gamer, 2017; Flechsenhar et al., 2018; Flechsenhar and Gamer, 2017; Rösler et al., 2017). For example, End and Gamer (2017) compared eye movements of the participants from a free-viewing paradigm with predicted fixations from a visual saliency model. They showed that these low-level saliency approaches can be further improved when taking social features such as heads into account. Even introducing a *face channel*, i.e., a specific channel for higher-level semantic information based on a face detection algorithm, boosts the performance across images containing social aspects (Cerf et al., 2008). Interestingly, it does so also for images that do not depict any humans. These authors even conclude that faces can be considered to be processed in a way similar to a bottom-up saliency path in human vision, in contrast to the classical notion.

From a physiological perspective, a dedicated distributed neural system for face perception has been discussed in the literature for years (Haxby et al., 2000). In general, evidence suggesting specific brain regions for the processing of social stimuli are comparably old (for example, Perrett et al., 1992). These studies found specific *face* cells in primate brains (Gross et al., 1972; Perrett et al., 1992) and are backed by several more recent findings (Pinsk et al., 2009; Schultz and Pilz, 2009). In fact, just recently, the existence of a third path in the well-established dichotomized (i.e., the dorsal and ventral) two-pathway model for visual object recognition was suggested (Pitcher and Ungerleider, 2021). Pitcher and Ungerleider (2021) argue that the cumulative evidence across the literature indicates an extension by a third stream for social information. Originally the model proposes a ventral stream that projects along the ventral brain surface. That stream is relevant for the processing of the identity of an object. It is considered relatively slow, processing mainly foveal and parafoveal input. The second original stream is the dorsal stream which is relevant for the location of an object. This stream is considered to be fast and processes visual input from the whole retina. However, newer evidence (e.g., Hein and Knight, 2008) provide evidence for a third pathway computing actions of dynamic faces and bodies and might be specific for social perception (Pitcher and Ungerleider, 2021). Like the other two paths, this path is assumed to originate from V1 as well. It projects into the motion-selective area V5 and across the anterior superior temporal sulcus towards the posterior temporal sulcus. As social attention evolved from general attentional research, (Pitcher and Ungerleider, 2021, p. 9) conclude that the former terms are "inadequate when it comes to describing the complexity and nuances of even basic social interactions".

Social attention, and social cognition in general, put a lot of emphasis on the processing of facial features, i.e., the eyes or emotional expressions. However, it is by no means restricted to these aspects. In fact, it covers humans in their full social capabilities (Emery, 2000). As an illustration, studies are also conducted about the integration of body postures with facial expressions (e.g., Azarian et al., 2017; Hietanen, 2002). Still, facial expressions play arguably the most critical role in human interactions. They are also assumed to be processed with the highest priority among social cues (Perrett et al., 1992). Therefore, they are in the focus of most in social attention research. All in all, compelling evidence from behavioral, computational, and neurobiological studies shows that social information is processed outside the traditionally assumed dichotomy of cognitive processes.

1.4 Social attention and gaze cueing

The previous section showed that social attention plays a central and unique role in human cognition. It is also of paramount importance for social behavior. Especially the eyes play a significant role (Emery, 2000), with their so-called dual function of gaze (Gobel et al., 2015; Nasiopoulos et al., 2015). The dual function of gaze refers to the fact that the eyes are capable of simultaneously perceiving and sending information, something that can be described as the mutual properties of eye contact. Such reciprocity can otherwise only be achieved by physical contact, and both are the most intimate interpersonal encounters (Heron, 1970). This reciprocity in eye gaze is also a remarkable tool in social interactions (Gobel et al., 2015). Across several lines of research, it could be shown that eye gaze is relevant for the initiation (Bayliss et al., 2013), maintenance (Böckler et al., 2011) and regulation (Thönes and Hecht, 2016; Wu et al., 2013) of social interactions. It is also highly relevant for synchronizing human behavior, e.g., in conversations (Oertel et al., 2012) or simultaneous tasks requiring close coordination (Kawase, 2013). On top, several psychiatric and neurodevelopmental disorders are associated with malfunctioning social attention. For example, increased vigilance and at the same time avoidance of eye contact in patients with social phobia (Boll et al., 2016; Wieser et al., 2009), reduced social attention in people from the autism spectrum disorder (Jones and Klin, 2013; Nation and Penny, 2008), or engagement in eye contact in ambiguous stimuli in patients with schizophrenia (Hooker and Park, 2005; Tso et al., 2012).

The first investigation of direct gaze on social interactions was already conducted in the 70s. The research at that time investigated research questions such as the influence of direct gaze on flight behavior (Ellsworth et al., 1972) or how it causes intimacy (Argyle and Dean, 1965; Ellsworth and Ross, 1975). In general, research at this time primarily focused on the social aspects of interactions, not so much on cognitive processes. Kleinke (1986) provided an overview of the gaze literature predating the field of social cognition. He identified five central gaze functions: (1) providing information, (2) regulating interactions, (3) expressing intimacy, (4) exercising social control, and (5) following goals. Notably, there is considerable overlap with recently identified functions of direct gaze on human cognition (Conty et al., 2016): (1) effects of attention capture by direct gaze (Böckler et al., 2014; Hietanen et al., 2016; Mares et al., 2016) (2) enhancement of self-awareness (e.g., for regulating interactions as shown in Hietanen et al., 2008), (3) memory effects (as shown in Vuilleumier et al., 2005, where mutual gaze facilitates recognition memory), (4) activation of pro-social behaviors (e.g., as shown in a dictator game study where social behavior is fostered, Nettle et al., 2013), (5) positive appraisals of others (e.g., as mutual gaze duration increases likeability, Kuzmanovic et al., 2009). This overlap highlights the close relationship between social cognition and behavior.

Just as direct gaze, following another person's gaze is an equally important skill in humans. For example, it is relevant for language acquisition (Brooks and Meltzoff, 2005) and for social learning in general (Tomasello et al., 2005). As such gaze following has a central function in interactions with conspecifics. It reveals the attentional focus of others and provides valuable information about the environment. Additionally, gaze also provides information about others state of mind, particularly in combination with facial expressions (Baron-Cohen, 1991; Perrett and Emery, 1994). For example, gaze can be used to conceal communication. Explicitly making eye-contact with a friend and looking towards the door might result in shift of the friends attention towards the door. Using a cold and neutral facial expression, this might indicate that one wants to leave the party. In contrast, the same gaze towards the door with eyes wide open and a happy facial expression might indicate an upcoming surprise and the unspoken message to focus the door. Thus, gaze provides necessary information for successful social interaction and navigation through our social environment. It is of central relevance for understanding social cognition (Emery, 2000; Frischen et al., 2007a; Nummenmaa and Calder, 2009; Shepherd, 2010).

The gaze cueing task is the most fundamental task to investigate basic social cognition. A frequency analysis of the term "social cognition" in search results from the *Web of Science* reveals that the first studies using this task in 1998/99 predate the prominent use of the term starting in the early 2000s (see Figure 1.1). Gaze cueing can be seen as one of the first empirical research designs to systematically investigate social cognition in the laboratory. The origin of the gaze cueing task lies in the Posner cueing task (Posner, 1980). In spatial cueing tasks directional cues are assumed to shift participants' covert attention towards a cued location. The gaze cueing task is a special form of a Posner cueing task, and its defining aspect is a face as a central cue. The most widely used measures are reaction times from manual responses to different types of trials. Alternative measures include saccadic eye movements or error rates. In valid trials, a face *looks* to the location where the target appears. In invalid trials, a face *looks* to the location where the target does not occur (see Figure 1.2 for an illustration). Participants' task is to respond to a given target as fast as possible. Faster responses to validly cued targets are generally reported. This difference in reaction times between trials with a valid and an invalid cue is called the gaze cueing effect. It is widely assumed that the overall facilitation in reaction times is due to a covert shift of attention (among others, e.g., Friesen and Kingstone, 1998; and for the general Posner cueing task, see Posner, 1980).

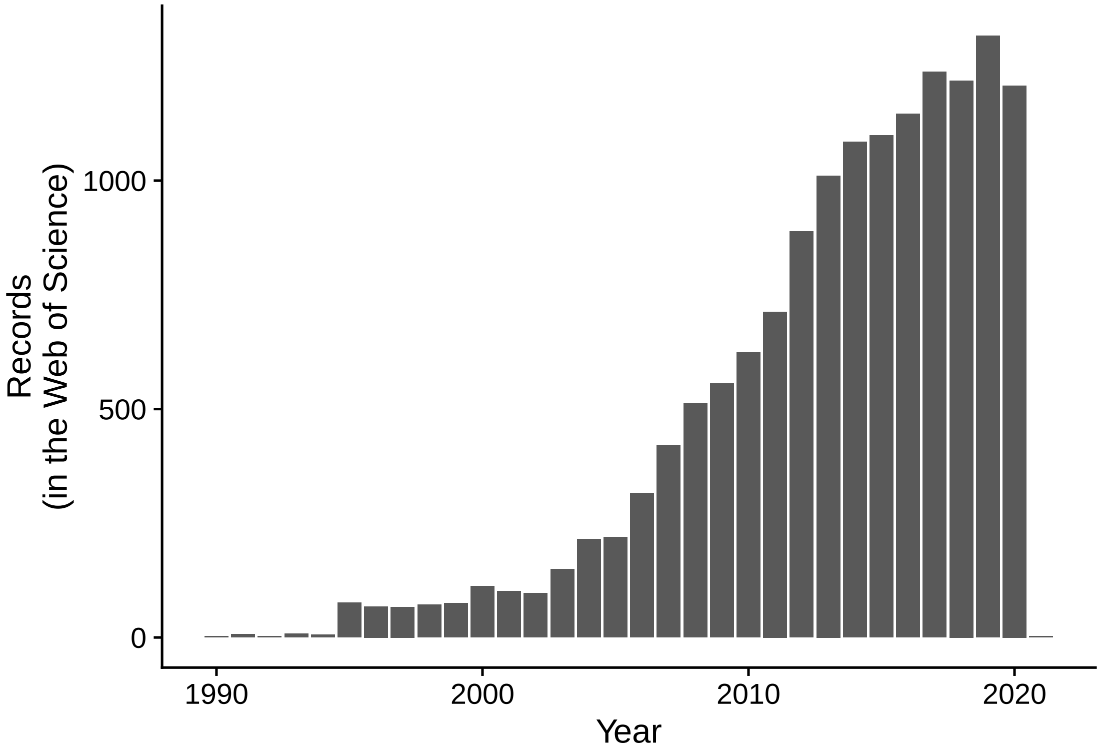

Figure 1.1: Histogram of Records in the Web of Science of the term *social cognition*.

Before gaze cueing studies became popular, studies on eye contact and gaze broadly discussed and focused on higher cognitive processes (e.g., in the context of liking and attractiveness in conspecifics, for a contemporary review, see Kleinke, 1986). The gaze cueing task was initially used and continues to be used for the systemic and controlled investigation of the foundations of social cognition and underlying mechanisms of attentional shifts in response to social stimuli. Nowadays the task is also used to address to research questions with regard to more complex social aspects (e.g., Strick et al., 2008, on attractiveness; or King et al., 2011, on trustworthiness). Until today, gaze cueing tasks are used in countless facets and variations across most subfields of psychology, but the classic paradigm remains relevant for basic social cognition research. However, it has also evolved to less constraint gaze cueing setups (e.g., free viewing paradigms as used in Großekathöfer et al., 2020). In this vast field multiple comprehensive reviews have been published on gaze cueing (Dalmaso et al., 2020b; Frischen et al., 2007a; Langton, 2000) and related topics (e.g., on the neurophysiology of gaze cueing: Perrett et al., 1992; and Emery, 2000; on clinical and developmental characteristics of gaze cueing: Nation and Penny, 2008). Here, we attempt to add to these perspectives a systematic review of the literature with a quantitative analysis of the published gaze cueing literature (see Study 1 from Chapter 4). We aim to address open research questions from a new methodological perspective on the matter.

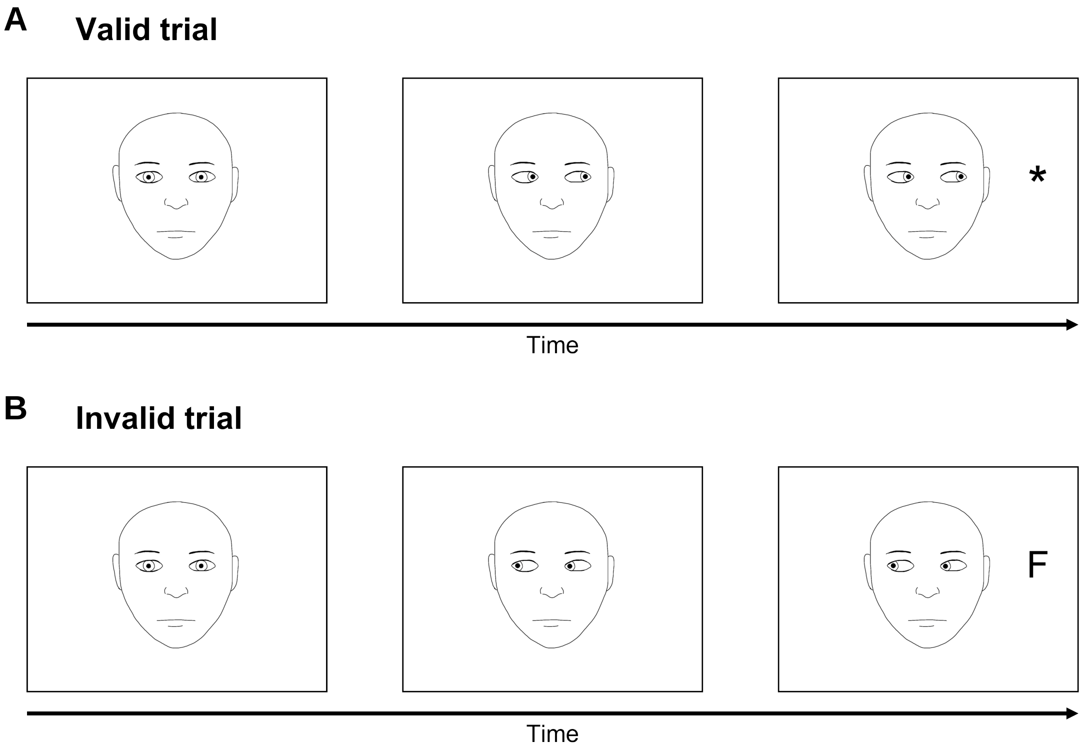

Figure 1.2: Schematic sequence of two classic gaze cueing trials. (A) A valid trial with a typical target used for detection or localisation tasks, depending on the instructions given to the participants. (B) An invalid trial of an identification task, e.g., typically between two or more letters or symbols. Depicted faces are modified line drawings by Barry Langdon-Lassagne, distributed under a CC-BY 3.0 license.

As mentioned before, the gaze cueing design evolved remarkably over time and addressed a plethora of research questions. Our meta-analysis focused exclusively on research applying the classic gaze cueing design (see Methods section of Study 1 for our definition). Thus, we systematically capture researchers' varying use of the basic gaze cueing task and extracted theoretically relevant moderators from those classic research designs.

Per definition, gaze cueing depends on a face. What constitutes a sufficient representation of a face? The very first studies on gaze cueing already covered a range of facial representations. On the lowest abstraction layer are cartoon faces drawn only with lines. Interestingly, such cartoon faces elicit gaze cueing, although they only provide an unnatural representation of a face (e.g., Friesen and Kingstone, 1998; or for a more recent study see, e.g., Takao et al., 2018). On a less reduced abstraction layer are photographs of faces. Again, gaze cueing is effective with such complex facial representations (Hietanen, 1999; Langton and Bruce, 1999).

Line drawings of faces come with various desirable features (Prazak and Burgund, 2014). They allow a perfectly balanced visual appearance (such as luminance, vertical symmetry, ...). They only introduce minimal to none confounding possibilities, and they enable the investigation of prototypical facial features and expressions. Some authors argue, that these prototypical features facilitate cognitive processes as they reduce ambiguity (Lundqvist et al., 1999; Öhman et al., 2001). On top, line drawing or cartoon faces have the advantage that they are completely in the researchers' control and that features can be changed in strict isolation. This is in strong contrast to photorealistic representations of faces which easily introduce confounding factors. For example, humans derive certain personality traits (Little and Perrett, 2007) or sexual behavior (Boothroyd et al., 2008) from static faces alone. Realistic faces might activate stereotypes (Becker et al., 2007) or transport persons' characteristics like trustworthiness (Strachan et al., 2017). These constraints are at the same time problematic but also a valuable source for investigating such factors. In the first place, they often enable the research of other, higher aspects of social interactions. An approach to address the problems introduced by photorealistic stimuli was the development of validated face databases, some including faces with directional gaze information (for example, the Radboud Faces Database Langner et al., 2010). An advantange of photorealistic stimuli is that they closely resemble what humans are encountered within social interactions in reality. Such ecologic validity is known to be especially relevant in social cognition research (Risko et al., 2012). In the end, both stimuli have their strengths and weaknesses.

Interestingly, however, an systematical comparison of the stimulus types' influence on gaze cueing has, to our knowledge, never been conducted. Until now, it remains an open question of whether gaze cueing effects differ depending on the approximation of a real face. Such a differences might yield important theoretical implications about feature-based or holistic gaze processing in social cognition. For example, in research investigating facial expression recognition, the difference between cartoon and photorealistic faces is discussed to be the reason for diverging findings in the literature (Prazak and Burgund, 2014). Thus, it might similarly reveal differences and provide further insights into the cognitive processes underlying gaze cueing. These insights might also be relevant for planning a study by providing a basis for informed decisions on which kind of stimuli researchers should favor in their research designs, for theoretical but also statistical reasons (i.e., for justifying the sample size).

As mentioned above, photographs of faces transport an impressive amount of visual information. Which information is relevant for gaze cueing? Directional information can be read from head orientation and pupil position. In fact, these sources are closely linked, often aligned in the same way, and basically do not appear in isolation. Unsurprisingly, gaze cueing can be found for eye cues with a forward-facing head and eye cues in a congruently aligned head (for example, Hietanen, 1999). However, theoretical models from comparative neurological studies suggest that eyes take precedence over the head (Perrett et al., 1992), which should be observable in behavioral measures. Some studies that systematically investigated this hypothesis found that gaze cueing is reduced when elicited by head orientation (Hietanen, 2002, 1999), thus supporting the findings from macaques (Perrett et al., 1992). This is remarkable considering the pupil size and the conditions when the pupil is visible (compared to head orientation). Other studies, however, reported robust and equivalent effects of gaze cueing with head orientation (e.g., Langton and Bruce, 1999; Xu and Tanaka, 2014). These research findings question the idea of eyes as the dominant feature. Our meta-analysis further investigates this research question. A meta-analysis is particularly suitable for this research question as in the vast majority of gaze cueing studies, the cue information is not explicitly manipulated. Most studies used eyes as directional cues, which might not be surprising considering that simple cartoon faces are not easily shown in profile, i.e., with head orientation. A meta analysis allows the comparison of both experimental manipulations between studies. From a systematical comparison of the manipulation across studies, we aim to draw inferences beyond what was motivated by the authors of the included studies.

Additionally, we were interested in the temporal dynamics of attention during gaze cueing. Two effects are discussed to be of relevance here: Inhibition of Return (IOR, Posner and

Cohen, 1984; Wang and Klein, 2010) and the gap effect (Fischer and Ramsperger, 1984). In general, the assumption is that differences in reaction times provide insights into the temporal flow of attention and cognitive processes. The IOR describes an aftereffect of attentional shifts to facilitate the detection of new events (Posner et al., 1985). Reaction times in a specific time window are slower to a target at a previously attended, i.e., cued location. This is assumed to indicate an inhibitory process. For classical Posner cueing tasks, a reliable IOR is found from 200 ms to 1000 ms (for a meta analysis, see Samuel and Kat, 2003), which couldn't be found for gaze cueing tasks (Friesen and Kingstone, 1998; Frischen et al., 2007b; Frischen and Tipper, 2004; Langton and Bruce, 1999). Therefore, it is discussed as an exclusive property of facial stimuli, which might highlight the relevance of such gaze cues in human cognition (Frischen et al., 2007b; Frischen and Tipper, 2004). However, some studies show an IOR at the before mentioned time window (Friesen and Kingstone, 2003; Newport and Howarth, 2009; or Gayzur et al., 2013), whereas others argue that very long SOAs are crucial for an IOR in gaze cueing (Frischen and Tipper, 2004).

The gap effect refers to faster responses when a central cue disappears before target presentation. Again the gap effect is failed to be observed in gaze cueing studies, i.e., temporal overlap between cue and target presentation (from here on referred to as cue-target overlap) does not facilitate the gaze cueing effect. Thus, it is discussed as a distinguishing factor (Friesen and Kingstone, 2003; Gayzur et al., 2013). Furthermore, IOR seems absent when cue and target have temporal overlap (Collie et al., 2000; Frischen and Tipper, 2004; Green et al., 2013). Thus, albeit not widely discussed, the gap effect might be as relevant as the IOR in gaze cueing. Overall, studies used overlapping and non-overlapping cue-target designs alike. Together with various SOAs, we can further investigate the IOR, gap effect, and an interaction from a meta-analytic perspective, although only two studies explicitly manipulated the relevant parts of the research design.

As the introductory example illustrates, faces transport much more than directional information, for example, emotions. These facial expressions can change the meaning of a gaze cue. For example, a fearful face might elicit a different reaction than a happy face. Combined with a directional gaze cue, the former might indicate a source of threat and thus marks an important warning. In contrast, the latter indicates a source of joy. The underlying question is whether and how other facial information is integrated with the processing of directional cues. Previous research showed a general processing advantage for emotions like anger or fear. Both emotions are discussed to be more relevant than other emotions (i.e., anger or fear, Eimer and Holmes, 2002). It might be evolutionary beneficial to facilitate responses to particularly relevant emotional directional cues. More specifically, it is presumed that fearful faces elicit stronger gaze cueing effects, i.e., faster responses to targets compared to non-fearful faces. Two theoretical explanations are discussed as mechanisms for the facilitated processing of gaze cues. First, fearful faces indicate a threat, which might be costly to overlook, especially compared to overlooking happy faces (Fox et al., 2007). Another theory claims that the bigger eyes in fearful expressions increase the saliency of the gaze cue and thus facilitate gaze cueing (Tipples, 2006,0). The idea that emotions are relevant in gaze cueing originated early in the literature and is still under empirical investigation. Up to today, studies did not accumulate convincing evidence for these hypotheses. Some studies did not find evidence (Hietanen and Leppänen, 2003; Graham et al., 2010; Bayliss et al., 2007) whereas others did (Tipples, 2006; Bayless et al., 2011; Neath et al., 2013; Pecchinenda et al., 2008; Ponari et al., 2013; Lassalle and Itier, 2015). Additionally, some studies even assume moderation by personality traits and states (e.g., introversion/extraversion, Ponari et al., 2013; or anxiety, e.g., Fox et al., 2007; and Mathews et al., 2003) others again, do not support such findings (for anxiety Holmes et al., 2010). A meta-analysis in this context offers excellent potential to quantitatively summarize such competing evidence.

Gaze cueing in the last 20 years produced a remarkable body of evidence, and a meta-analysis is capable of synthesizing these research findings and at the same time offering enhanced objectivity and reliability (Anderson and Maxwell, 2016; Gurevitch et al., 2018). On that basis, we were first interested in the estimated *true* average gaze cueing effect and its accompanying heterogeneity by combining 58 gaze cueing studies within a multi-level mixed-effects meta-regression. Second, we analyse manipulations that might influence the magnitude of the gaze cueing effect that were either not investigated by the original studies or with competing evidence. Third, we were interested in the temporal properties of the gaze cueing effect. Overall, our goal is to stimulate new hypotheses and point towards future research directions for gaze cueing studies and social cognition in general. Furthermore, we aim to provide justified estimates that can be used for sample size justification in future studies.

Direct gaze and gaze following also influence higher-order social behavior. As stated before, some of these studies even predate the field of social cognition (see for example Argyle and Cook, 1976; Ellsworth and Ross, 1975; Ellsworth et al., 1972). However, these phenomena are still subject to research within the field of social cognition today. Such studies underline that direct and averted gaze serve as a tool for social evaluation and regulation. For example, the hierarchy of group members is reflected in the attentional resources spent. Foulsham et al. (2010) altered participants' viewing patterns of group discussions by manipulating participants' believe whether a high-rank or low-rank group member was later to observe the viewing behavior. The authors argue that participants became aware of the signaling aspect of gaze. Indeed, they found that participants were very sensible to who might read their signals from the eyes: the higher rank group members in the video received most attention but only when participants believed that these candidates did not see their viewing patterns. Consequentially, they might have avoided confrontation with higher ranks by avoiding eye contact by proxy. Furthermore, it highlights that these processes are also affected by social norms. Here regarding social dominance, but others have also provided evidence for social desirability (Risko and Kingstone, 2011; Roepstorff and Frith, 2004) or the possibility to interact (Laidlaw et al., 2011). After all, gaze following is not only a "Window into Social Cognition" (Shepherd, 2010, p. 1) but also has critical functions in social behavior. Thus examining the functionality of gaze need to dissociate these signaling and reception properties of gaze (Gobel et al., 2015).

All in all, a wealth of evidence from behavioral, computational, and neurobiological studies suggests specific cognitive processes for social attention. These processes share features of bottom-up processing, e.g., in their rapid and reflexive processing in social attention. However, they are at the same time heavily affected by the social environment (e.g., norms). In sum, cognitions in social situations have unique properties that confront social attention research with special requirements.

Chapter 2

Social attention and ecological validity

2.1 Basic concepts of validity

Science, like religion and the arts, is a social approach to understanding the world. These approaches have in common that all use a shared meaning system consisting of concepts and relationships. Science differs from the other approaches in the way these concepts are validated. It is the sole approach using *systematic empirical validation* for inferences about the world (Jaccard and Jacoby, 2020). In modern psychology, the empirical validations of concepts and relationships come from observations obtained by experiments, tests, or questionnaires. These observations have to meet specific criteria to be considered scientific. The criteria are most prominently defined in the classical test theory, a framework for empirical validation that is widely accepted (Lord et al., 1968; Novick, 1966).

According to the classical test theory, an observation is composed of the true (aimed for) construct value, random error, and systematic error. When an observation is affected by random error, reliability is said to be limited. Consequently, this source of error randomly increases or decreases the observed value to an unknown degree on every measurement. So measuring a person twice might lead to different observations, a difference that might be attributed to random error. Accordingly, an observation free from random error is reliable, and measuring a person twice results in identical results (Jaccard and Jacoby, 2020). Reliability can be assessed with specific statistical tests such as a test-retest or a split-half test. Such tests provide metrics from 0 to 1, indicating the amount of random error between observations. An entirely reliable observation has a reliability score, e.g., calculated as Cronbach's α, of 1 (Cronbach, 1951).

Interestingly, two historic streams of psychology research use different definitions for *reliability* (Cronbach, 1957; Hedge et al., 2018). One stream, correlational research, aims to distinguish individuals by capturing between-subject variances, such as personality psychology and intelligence research. In these fields, the term is correctly used as defined above. In contrast, the other stream, experimental research (such as social cognition research), aims to identify cognitive mechanisms by capturing within-subject variance. The term is often used in this stream to describe the *replicability* of an effect in another experiment. In the proper sense, cognitive tasks often have low reliability (Hedge et al., 2018, and sometimes even low replicability, Open Science Collaboration, 2015). The lack of reliability of these experiments comes by design. Their main purpose is often to uncover shared cognitive mechanisms by optimizing experiments for group differences aiming for low variance between subjects (Hedge et al., 2018).

Contrary to reliability, validity is more difficult to assess. According to classical test theory, reliable observations are considered valid when they are not biased by systematic error. Systematic errors can occur for various reasons, and, in fact, the main task for science is uncovering sources of systematic error. Only the to be observed construct should be reflected in the observation and deviations might indicate potentially relevant moderators or an invalidly conceptualized construct. Thus, the validity of observations, and therefore of an experiment is of vital importance to science to uncover moderators. Without validity, it is unknown what was observed or measured. However, there are no commonly agreed-upon methods or metrics to test validity (Jaccard and Jacoby, 2020). Until today, several types of validities with overlapping definitions have been proposed in psychology research (Brunswik, 1947; Campbell, 1957; Holleman et al., 2020; Mook, 1983; Schmuckler, 2001; Shamay-Tsoory and Mendelsohn, 2019). Despite this ongoing debate, there is somewhat consensus about the common idea to assess validity. Generally, the higher the correlation of observations with known measures (e.g., convergent and concordant validity), the more an observation is considered valid. Alternatively, correlations within a meaningful theoretical network (e.g., construct validity) or the absence of such correlations (e.g., discriminant validity) can also hint for validity (Jaccard and Jacoby, 2020).

In recent years a debate unfolded about the ecological validity of social attention experiments. Ecological validity can be seen as a special case of convergent validity. In most cases, social behavior is observed in the laboratory and then correlated with well-established patterns of natural human behavior. Thus, it describes the proximity of the observed human behavior, with behavior considered natural (Schmuckler, 2001). In that regard, especially laboratory studies were questioned regarding their ecological validity. Laboratory studies in psychology are designed to control for random and systematic error in observations (Hedge et al., 2018). Contemporary social cognition research makes use of the investigation of human social behavior with computerized experiments, such as the gaze cueing paradigm (see Section 1.4). In contrast, early research often required experimental interactions with the experimenter (e.g., as in Argyle and Ingham, 1972). Such computer experiments have long known advantages for precision and reliability (Aaronson et al., 1976). The prominent critic these laboratory studies have to face is that behavior from the laboratory seems not or only under specific conditions to generalize towards natural environments.

2.2 Ecological validity in gaze cueing paradigms

This section also appeared largely in (Großekathöfer et al., 2020, p. 135 – 137).

Humans in their social environment rely on the information conspecifics provide. This does not only hold for reading explicit signals, such as verbal communication, but also for implicit signals, such as eye gaze or nonverbal cues. Specifically, if an individual looks into a certain direction, this information is often read spontaneously by an observer who redirects his or her attention towards the referred object or location. Such guidance of someone else's attention is called gaze following. As a consequence, joint attention is established.

The most frequently used paradigm to investigate such attentional shifts is the so-called gaze cueing paradigm (Friesen and Kingstone, 1998; Driver et al., 1999; Langton, 2000; for a review see Frischen et al., 2007b). This paradigm has been inspired by classical studies on spatial attention by Posner (1980) and consists of a centrally presented face with varying gaze directions. This face is then followed by a subsequently presented target at either the cued location (i.e., the location that the face is looking at) or an uncued location (i.e., a location that is not being looked at by the face). Studies using this gaze cueing paradigm have demonstrated that gaze cues facilitate target processing as evident in smaller reaction times to targets at cued as compared to uncued locations (Frischen et al., 2007b). The paradigm was also used to show that gaze following is shaped by high-level social cognitive processes like group identity (Liuzza et al., 2011), theory-of-mind (Cole et al., 2015; Teufel et al., 2009; Wiese et al., 2012; Wykowska et al., 2014) or physical self-similarity (Hungr and Hunt, 2012).

However, even though gaze cues are crucial for joint attention, this standard gaze cueing paradigm can be criticized for lacking ecological validity. Whereas in the real world, gaze signals occur within a rich context of competing visual information, gaze cueing studies typically used isolated heads (Friesen and Kingstone, 1998; Langton, 2000) or even cartoon faces (Driver et al., 1999; Ristic and Kingstone, 2005) as gaze cues (for an overview see: Risko et al., 2012). Although gaze cueing was also found with more naturalistic stimuli (Perez-Osorio et al., 2015), in a recent study in which Hayward et al. (2017) compared attentional measures of gaze following from laboratory (classical gaze cueing) and real world (real social engagement) settings, they did not find reliable links between those measures.

As a compromise between rich but also less controlled field conditions and standardized but impoverished laboratory studies, complex naturalistic scenes were used to investigate gaze behavior in laboratory settings (e.g., Fletcher-Watson et al., 2008; Perez-Osorio et al., 2015; Zwickel and Võ, 2010). To specifically explore the influence of gaze cues, Zwickel and Võ (2010) and Perez-Osorio et al. (2015) used pictures of a person (instead of isolated heads or faces) as a directional cue within a naturalistic scene. Zwickel and Võ (2010; in contrast to the gaze cueing task chosen by Perez-Osorio et al., 2015) used a free viewing instruction, meaning that participants had no explicit task to fulfill but should just freely explore the pictures. The authors argued that the lack of a specific task puts gaze following to a stricter test since previous studies frequently used target detection tasks (e.g., Langton et al., 2018) or comprised specific instructions such as asking participants to understand a scene (Castelhano et al., 2007). Consequently, in those latter studies, it remains unclear to what degree gaze following occurred spontaneously or was caused by the specific task at hand. In detail, Zwickel and Võ (2010) presented participants multiple 3D rendered outdoor and indoor scenes for several seconds that always included two clearly visible objects as well as either a person or a loudspeaker that was directed towards one of these objects. The loudspeaker, which also represents an object with a clear spatial orientation, served as a control condition to ensure that gaze cueing effects are due to the social meaning (i.e., the direction of the depicted person's gaze) as compared to a mere following of any directional cue. The results of the study showed that participants fixated the cued object remarkably earlier, more often and longer than the uncued object. By showing that leaving saccades from the head most often landed onto the cued object, the results gave further evidence for the direct influence of eye gaze on attentional guidance. Crucially, similar effects were not obtained for the loudspeaker. The cued objects were not just focused because they might have been salient by themselves (e.g., due to positioning), or because they were cued by another object, but became more salient by the person's reference. To sum up, Zwickel and Võ (2010) provide convincing evidence that joint attention is a direct consequence of gaze cues and gaze following, it happens spontaneously and has high relevance even in situations that are more naturalistic (i.e., involve complex scenes and the absence of explicit tasks) than classical gaze cueing studies based on variations of the Posner paradigm.

In the current study, we were first interested in whether the previously reported effects hold when using a different set of stimuli. Replication in itself is a core concept of scientific progress (Schmidt et al., 2009) and thus relevant for assessing the stability of effects. Nevertheless, our motivation was also to improve certain aspects of the study and at the same time extending this line of research. Due to their low resolution and reduced richness of details, the 3D rendered scenes used by Zwickel and Võ (2010) did not allow for an assessment of the depicted person's gaze direction. As a consequence, the observed cueing effects could be due to directional information inferred from both the body and head of the person. We therefore developed a new set of photographic stimuli that had sufficient resolution to also allow for perceiving gaze direction with clearly visible eyes of the depicted person. These photos always included a human being who directed his/her gaze towards one of two objects that were placed within reaching distance. In order to be consistent with the study of Zwickel and Võ (2010), the depicted person's head and body were congruently aligned with his/her eye gaze. Second, in order to extend this line of research, we manipulated top-down attentional processes by task instruction to explore the susceptibility of gaze following effects in naturalistic scenes. Earlier research showed that social attention can be influenced by multiple factors like social status of the observed persons (Foulsham et al., 2010), possibility to interact (Hayward et al., 2017; Laidlaw et al., 2011) or social content (Birmingham et al., 2008b). Together with Zwickel and Võ (2010), these studies have in common that they manipulate viewing behavior of the participant by manipulating the stimuli or environment. In contrast, in the present study, we tried to modulate viewing behavior via task instructions (for a similar procedure see Flechsenhar and Gamer, 2017). Specifically, half of the participants received an instruction before the viewing task, that they should try to remember as many objects from the scenes as possible (explicit encoding group). The other half of the participants (free viewing group) merely received the instruction to freely explore the pictures and the memory test that was accomplished after the experiment was unannounced and therefore reflected spontaneous encoding of the respective scene details. The motivation for this manipulation was twofold. First, it was thought to test the robustness of gaze following against top-down processes by discouraging observers to utilize the information provided by eye gaze. Second, it allowed for examining gaze following effects on memory.

We expected to replicate the findings of Zwickel and Võ (2010) in the free viewing group. Specifically, we anticipated to observe an early fixation bias towards cued objects, an enhanced exploration of these details (i.e., more fixations and longer dwell times) and more saccades leaving the head towards the cued as compared to the uncued object. The instruction in the explicit encoding group was thought to induce a more systematic exploration of the presented scenes resulting in higher prioritization of both objects and reduced cueing effects. Furthermore, we anticipated a generally enhanced recall performance in the explicit encoding group. Due to the expected difference in attentional resources spent on the cued and uncued object in the free viewing group, memory performance of the cued object was expected to be better compared to memory performance of the uncued object. Finally, as previous studies showed a strong preference of fixating the head over body and background regions in static images (End and Gamer, 2017; Freeth et al., 2013), we expected to see a similar bias in the current study regarding dwell times, number of fixations and fixation latency. Additionally, we hypothesized that the prioritization for the head decreases when participants follow specific exploration goals such as in the explicit encoding group of the current study (cf., Flechsenhar and Gamer, 2017).

Please see Study 2 from Chapter 5 for the methods, results, and interpretation of the empirical work.

Alternatively to enriching laboratory settings to increase ecological validity, field studies are naturally thought to score high on ecological validity. In fact, when assessing the ecological validity of behavior observed in the laboratory, it is often correlated with natural behavior measured in field studies. Thus, for example, the common notion in social cognitive research is that social stimuli are processed preferentially (see above, or, e.g., End and Gamer, 2017). However, measurements from field studies often point in the opposite direction (e.g., Laidlaw et al., 2011).

2.3 Ecological validity and VR

This section also appeared largely in Großekathöfer et al. (in press).

Social cognition research places great hope in virtual reality (VR) to overcome limitations of laboratory studies and resolve discrepancies between findings obtained within restricted laboratory contexts and naturalistic situations (Parsons et al., 2017; Risko et al., 2012; Rubo and Gamer, 2021; Zaki and Ochsner, 2009). These discussions are based on the critique that social cognition research frequently involves simplified stimuli that do not represent reality, which is multimodal, dynamic, and contextually embedded (Zaki and Ochsner, 2009). One area of research where these considerations became especially prominent in recent years is the field of social attention. In general, attentional shifts towards humans beings due to their sole presence in the visual field are well documented (Birmingham et al., 2008b; End and Gamer, 2017; Großekathöfer et al., 2020; Rösler et al., 2017). However, such preferred visual exploration of conspecifics seems highly reduced in reality (Horn et al., 2021; Laidlaw et al., 2011; Rösler et al., 2021). As a consequence, researchers sought more appropriate research designs that approximate real social environments but at the same time still provide experimental control (Risko et al., 2012,1). A solution often discussed in this context is VR since it allows for multimodal, contextually embedded, and dynamic stimulus presentation (Parsons et al., 2017). In principle, it can enable researchers to observe natural viewing behavior in the laboratory without losing experimental control.

The use of VR for examining social attention is a rather recent development. It has not yet been extensively used to assess attentional prioritization of human beings (for an exception, see Rubo and Gamer, 2021). An experimental design that has been more frequently applied in this domain concerns the examination of social attention in the real world using mobile eye-tracking glasses and comparing these findings to a presentation of video recordings on a computer screen to either the same (Foulsham et al., 2011) or another participant (Rubo et al., 2020). These studies provided initial evidence that attentional allocation towards human beings differs between laboratory and real life conditions. For example, Foulsham et al. (2011) found generally low fixation probabilities on persons, which were further reduced in real life conditions when people were near the observer or remained in the visual field for longer durations. Although Rubo et al. (2020) did not confirm a general avoidance of gaze towards conspecifics, they found an increased exploration of people located in the observer's vicinity. This bias, however, was less pronounced in the real world as compared to the laboratory situation.

Although such studies provide initial evidence for crucial differences between laboratory and field conditions, they are not without limitations. First, participants accomplished different tasks in both contexts (e.g., walking around vs. watching a video) which might induce different patterns of visual exploration (e.g., for avoiding obstacles when planning walking routes). Second, head and body movement were restricted in the laboratory and it is well known that saccadic eye movements differ substantially between conditions with restrained as compared to unrestrained head movements (for a review see Freedman, 2008). Third, previous studies involved presenting videos to participants in the laboratory context that were recorded by a head-fixed camera of the same or another participant in the field. Thus, participants in the laboratory could not freely decide where to orient their attention. All in all, these limitations may restrict the generalizability of findings and undermine conclusions that were based on a direct comparison of visual exploration patterns between laboratory and field conditions. Please note that although some of these problems might be addressed by including the video presentation into the real environment itself (Laidlaw et al., 2011), other problems such as the limitation of the field of view (FOV) to the previous recording condition persist. Moreover, such settings might be limited to certain experimental situations where a video playback in the surrounding is not considered unnatural or strange.

In the current study, we designed a novel experimental setting to solve these issues and provide a rich and ecologically valid viewing situation (Shamay-Tsoory and Mendelsohn, 2019; but see Holleman et al., 2020, for a critical comment). Specifically, we presented participants with spherical videos[1] of public places using a head-mounted display (HMD) with an integrated eye-tracker. Such stimulation has several advantages compared to previous screen-based experiments. First, it enables participants to actively and freely experience an environment including unrestricted head movements and some degree of body movement (e.g., turning around). Second, the participant's perspective is contextually embedded in the scene, i.e., she cannot look behind the scene. Whereas in traditional screen-based experiments, participants can evade the stimulation by looking around, such behavior is impossible within the HMD-based presentation of spherical videos. And third, the currently proposed viewing situation enables experimental control over the stimulation, which has been proposed to be one major advantage of VR above field examinations (Parsons et al., 2017).

Compared with 3D rendered virtual scenes, spherical videos come with a number of advantages but also have some limitations. The main advantage is that rich naturalistic stimuli can be generated remarkably faster, cheaper, and easier as compared to the extensive and costly development of 3D worlds. This seems especially true when these scenes include human beings. The main limitations are that interactions with the virtual environment, scenes that hurt physical laws (such as gravity), or scenes with naturalistic 3D properties (i.e., including stereoscopic vision) can hardly be realized with spherical videos. Another challenging aspect for VR in general is movement. Active, self-paced, and continuous movements are difficult, costly, and demanding to include, even in 3D rendered scenes. Since this is basically a form of interaction with the environment, it is impossible to realize with spherical videos. A prominent solution to overcome such problems in 3D rendered scenes is passive movement (e.g., teleportation to a new location) which might also be realized with multiple spherical videos to some degree. After all, the decision on how to realize a VR scene needs careful considerations but it seems plausible to assume that being contextually embedded and em-

---

[1] Spherical videos, also referred to as 360° videos, are videos recorded with multiple cameras to cover the whole surrounding at a specific location (see Methods section of Study 3 for more details). When we refer to spherical videos in the current article, we imply its presentation through an HMD with head tracking enabled. Please note that such spherical videos can also be watched on standard monitors and smartphones. Depending on the software, navigation is then enabled through keyboard, mouse, or device movements.

powered to actively experiencing an environment should reduce demand characteristics and elicit a more natural viewing behavior.

The current study aimed at examining the suitability of spherical videos for investigating social attention, and we were interested in better understanding the boundaries of typical laboratory settings. Therefore, we specifically compared visual exploration pattern of participants when viewing spherical videos of five public places in the laboratory to their behavior when visiting the same spots in the real world. We chose to examine participants' behavior at several locations in order to ensure generalizability of findings across situational characteristics and to permit assessing the reliability of the current method by estimating the consistency of viewing patterns across the different locations in the video as well as the real life condition. Moreover, we specifically compared viewing behavior between conditions to determine to what degree measures of social attention generalize from the laboratory to field contexts. Although we are convinced that the currently used spherical videos have some advantages over previously used stimulation conditions, they still differ from the real world since participants cannot socially interact with pedestrians in the video and do not have to follow certain social norms when viewing the scenes in the laboratory (e.g., staring will not have consequences, Ellsworth et al., 1972). Since both factors are suspected to play a critical role in attentional allocation towards conspecifics (Laidlaw et al., 2011; Foulsham et al., 2010; see also, Gobel et al., 2015, for a discussion on the dual function of gaze), we expected a reduced amount of social attention in the real world as compared to the viewing of spherical videos. Finally, for exploratory purposes, we related the currently observed viewing behavior to questionnaire data on autistic personality and social anxiety traits.

Please see Study 3 from Chapter 6 for the methods, results, and interpretation of the empirical work.

Chapter 3

The present work

Research on social attention faces high demands to yield generalizable results that are also valid outside artificially limited conditions. In total, social attention research clearly points out that findings from traditional computer experiments are too often only valid in narrow boundaries. This limitation can be addressed by improving the ecological validity of an experiment. The main objective of the present work is to examine the status quo of the field and propose that VR is a fruitful tool for solving pressing issues in social attention research.

The first study systematically reviews the published literature on the gaze cueing paradigm and evaluates the findings statistically. A meta-analysis is the preferred method to accumulate evidence across multiple studies. It provides an exhaustive overview of the literature and allows follow-up investigations regarding the size and variability of the *true* effect size. Furthermore, it can assess biases in the published literature and provide distinct analyses of moderator relationships. In that regard, a meta-analysis is unrivaled across scientific methods to summarize the findings of a field and is an outstanding tool for preparing the ground for future studies.

The meta-analysis is accompanied by an empirical study assessing gaze cueing in a gradually more ecologically valid setting. Specifically, in a less restrained gaze following setup, the study investigates the generalizability of the gaze cueing effect. Specifically, it assesses the impact of more ecological valid environments, namely contextually embedded stimuli. Additionally, the research design is also designed to test the robustness of gaze cueing against top-down modulation. As such, the study provides important insights into potential moderators but also on the reflexiveness of social attention.

The last present work highlights the versatile benefits of VR technology for social cognition research. This claim is supported by a study complying with an often formulated and demanded development: The study of social attention in VR. Specifically, the viewing behavior in the real world is compared with a close approximation of it in the laboratory using spherical videos shown in an HMD. This study aims at providing insights regarding two critical issues for social attention. First, the visual approximation (including the enabled exploration behavior) between the laboratory and the field can not be reached by classical computer experiments. Thus, the study allows a dedicated evaluation of diverging findings in social attention. Second, the application of spherical videos and HMDs is still novel in social attention research. Therefore, the study might also represent a milestone for introducing VR in social attention research.

Part II

Empirical investigation

Chapter 4

Study 1

Multi-level meta-regression of gaze cueing effects

4.1 Brief summary

The gaze cueing effect describes the facilitation of reaction times to target stimuli in response to a directional facial gaze cue. Gaze cueing studies were among the first approaches to investigate social cognition systematically in highly controlled laboratory settings. In the last two decades, manifold evidence was collected to explain moderating factors for attentional orientation (e.g., the relevance of emotional faces or the precedence of the eyes over the head) and to capture temporal properties of attention (e.g., Inhibition-of-Return). Our systematic review includes a multi level mixed effect meta-regression of 58 studies with 211 observations from 83 samples and covers 2577 unique participants. Additional to summarizing previous lines of research, we also address novel research questions such as the degree of photorealism of the facial representation. In general, we find an average gaze cueing effect of $M = 13$ ms (95% CI [10.25 ms, 15.54 ms]) accompanied by surprisingly large heterogeneity ($I^2 = 87\%$). We do not find evidence for the facilitation of the gaze cueing effect for specific emotional faces. In general, gaze cueing seems to be robust amongst most experimental moderators. However, we find evidence for the assumed preferential processing of eye cues (over head

cues). Overall, our meta-analysis provides a systemic review and examination of the gaze cueing literature and reveals important areas for future research.

4.2 Methods

4.2.1 Literature search and coding procedure

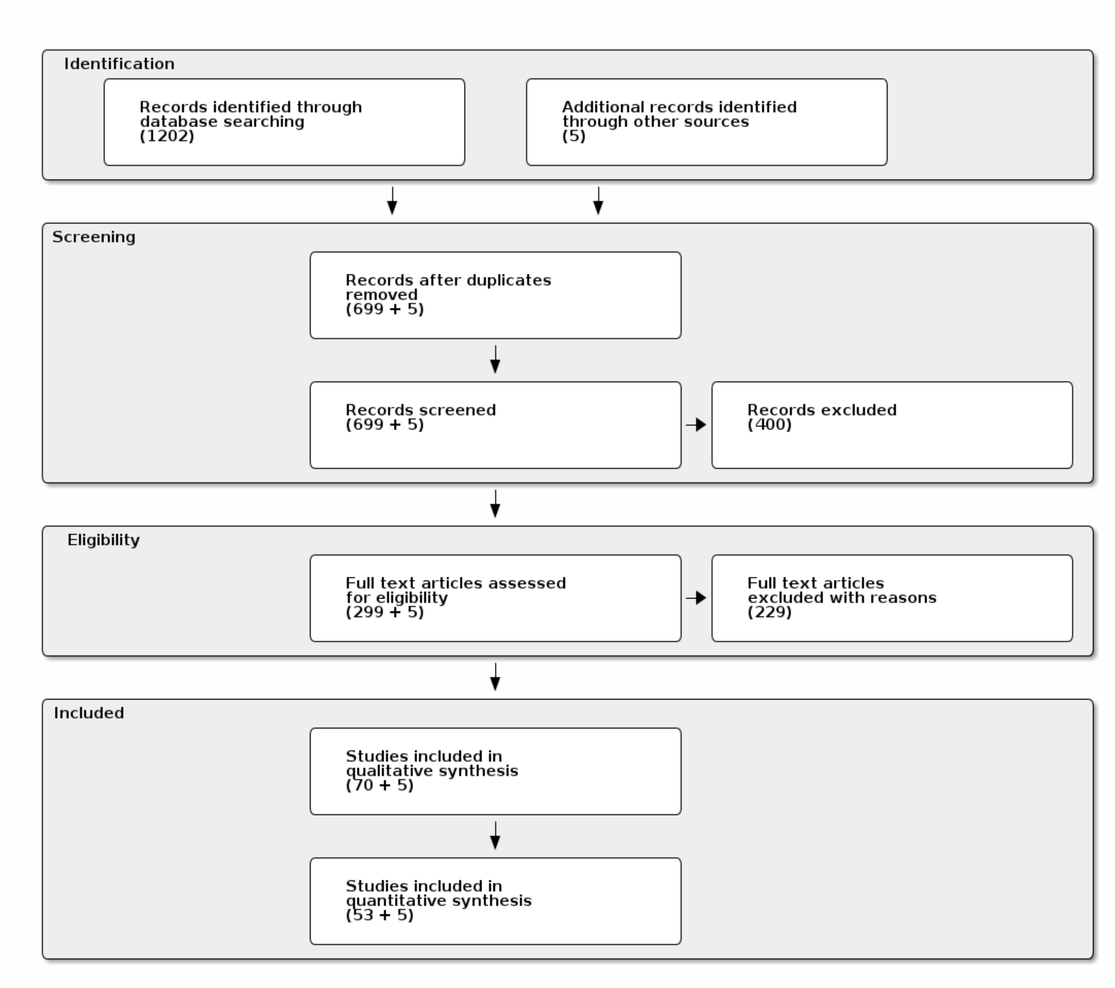

Figure 4.1: PRISMA-style flowchart showing the study selection for meta-analysis on the gaze cueing effect

Figure 4.1 shows a PRISMA-like flow chart of our article selection process (Moher et al., 2015). We obtained studies from three sources. First, on the 15th of March 2018 and on the 4th of July 2018 we identified $i = 1202$ relevant published articles across four (meta) hosts of relevant publication databases (*PubMed*, *Web of Science* (former *ISI web of knowledge*), *ProQuest* (hosting PsycINFO, and PsycARTICLES) and *Ebsco* (PSYNDEX)). We used the following keywords: "('gaze cueing') OR ('gaze cuing') OR ('face perception' AND (cueing OR cuing) AND attention) OR ('gaze perception' AND (cueing OR cuing) AND attention) OR ('posner cueing' AND attention AND (gaze OR face OR social OR joint)) OR ('spatial cueing' AND attention AND (gaze OR face OR social OR joint)) OR (attention AND (gaze OR face) AND (orienting OR orientation OR following) and reflexive)". Second, from the data base results we contacted the 10 most often occurring authors[1] and asked for unpublished data (i = 3 additional contributions). Third, we included unpublished data conducted in the work group of the last author (i = 2 additional contributions).

We trained two raters on the first 70 studies and achieved high inter-rater agreement on 50 studies for exclusion reason ($\kappa = 0.82$), as well as on the exclusion decision ($\kappa = 0.81$, Fleiss, 1971). We kept articles with divergent exclusion decisions. Remaining studies were rated only by one of the two raters. We excluded (1) animal studies ($i = 17$), (2) infant, child and adolescent studies (mean age of sample below 18 years, $i = 52$), (3) studies with a clinical sample ($i = 69$), (4) non-experimental studies ($i = 36$) and (5) studies with a clear lack of gaze cueing ($i = 226$).

Overall, the raters screened title and abstracts of $i = 699$ articles. In the end we identified $i = 299$ (42.78%) studies to be potentially relevant. On top we had access to $i = 5$ unpublished studies.

For eligibility one rater inspected the full text of the articles. Not eligible for our analysis were: (1) Studies that matched any exclusion criteria ($i = 57$). (2) Observations not coming from a classical gaze cueing task, which we defined as a single, static, upright and complete human face as a central, directional and non-informative cue followed by a single horizontally-peripheral visual target ($i = 104$). (3) Participants tasks was either to detect, localize or identify the target (not eligible: $i = 41$). (4) We did not consider observations of reactions to targets other than manual key responses ($i = 14$). (6) Manipulated participants e.g., due to emotion inducing ($i = 15$) as well as (7) manipulated stimuli, e.g., due to inducing trustworthiness of the faces ($i = 21$) were also not eligible. On top, we identified article duplicates ($i = 15$).

---

[1] Authors in descending order: Steven Tipper, Andrew P. Bayliss, Alan Kingstone, Jari Hietanen, Nathalie George, Chris Friesen, Paola Ricciardelli, Gustav Kuhn, Stephen Langton and Jelena Ristic

All in all, we found $i = 75$ eligible articles. Due to missing information (e.g., because of untraceable raw reaction time means) we were able to include $i = 58$ articles into our meta analysis. As articles reported multiple gaze cueing effects we collected $j = 83$ independent samples and $k = 211$ observations of the gaze cueing effect. Overall our meta-analysis covered $l = 2577$ unique participants.

4.2.2 Moderator variables

We extracted six a-priori specified moderators. From the included studies we documented (1) the task type, (2) the stimuli used to induce the gaze cueing effect, whether it was a cartoon face or photography of a face, (3) the temporal overlap between cue and target onset, (4) the cue type, (5) the emotion of the facial cue and, (6) stimulus onset asynchrony for each given experiment.

One of three different reaction time tasks is typically used to measure the gaze cueing effect. These tasks require different decisions as fast as possible from the participants by keypress. Participants execute either a detection, localisation, or identification task in a given experiment. In detection and localisation tasks, participants react only to a single type of target (see 1.2A). In detection tasks, participants indicate by key press the mere appearance of the target. In contrast, in the localisation task, participants respond to the target's location by one of two keys (e.g., left or right arrow key). In identification tasks, participants report the identity of the target (e.g., a common choice is the identification of E and F, see 1.2B). The stimuli used to elicit gaze cueing were categorized into one of two categories: cartoon face or photography of a face. The temporal overlap between cue and target indicates whether cue offset was after target onset. Without temporal overlap, the cue disappeared before target onset. This was documented when the entire face disappeared, the gaze became straight again, or the cueing eyes disappeared within the face (such disappearing pupil was sometimes used with cartoon faces, e.g., in Gayzur et al., 2013). With temporal overlap, the cue remained after target onset. We also coded two categories of cue types. One category included experiments where the directional cues were eye cues with a forward-facing head. The other category included eye cues with the head congruently aligned. The emotion types were adopted by the category of the original authors. The categories included “none/neutral”, “happy”, “angry”, “fearful”, “disgust”, and “surprised”. Consequently, some “none/neutral” labeled stimuli might also qualify for “happy” faces. Especially when studies were not investigating the emotional influence on gaze cueing, they might have used a (slightly) smiling face as an emotionally “neutral” face. As a continuous moderator, the last documented experimental condition was the SOA in milliseconds (ms).

Additionally, we collected study and sample characteristics from each included study. Study characteristics include the publication year. Sample characteristics included the number of trials used, the mean age, and the proportion of females in the sample.

4.2.3 Statistical analysis

The effect size of interest i.e., the gaze cueing effect, is the raw mean difference (RMD) of the mean reaction times of participants μ from an observation k to uncued and cued trials: $\mu_k = \mu_{k_{uncued}} - \mu_{k_{cued}}$ in milliseconds (ms). A positive μ_k indicates on average faster reaction times for cued trials. Although often not applicable, *RMD*s are the preferred way to calculate and report meta-analysic effects, e.g., due to intuitive interpretation of the results (Bond, Wiitala, & Richard, 2003).

Most studies included in our meta-analysis reported multiple gaze cueing effects. To account for the dependencies structure of the data we used a multi-level model with random effects representing a hierarchical structure for observed gaze cueing effects from samples in articles (Moeyaert et al., 2016; Noortgate and Onghena, 2003). Although advanced and robust methods exist to account for dependencies within sampling errors (e.g. Hedges et al., 2010; Noortgate and Onghena, 2003; for an overview see Moeyaert et al., 2016), the preferred way is to model dependencies explicitly (Berkey et al., 1998; Kalaian and Raudenbush, 1996; Raudenbush et al., 1988). This is possible using a variance-covariance matrix representing the sampling errors for each observation k in sample j (Gleser and Olkin, 2009; López-López et al., 2017). Although this process is well documented, often the relevant information are not reported in articles (Cheung, 2019; Gleser and Olkin, 2009).

We calculated the covariances[2] of the sampling errors of observations k from the same sample j. Additionally a meta-analysis crucially depends on correct estimation of the sampling

---

[2] with: $COV_{k_1,k_2} = \sigma_{k_1} \times \sigma_{k_2} \times r_{k_1,k_2}$, where r_{k_1,k_2} is the correlation between reaction times to different manipulations of a sample with σ_k’s standard deviations

variance (Moeyaert et al., 2016). Since the gaze cueing effect is a difference score, we calculated the sampling variance[3] V_k for the observation k. With the correlations between observations of a sample k (r_{k_1,k_2}) and the correlations between cued and uncued reaction times from a sample ($r_{k_{uncued,cued}}$) we were able to perform our analysis with the full variance-covariance matrix. We had to contact all corresponding authors to receive the relevant information.

In the end, we had access to the relevant information of a total of $l = 1063$ (i.e., 41.25%) unique participants from $k = 88$ (i.e., 41.71%) samples distributed over $j = 36$ (i.e., 43.37%) experiments from $i = 26$ (i.e., 44.83%) articles. From these samples we calculated the full variance-covariance matrices. For samples missing the relevant information we took the mean of the available correlations, respectively, namely $\bar{r}_{k_{uncued,cued}} = 0.92$ and $\bar{r}_{k_1,k_2} = 0.10$ by Fisher-z and back transforming to calculate the variance-covariance matrix with an informed estimate.

To estimate similarity from experiments (within-cluster) from to the same articles (between-cluster) in the multi-level structure we calculated the intraclass correlation ρ as the proportion of variance at the article level (Nakagawa et al., 2017).

As in most meta-analyses, the vast majority of accessible observations comes from published articles. This is problematic because a special kind of observations, e.g., studies with significant findings, are more likely to get published (Rothstein et al., 2005). As a consequence gaze cueing is likely overestimated in the published literature. Additionally, such a bias hurts the assumptions for random effects meta-analysis that observations are a random selection of an infinite pool of observations (Hedges and Vevea, 1998). Here, we assess the influence of publication bias by assessing whether the meta-analytic results are robust against the precision of the estimate. We followed a common choice and included the standard error (SE) as a moderator in our meta-analytic model to test this robustness (known as Egger's Test; Egger et al., 1997). Additionally this model also allows us to cautiously interpret the intercept as *corrected* for publication bias by extrapolating to perfect precision (i.e., SE = 0, a logical extension of methods known as PET/PEESE, Stanley and Doucouliagos, 2013).

For the moderators we conducted separate mixed-effects multi-level meta regressions

---

[3] with: $\sigma_k^2 = (\frac{\sigma_{k_{uncued}}}{N_k})^2 + (\frac{\sigma_{k_{cued}}}{N_k})^2 - (2 \times r_{k_{uncued,cued}} \times \frac{\sigma_{k_{uncued}}}{\sqrt{N_k}} \times \frac{\sigma_{k_{cued}}}{\sqrt{N_k}})$, where $r_{k_{uncued,cued}}$ is the correlation between uncued and cued trials within one manipulation of a sample with $\sigma_{k_{uncued}}$ and $\sigma_{k_{cued}}$ standard deviations and a sample size of N_k. Note, that we again assume dependency, now between $\mathrm{rt}_{\mathrm{uncued}}$ and $\mathrm{rt}_{\mathrm{cued}}$ for subject l.

to assess their impact on the gaze cueing effect. We included dummy coded categorical moderators in the regression model. To test whether predictors have an influence on gaze cueing we calculate a Wald-type χ^2 test of the according model coefficients, also known as omnibus test. A significant test indicates deviation from the intercept for at least one of the coefficients. These results were followed-up with pairwise comparisons. As a goodness-of-fit measure we used (pseudo) R^2 as improvement from null (i.e., a random effects intercept-only) model to the fitted model (Cox, 1989; Raudenbush, 2009)[4].

For all analyses we report significance of the model coefficients at a conventional level of $\alpha = .05$. Furthermore, we report predicted gaze cueing effect with a 95% confidence interval (95% CI) for all coefficients. For SOA as a sole continuous predictor we report predictions for a common short and a common long SOA instead of reporting the negligible intercept (i.e., at 0 ms). All analyses reported are for random effects models using restricted maximum-likelihood estimation and the Knapp and Hartung (2003) adjustment of standard errors. We used *R* (version 3.6; R Core Team, 2020) and the package *metafor* (Viechtbauer, 2010) for all meta-analyses. Pairwise comparisons were conducted using the package *multcomp* (Hothorn, Bretz, & Westfall, 2008). For data processing we mainly used *tidyverse* packages (Wickham et al., 2019). Data and analysis scripts are available at: `https://osf.io/mbrfp/`

4.3 Results

Finally, our meta-analysis covered $i = 58$ studies reporting gaze cueing effects for $j = 83$ samples and $k = 211$ observations from the samples.

First, we report the results of the model assessing a publication bias in the extracted studies. Second, we report the results of a multi-level random effects intercept-only model for gaze cueing effects, assessing the average true gaze cueing effect estimated from our data. Third, we report moderator assessments from multi-level mixed effect meta-regressions.

---

[4] $pseudo\ R^2 = \frac{\tau_I^2 - \tau^2}{\tau_I^2}$, where τ_I^2 is the residual variance of the according random effects model i.e., the intercept only model using the same data.

4.3.1 Publication bias

Our model assessing the publication bias in the given studies does not indicate a relationship between the standard error of a study and the gaze cueing effect ($t(191) = 0.39$, $p = .697$, see 4.2). Additionally, the models intercept can cautiously be seen as an extrapolation to a observation with SE $=$ 0 from an infinite sample size. The intercept is estimated to be different from 0 ($t(191) = 4.82$, $p < .001$). Thus, we find a gaze cueing effect after *correction* for publication bias of 11.86 ms (95% *CI* [7.00 ms, 16.72 ms]).

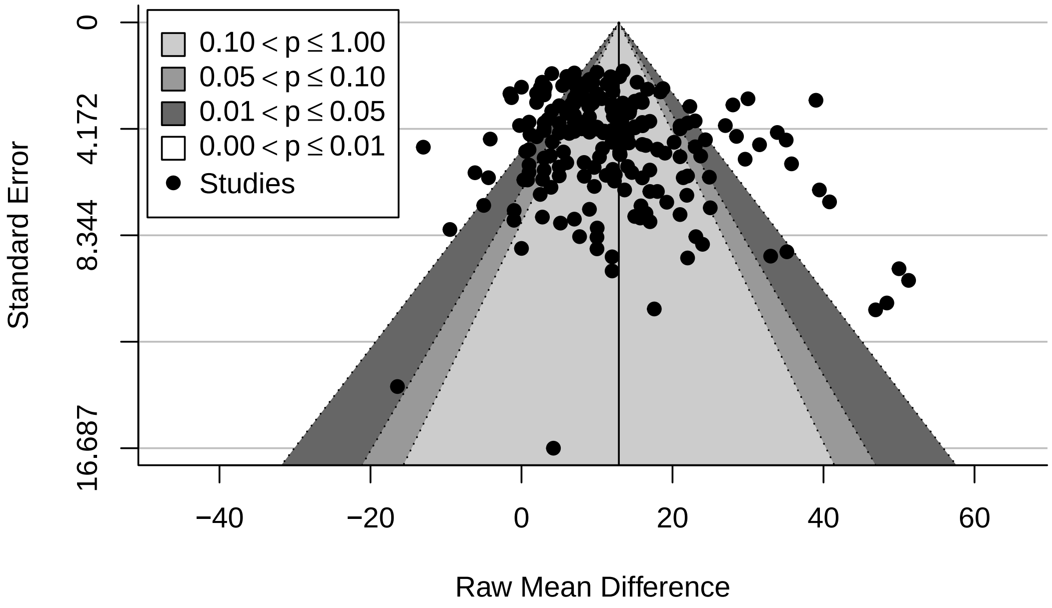

Figure 4.2: Funnel plot of the relationship between expected standard error and observed RMD. Note that visual interpretation of the funnel plot is impeded by clusters of points that indicate the multi-level structure of the data. However, such patterns also serve as indications of publication biases.

4.3.2 Analysis of the gaze cueing effect

We fit a multi-level random effects model to test sampling variability in the observed gaze cueing effects. We allowed for random differences between estimated true effects of articles for unconditional inference about the potential population of articles from which we assume the given studies to be a random sample. Our model showed faster reaction times to cued as compared to uncued trials. It estimates an average true gaze cueing effect of $M = b_0 = 12.89$ ms, 95% *CI* [10.25 ms, 15.54 ms], $t(210) = 9.59$, $p < .001$. The Q-test for heterogeneity indicates non-equal residual variances across observations, $Q(210) = 1{,}092.56$, $p < .001$. Overall, $I^2 = 86.79\%$ of the variation of the effect sizes indicates a large heterogeneity of the effect sizes relative to sampling error, which indicates an almost three times larger variance among the effect sizes than the reported sampling variance. The largest part of this residual variance can be accounted to the article as indicated by an high intraclass correlation of $\rho = 0.72$. This can be interpreted as high similarity in the data between gaze cueing effects that stem from to the same article.

4.3.3 Moderator analysis

With the mixed effect meta-regressions, we test experimental manipulations of the gaze cueing effect. Each article uses a nearly unique set of manipulations. We tried to include as many observations as possible. However, sometimes information was missing, or the manipulation was not done with the given sample. Therefore the number of observations k in our meta-regression models varies between analysis.

Task types

Our first meta-regression model included the task as a predictor for gaze cueing. Our model ($k = 202$) suggests no heterogeneity in detection, localization and identification tasks eliciting gaze cueing ($F(2, 199) = 1.40$, $p = .250$, see Table 4.1). For example, the model predicts the average gaze cueing effect for a detection task of $M = 11.59$ ms (95% *CI* [7.45 ms, 15.72 ms]), for a localisation task of $M = 14.48$ ms (95% *CI* [10.95 ms, 18.02 ms]) and, for a identification task of $M = 11.13$ ms (95% *CI* [7.44 ms, 14.81 ms]).

Photograph vs. cartoon stimuli

The model to test whether gaze cueing effects elicited by photographs differ from cartoons did not reveal a significant effect ($F(1, 200) = 0.21$, $p = .645$, see Table 4.2). For comparison, the model predicts a gaze cueing effect for cartoon faces of $M = 13.58$ ms (95% *CI* [8.71 ms,

Table 4.1: Estimated coefficients for the meta-regression model for task type.

Model	k	Coefficients	*beta*	lower *CI*	upper *CI*	*t*	*p*
Task	202						
		Intercept	11.59	7.45	15.72	5.53	$< .001$
	95	T:local	2.90	-1.54	7.33	1.29	.200
	62	T:ident	-0.46	-4.87	3.95	-0.21	.836

Note. The meta-regression is based on dummy contrasts with task type detection as the reference group. T:local: Task type localisation, T:ident: Task type identification.

Table 4.2: Estimated coefficients for the meta-regression model for stimulus type.

Model	k	Coefficients	*beta*	lower *CI*	upper *CI*	*t*	*p*
Stimuli	202						
		Intercept	13.58	8.71	18.46	5.49	$< .001$
	119	S:Photo	-1.35	-7.13	4.42	-0.46	.645

Note. The meta-regression is based on dummy contrasts with cartoon faces as the reference group. S: stimulus type.

18.46 ms]). For trials with photographic faces the average gaze cueing effect is estimated to be $M = 12.23$ ms (95% *CI* [8.95 ms, 15.51 ms]).

Head vs. eye cues

We find a difference in gaze cueing effects for the cue type ($F(1, 200) = 13.52$, $p < .001$, see Table 4.3). Gaze cueing elicited from heads was predicted to be significantly weaker ($t(200) =$ -3.68, $b_1 =$ -7.38, $p < .001$). Gaze cueing elicited by the eyes is predicted to be $M = 13.25$ ms (95% *CI* [10.50 ms, 16.00 ms], $t(200) = 9.51$, $p < .001$). Although weaker, gaze cues elicited from head cues are still significant, but the effect is estimated to be only $M = 5.87$ ms (95% *CI* [1.34 ms, 10.41 ms], $t(200) = 2.55$, $p = .020$).

Emotions

The model assessing the influence of emotions on gaze cueing indicates no differences between emotions ($F(5, 186) = 0.94$, $p = .459$, see Table 4.4). Model predictions of the gaze

Table 4.3: Estimated coefficients for the meta-regression model for cue type.

Model	k	Coefficients	*beta*	lower *CI*	upper *CI*	*t*	*p*
Cue	202						
		Intercept	13.25	10.50	16.00	9.51	$<$ .001
	19	C:Head	-7.38	-11.33	-3.42	-3.68	$<$ .001

Note. The meta-regression is based on dummy contrasts with eye cues as the reference group C: cue type.

Table 4.4: Estimated coefficients for the meta-regression model for emotions.

Model	k	Coefficients	*beta*	lower *CI*	upper *CI*	*t*	*p*
Emotion	192						
		Intercept	12.75	9.77	15.73	8.44	$<$ .001
	11	E:Happy	0.12	-6.49	6.74	0.04	.970
	7	E:Angry	-2.36	-9.17	4.46	-0.68	.496
	10	E:Fearful	-0.09	-6.71	6.52	-0.03	.978
	1	E:Disgust	2.06	-15.09	19.21	0.24	.813
	2	E:Surprised	-1.00	-8.26	6.25	-0.27	.785

Note. The meta-regression is based on dummy contrasts with neutral faces as the reference group E: facial emotion.

cueing effect are for neutral faces M = 12.75 ms (95% *CI* [9.77 ms, 15.73 ms]), happy faces M = 12.88 ms (95% *CI* [6.42 ms, 19.34 ms]), angry faces M = 10.40 ms (95% *CI* [3.74 ms, 17.05 ms]), fearful faces M = 12.66 ms (95% *CI* [6.19 ms, 19.13 ms]), disgusted faces M = 14.82 ms (95% *CI* [-2.22 ms, 31.86 ms]), surprised faces M = 11.75 ms (95% *CI* [4.63 ms, 18.87 ms]).

Temporal properties

For SOA, we found that with longer SOA the gaze cueing effect slightly decreases ($F(1, 203)$ = -3.92, β_{SOA} = -0.004, $p < .001$, see Table 4.5). The model predicts for a SOA of 250 ms a gaze cueing effect of M = 12.92 ms (95% *CI* [10.20 ms, 15.63 ms]) and for a SOA of 1000 ms a gaze cueing effect of M = 9.94 ms (95% *CI* [7.00 ms, 12.88 ms]). As discussed in the literature, we also tested whether a temporal overlap of the face stimulus and target onset had an influence on gaze cueing dependent on the SOA (see Table 4.6). The omnibus

Table 4.5: Estimated coefficients for the meta-regression model for stimulus onset asynchrony.

Model	k	Coefficients	*beta*	lower *CI*	upper *CI*	*t*	*p*
SOA	205						
		Intercept	13.91	11.09	16.73	9.73	$< .001$
		SOA	0.00	-0.01	0.00	-3.92	$< .001$

Note.

test indicates differences in the coefficients ($F(3, 192) = 7.28$, $p < .001$). In fact we found an interaction between SOA and temporal overlap (see Figure 4.3). Specifically, we find a stronger decline of the gaze cueing effect over SOA in reaction times for experiments without cue-target overlap ($t(192)$ = -4.25, β_{SOA} = -0.01, $p < .001$), e.g. from 250 ms SOA M = 13.13 ms (95% *CI* [8.89 ms, 17.37 ms]) to 500 ms M = 10.52 ms (95% *CI* [6.34 ms, 14.70 ms]). In contrast, in experiments with cue-target overlap the gaze cueing effect does not decline over time ($t(192)$ = -1.91, M = 0.00, p = .200), e.g., from 250 ms SOA M = 12.67 ms (95% *CI* [9.42 ms, 15.91 ms]) to 500 ms SOA M = 12.13 ms (95% *CI* [8.91 ms, 15.34 ms]). This model offers interesting and testable predictions: For experiments without cue-target overlap, after 1,507.02 ms of SOA the gaze cueing effect vanishes (For SOA of 1,507.02 ms, the predicted gaze cueing effect is $M = 0$, 95% *CI* [-6.72 ms, 6.72 ms]). In contrast, with cue-target overlap our model predicts the gaze cueing effect to vanish at a SOA of 6,095.72 ms ($M = 0$, 95% *CI* [-13.07 ms, 13.07 ms]). Note, that the predicted slope (i.e., decline) for experiments with cue-target overlap is not significant.

4.3.4 (pseudo) $\mathbf{R^2}$

None of the mixed effect meta-regressions improved the fit of the data compared to an according null model. One exception is the model assessing cue type, which seems to be slightly better than the according null model (R^2 = 0.86%, whereas for all other models R^2 = 0%).

Table 4.6: Estimated coefficients for the meta-regression model for stimulus onset asynchrony times cue-target overlap.

Model	k	Coefficients	*beta*	lower *CI*	upper *CI*	*t*	*p*
SOA $\times O$	196						
		Intercept	15.74	11.12	20.37	6.71	$<$ .001
	137	O:Yes	-2.53	-7.77	2.70	-0.95	.341
		SOA	-0.01	-0.02	-0.01	-4.25	$<$ .001
		O:Yes x SOA	0.01	0.00	0.01	3.06	.002

Note. The meta-regression is based on dummy contrast with no cue-target overlap at stimulus onset asynchrony (SOA) of 0 ms as the reference group. O: cue-target overlap.

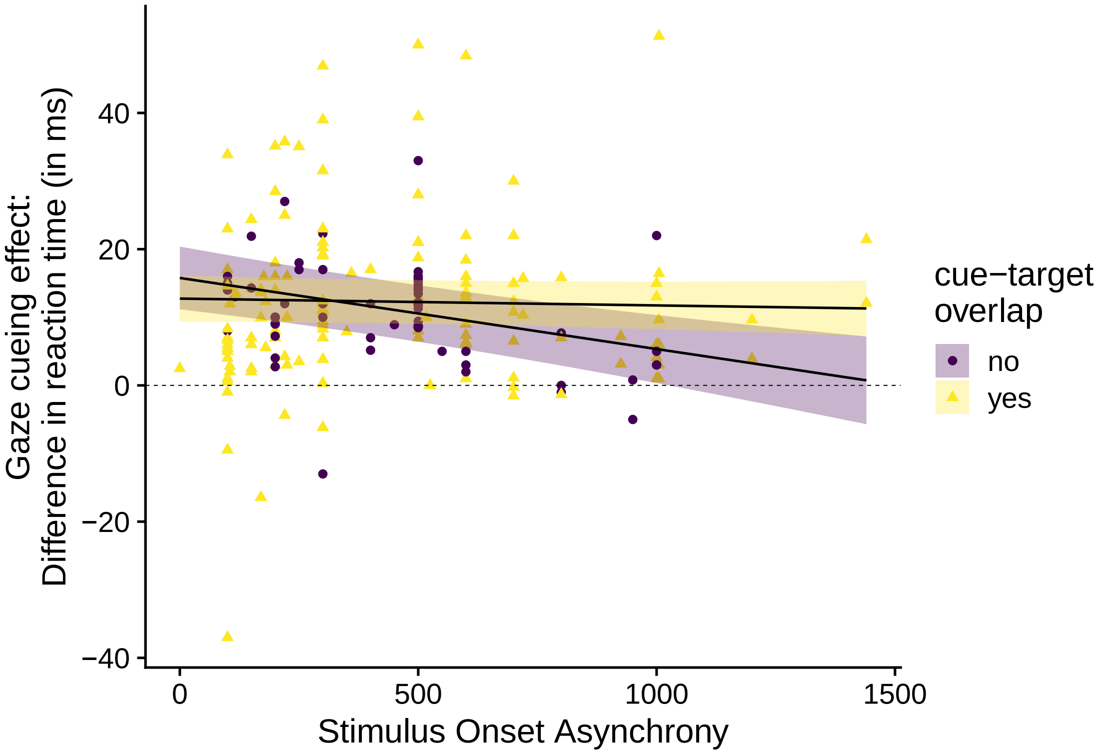

Figure 4.3: Interaction plot for the meta-regression model showing estimated marginal means as a function of stimulus onset asynchrony and cue-target overlap. Error regions around regression slopes depict the 95% confidence intervals at the predicted values.

4.4 Discussion

This meta-analysis used a mixed effect multi-level approach to analyze 58 articles (covering 2577 subjects) conducted on the gaze cueing effect. Our meta-analysis, which is the first on gaze cueing, yields four main findings. First, across all studies, we find a significant average gaze cueing effect. Second, the gaze cueing effect was accompanied by considerable heterogeneity, which experimental moderators could not explain. Third, we find no evidence for specific processing of emotional or different abstraction layers of faces but evidence for the precedence of the eye over head cues. Fourth, as long as the facial stimulus remains, a gaze cueing effect can be observed. In the following, we discuss the implications of these findings.

The average gaze cueing effect that was estimated from the included studies amounted to 13 ms. Thus, studies find overall faster reaction times to cued trials than uncued trials across all experimental manipulations. Interestingly, observations seem to be very similar when they come from the same article. One reason might be that cognitive tasks that are designed to increase within-subject variance perform poorly on intrapersonal reliability (Hedge et al., 2018). Thus participants are not measured reliably, and measures from the same sample do not correlate highly. Additionally, we find considerable heterogeneity which could not be accounted for by any of the included moderators. Such heterogeneity, if unexplained, is cumbersome. Experimental manipulations are typically designed to explain variance. With large heterogeniety psychological effects become very hard to replicate. This is especially true when it falls together with small samples and low power as given in psychological research in general (Stanley et al., 2018), which is also the case for the given set of articles. However, considerable heterogeneity is often associated with psychological phenomena (Erp et al., 2017) and is a general challenge for psychological research (Stanley et al., 2018). Even within carefully conducted so-called direct replications and their original studies, heterogeneity (and non-replicability) can be observed to a non-trivial amount (Eerland et al., 2016; Hagger et al., 2016). Considering that articles included in our meta-analysis do not aim to be replications, not even conceptual ones, heterogeneity is expected to a certain degree. Still, it remains unsatisfying not to capture the given heterogeneity with the extracted experimental manipulations.

With the given estimate of the gaze cueing effect, we were also able to post-hoc determine the average power across studies. Given our estimate is correct, the included studies have a mean power of 34% to find a significant effect. That aligns with estimations for psychological research. For example, Bakker et al. (2012) estimated an average power of 35% for psychological studies, and Stanley et al. (2018) estimated 36%. From such underpowered studies an inconsistent body of literature might evolve (Maxwell, 2004). The inconsistent results for specific moderators (e.g., emotions and SOA) in the primary literature might be a direct consequence of the low-powered studies. In fact, these inconsistencies were the prime motivation for us to conduct a meta-analysis. Therefore, we examined the moderating effects of several experimental manipulations. On the one hand, we investigated research questions that authors ignored in original studies (i.e., stimulus type). On the other hand, we investigate experimental moderators, for which the evidence is relatively weak (i.e., emotions) or with competing results (i.e., cues). A meta-analytic perspective allows us to address new research questions with the given literature to gain new insights. Overall, we were surprised by the considerable heterogeneity in the effect sizes and the incapability of experimental factors to explain it.

Research questions regarding the type of facial stimuli were largely neglected in gaze cueing studies (see Hietanen and Leppänen, 2003, for an exception). Authors did not empirically manipulate whether participants got cued by a cartoon or a more naturalistic photographic face. Theoretically, it is interesting whether top-down features are relevant for gaze cueing. For example, the sclera of the human eye is not only relevant for social cognition in general (Kobayashi and Kohshima, 2001) to be necessary for successful gaze cues, as gaze cueing seems to fail with inverted eyes (Ricciardelli et al., 2000). From that perspective, it was argued that cartoon faces might facilitate gaze cueing as the contrast between sclera and pupil is higher (Hietanen and Leppänen, 2003). Furthermore, the comparison can indicate whether directional facial information is processed holistically or feature-based (Prazak and Burgund, 2014). We found that this manipulation does not alter the gaze cueing effect: Whether gaze cueing was elicited by a cartoon face or a rich photograph had no effect on the reaction time. Thus, we find cumulative evidence to support the view that directional face cues are more holistically processed. Additionally, this finding is in line with previous results indicating that top-down processes can alter the processing of ambiguous stimuli just as well. When schematic faces can elicit the gaze cueing effect, it might be according to the higher-order interpretation. Evidence for this idea comes from a study with ambiguous stimuli which could represent a face with a hat and a car at the same time (Ristic and Kingstone, 2005). The authors showed that the ambiguous stimuli evoked a gaze cueing effect when framed as a face but not when introduced as a car. This effect was persistent. Once framed as a face, participants were unable to "unsee" those eyes again. The same mechanism could be in charge of cartoon faces, backed up by our data.

We find evidence that eyes elicit a stronger gaze cueing effect than head cues. This supports the claim that eyes take precedence over the head (Jellema et al., 2000; Perrett et al., 1992). Cues from eyes elicit a stronger gaze cueing effect than heads. Still, we find a gaze cueing effect, albeit weaker, for heads. So this can be seen in support of a hierarchy of social directional cues with primary processing of eyes followed by head cues (Perrett et al., 1992). Note, however, that head cues are only feasible with a stimulus with three-dimensional properties, such as photographic stimuli. Thus, gaze cueing studies using a head cue do use photographic stimuli. In contrast, a typical cartoon face has only two-dimensional properties. Consequentially it does not allow to alter the heads' orientation. However, as we discussed above, the kind of facial representation does not affect gaze cueing. Thus, authors might prefer eye cues, not for theoretical reasons, but because cartoon faces are more accessible and flexible (see above). On top of that, this is the only model improving the null model, although model performance increases only marginally.

According to our data, emotions seem to be irrelevant for attentional shifts elicited by gaze cueing. Our finding is in line with previous research, not finding evidence that gaze processing for attention shifts is dependent on facial expression processing (Hietanen and Leppänen, 2003; Holmes et al., 2010). This is also in line with findings on exogenous cueing tasks (Koster et al., 2006) or the dot-probe task (Reutter et al., 2019; Schmukle, 2005). In fact, we see a consistent gaze cueing effect across all tested emotions. Our analysis covered happy, angry, fearful, disgusted, and surprised faces. However, our finding is limited in that we only investigate a general effect of facial expressions on gaze cueing. Other authors have claimed several constraints in response to inconsistent findings in the literature. For example, attentional shifts are discussed to be dependent on specific personality traits (e.g., introversion/extraversion, Ponari et al., 2013; or autistic traits Lassalle and Itier, 2014), on participants' emotional states such as anxiety (e.g., Fox et al., 2007; and Mathews et al., 2003). In addition, other authors argue that gaze and emotions are initially processed separately (Pourtois, 2004) and are only integrated with later cognitive processes (Fichtenholtz et al., 2007). As a consequence, emotion-specific gaze cueing occurs only at long SOAs (Graham et al., 2010). Even others discuss sequence effects of the appearance of directional and emotional expressions (Lassalle and Itier, 2015). Albeit there is even inconsistent evidence across those more fine-grained hypotheses, we could not further investigate these possibilities with the given meta-analysis. Still, the often assumed and supposedly face-valid hypotheses of greater cueing effects for fearful faces do not hold across the included studies. Nevertheless, we provide additional evidence for separate cognitive processes for processing expressions and directional cues from faces (Pourtois, 2004). Future work needs to address whether other moderating factors play a nuanced role in gaze cueing.

Much was discussed over the attentional time flow in gaze cueing studies and how it might be unique for social cues (Frischen et al., 2007b; Frischen and Tipper, 2004). Our data suggest that whether it is a defining factor or not, in the end, it has to take into account whether the (facial) stimulus remains until target presentation. We do not find an IOR for gaze cueing tasks when the face remains, but for a disappearing face, an IOR can be observed around 1000 ms (see Figure 4.3). This is actually in line with classical Posner cueing tasks, which found an IOR for a similar time window (e.g., Samuel and Kat, 2003, report from a graphical meta-analysis an stable IOR between 300 and 1600 ms). In fact, the given data can be extrapolated, so when a face is gone, gaze cueing won't be observed after 1500 ms. In contrast, studies having the face remaining after target onset do find gaze cueing effects at that SOA.

Our meta-analysis comes with limitations. Like most meta-analyses, this meta analysis summarizes the published literature and includes only a few unpublished contributions. Thus, it is vulnerable to research that is less likely to get published, e.g., null findings (Kühberger et al., 2014). These biases can especially hurt a central assumption of random effects models of the included effect sizes being a random sample. Simulation studies show that these biases critically increase Type-I Errors and lead to overestimating the estimated effects (Carter et al., 2019). We do not see any indication of a publication bias. However, with the considerable heterogeneity we find, publication bias diagnostics are known to perform only poorly (Carter et al., 2019; Stanley, 2017). From that perspective, a multi-lab replication study might be advisable (Carter et al., 2019). Second, we could not include all potentially interesting factors in our meta-analysis that gaze cueing studies addressed. Most notably, among the ignored aspects are cultural differences (e.g., Takao et al., 2018), dynamic stimuli (e.g., Bayless et al., 2011), non-social cues (as used by classical Posner cueing tasks, e.g., arrows), and predictive designs (e.g., Bayliss et al., 2006). Additionally, the summary of the evidence necessitated by a meta-analytic approach might not respect some nuanced requirements capturing the complexity of the gaze cueing effect. This might be especially true for the analysis of the temporal properties of the gaze cueing effect. In the end, we applied a linear model, although theoretical considerations do not implicate that an IOR effect probably follows a linear trend. However, our statistical analysis is supported by the absence of descriptive patterns, as can be visually explored in 4.3). In general, direct and conceptual replications are the preferred approach for addressing more nuanced differences between scientific theories (Collaboration, 2015). Although, a single study rarely provides enough evidence to address research questions in a particular domain.

All in all, our meta-analysis shows a significant effect of 13 ms, which is, however, accompanied by considerable heterogeneity. This has severe consequences for detecting a publication bias, which was negative but might have failed due to the considerable heterogeneity of the included observations. It also impedes replication of gaze cueing studies. Furthermore, we find a difference in cue types suggesting dominant processing of eye cues over head cues. In contrast, emotions or the ecology of the stimulus do not affect gaze cueing. In the future, well-powered and (pre)registered reports should further investigate competing and more nuanced hypotheses. With the given analysis, we provide justified estimates for future gaze cueing studies.

Chapter 5

Study 2

Gaze cueing in naturalistic scenes under top-down modulation – effects on gaze behavior and memory performance

This chapter also appeared largely in Großekathöfer et al. (2020).

5.1 Brief summary

Humans as social beings rely on information provided by conspecifics. One important signal in social communication is eye gaze. The current study ($n = 93$) sought to replicate and extend previous findings of attentional guidance by eye gaze in complex everyday scenes. In line with previous studies, longer, more and earlier fixations for objects cued by gaze compared to objects that were not cued were observed in free viewing conditions. To investigate how robust this prioritization is against top-down modulation, half of the participants received a memory task that required scanning the whole scene instead of exclusively focusing on cued objects. Interestingly, similar gaze cueing effects occurred in this group. Moreover, the human beings depicted in the scene received a large amount of attention, especially during early phases of visual attention, even though they were irrelevant to the current task. These

results indicate that the mere presence of other human beings as well as their gaze orientation have a strong impact on attentional exploration.

5.2 Methods

5.2.1 Participants

The cueing effects in fixations and saccades that were obtained by Zwickel and Võ (2010) can be considered large (Cohen's $d_z > 0.70$). However, since effects of the top-down modulation implemented in the current study might be smaller, we used a medium effect size for estimating the current sample size. When assuming an effect size of Cohen's $f = 0.25$ at an α level of .05 and a moderate correlation of .40 between factor levels of the within-subjects manipulation object role (cued vs. uncued), a sample size of 66 participants is needed to reveal main effects of the object role or interaction effects between group and object role at a power of .95. Under such conditions, the power for detecting main effects of group is smaller ($1\text{-}\beta = .67$). As a compromise, we aimed at examining 90 participants (plus eventual dropouts) to achieve a power of .80 for the main effect of group and $> .95$ for main and interaction effects involving the within-subjects manipulation object role.

Finally, 94 subjects participated voluntarily. All participants had normal or corrected-to-normal vision and were recruited via the University of Würzburg's online subject pool or by blackboard. Participants received course credit or a financial compensation of 5€. All participants gave written informed consent. One participant was excluded due to problems with the eye-tracking data acquisition, resulting in a final sample of $n = 93$ for the analysis with 64 female and 29 male participants between 18 and 55 years ($M = 24.75$ years, $SD = 5.06$ years). Overall, participants scored very low for autism traits in the Autism-Spectrum Quotient scale (AQ-k, German version, Freitag et al., 2007, Range 0 to 23, $M = 5.75$, $SD = 3.69$). In the final sample, one participant had an overall score higher than 17 which might reflect the presence of an autistic disorder. However, since we did not specify an exclusion criterion regarding AQ-k values beforehand, we decided to keep this participant in the sample.

5.2.2 Stimuli and apparatus

The experimental stimuli consisted of 26 different indoor and outdoor scenes. In each scene, a single individual was looking at one of two objects that were placed within reaching distance. Thus, there was a total of 52 different objects across all scenes (see online supplement S1 for a complete list of all objects). The direction of the gaze (left/right) and the placement of the objects (object A and B left/right) were balanced by taking 4 photographs of each scene (see Figure 5.1 for an outdoor example). Similar to Zwickel and Võ (2010), we did not restrict the position of the individual in the photograph (i.e., the person could appear in the center or more peripherally) such that participants could not expect a specific spatial structure of the scene and the gaze cue. This created 104 unique naturalistic pictures in total. For each participant, a set was randomly taken from this pool containing one version of each scene, resulting in 26 trials. The number of stimuli with leftward and rightward gaze of the depicted person, respectively, was balanced within each participant. Eye movements were tracked with the corneal reflection method and were recorded with an EyeLink 1000plus system (tower mount) at a sampling rate of 1000 Hz. The stimulation was controlled via Presentation® (Neurobehavioral Systems). All stimuli had a resolution of 1280 x 960 pixels and were displayed on a 24" LG 24MB65PY-B screen (resolution: 1920 x 1200 pixels, display size: 516.9 x 323.1 mm) with a refresh rate of 60 Hz. The viewing distance amounted to 50 cm thus resulting in a visual angle of 38.03° x 28.99° for the photographs.

5.2.3 Design and procedure

The experimental design was a 2 x 2 mixed design. First, as a two-level between-subject factor each participant was either assigned to the free viewing or the explicit encoding group (instruction group). Additionally, as a two-level within subject factor object role was manipulated, with objects being cued or uncued by the depicted individual in the scene.

After arriving at the laboratory individually, participants were asked to give full informed consent. Then the eye-tracker was calibrated for each participant using a 9-point grid. According to the manipulation, half of the participants were told that there was a follow-up memory test for objects that were part of the depicted scenes. All participants were then told to look at the following scenes freely without specifying further exploration goals or

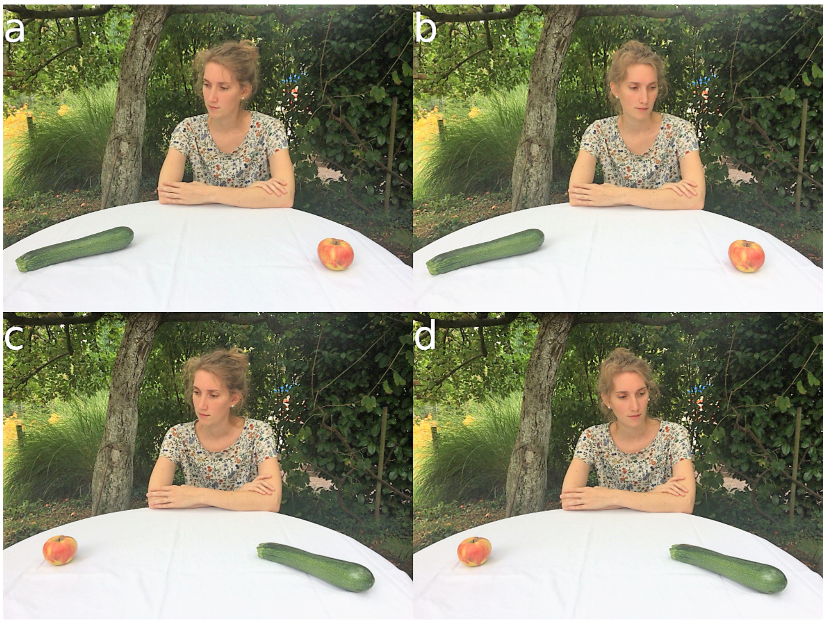

Figure 5.1: Example photographs of a single scene. Gaze direction and objects were balanced over participants. In total 104 photographs of 26 scenes were used. Please note that since we did not obtain permission for publishing the original stimuli, this image shows an example that was not used in the experiment but taken post-hoc in order to illustrate the generation of the stimulus set.

mentioning the content of the scenes. The presentation order of the pictures was randomized. Each trial started with the presentation of a fixation cross for one second, followed by the scene for 10 seconds. This interval was chosen based on our previous studies on social attention (End and Gamer, 2017; Flechsenhar and Gamer, 2017) and was slightly longer than the interval (7 s) that was used by Zwickel and Võ (2010). The inter-trial interval varied randomly between 1 and 3 seconds. After the last trial, participants filled in demographic questionnaires and completed the AQ-k. These questionnaires were used for characterizing the current sample of participants, but they were also introduced to reduce recency effects in the memory task that was accomplished afterwards. It took approximately 5-10 minutes to complete the questionnaires. Participants then were asked to recall as many objects from the scenes as possible and write them down on a blank sheet of paper. No time limit was given but after 10 minutes, the experimenter asked participants to come to an end. In fact, most participants stopped earlier and indicated that they did not recall further objects. Finally, participants received course credit or payment and were debriefed.

5.2.4 Data analysis

For data processing and statistical analysis, the open-source statistical programming language *R* (R Core Team, 2019) was used with the packages *tidyverse* (Wickham, 2017), *knitr* (Xie, 2015) and *papaja* (Aust & Barth, 2018) for reproducible reporting. All analysis and data is available at `https://osf.io/jk9s4/`. For the analysis of the eye-tracking data, EyeLink's standard configuration was used to parse eye movements into saccades and fixations. Saccades were defined as eye movements exceeding a velocity threshold of 30 °/s or an acceleration threshold of 8.000 $°/s^2$. Fixations were defined as time periods between saccades.

We determined the following regions of interest (ROI) by color coding respective images regions by hand using GIMP (GNU Image Manipulation Program): the cued object (average relative size on image: $M_{size} = 1.90\%$, $SD_{size} = 1.95\%$, average visual degree on image: $M_{dg} = 5.23$, $SD_{dg} = 2.90$), the uncued object ($M_{size} = 1.92\%$, $SD_{size} = 2.02\%$, $M_{dg} = 5.19$, $SD_{dg} = 2.85$), the head ($M_{size} = 2.01\%$, $SD_{size} = 1.50\%$, $M_{dg} = 5.74$, $SD_{dg} = 1.91$) and the body ($M_{size} = 7.76\%$, $SD_{size} = 4.11\%$, $M_{dg} = 12.55$, $SD_{dg} = 4.23$) of the depicted person.

Gaze variables of interest were calculated in a largely similar fashion as in Zwickel and Võ (2010). Specifically, we determined the cumulative duration and number of fixations on each ROI per trial. These values were divided by the total time or number of fixations, respectively, to yield proportions. As an additional measure of prioritization, particularly for early attentional allocation, we determined the latency of the first fixation that was directed towards each ROI. These measures allow for effective comparisons of prioritization between the two relevant objects and between the head and the body. To reveal direct relations between the head and the relevant objects, we calculated the proportion of saccades that left the head region of the depicted individual and landed on the cued and uncued objects, respectively. In order to analyze the influence of the experimental manipulations on the eye-tracking data, we carried out separate analyses of variance (ANOVAs) including the between-subject factor instruction group and the within-subject factor object role. ANOVAs were conducted on the dependent variables fixation latency and proportion of saccades from the head towards the object. To examine general effects of social attention, a separate ANOVA with the between-subject factor instruction group and the within-subject factor ROI (head vs. body region) was conducted on fixation latency.

Fixation durations and numbers of fixations were analyzed in more detail by additionally considering the temporal progression of effects. To this aim, we calculated relative fixation durations as well as relative numbers of fixations on each ROI for 5 time bins of 2 s each spanning the whole viewing duration. These data were analyzed using separate ANOVAs on relative fixation durations and numbers, respectively. The first analyses focused on the temporal progression of cueing effects and included the between-subject factor instruction group and the within-subject factors object role and time point. Subsequent analyses on general effects of social attention included the between-subject factor instruction group and the within-subject factors ROI (head vs. body region) and time point. In case of significant interaction effects, we calculated contrasts using *emmeans* (Lenth, 2019) as post-hoc tests with p values adjusted according to Tukey's honest significant difference method.

The memory test was scored manually by comparing the list of recalled objects to the objects that appeared in the scenes. We separately scored whether cued or uncued objects were recalled and ignored any other reported details. Afterwards, we calculated the sum of recalled objects separately for cued and uncued details. These data were analyzed using an ANOVA including the between-subject factor instruction group and the within-subject factor object role. To further assess the influence of visual attention on memory, we used a generalized linear mixed model (GLMM) approach implemented via *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Based on the ANOVA results (see below), we used a sequential model building strategy starting with model 1 including only instruction group as the main predictor of subsequent recall performance. In the second step we added z-standardized relative fixation duration in model 2a and analogously, the z-standardized relative number of fixations in model 2b. In the third step we added object role and corresponding interaction terms with the other factors to the previous models. We always tested the higher-order model against its lower-order counterpart using an ANOVA approach to examine if relative fixation duration and/or relative number of fixations had incremental value beyond group membership or interacted with object role in predicting recall performance.

For all analyses the a priori significance level was set to $\alpha = .05$. ANOVAs were computed with the package *afex* (Singmann, Bolker, Westfall, & Aust, 2019). As effect sizes, generalized eta-square ($\hat{\eta}^2_G$) values are reported, where guidelines suggest .26 as a large, .13 as a medium and .02 as a small effect (Bakeman, 2005). Greenhouse-Geisser correction was applied in all repeated-measures ANOVAs containing more than one degree of freedom in the numerator to account for potential violations of the sphericity assumption (Greenhouse and Geisser, 1959).

5.3 Results

5.3.1 Gaze following

A significant main effect of object role in the analysis of fixation latencies indicates earlier fixations on cued compared to uncued objects ($F(1, 91) = 54.84$, $p < .001$, $\hat{\eta}^2_G = .171$; $M_{lat,cued} = 1852ms$, $M_{lat,uncued} = 2413ms$). The main effect of instruction group was also significant, with earlier fixations on both objects in the explicit encoding ($F(1, 91) = 34.91$, $p < .001$, $\hat{\eta}^2_G = .202$; $M_{lat,mem} = 1821ms$) compared to the free viewing group ($M_{lat,free} = 2444ms$). The interaction effect failed to reach statistical significance ($F(1, 91) = 3.48$, $p = .065$, $\hat{\eta}^2_G = .013$; see Figure 5.2 A).

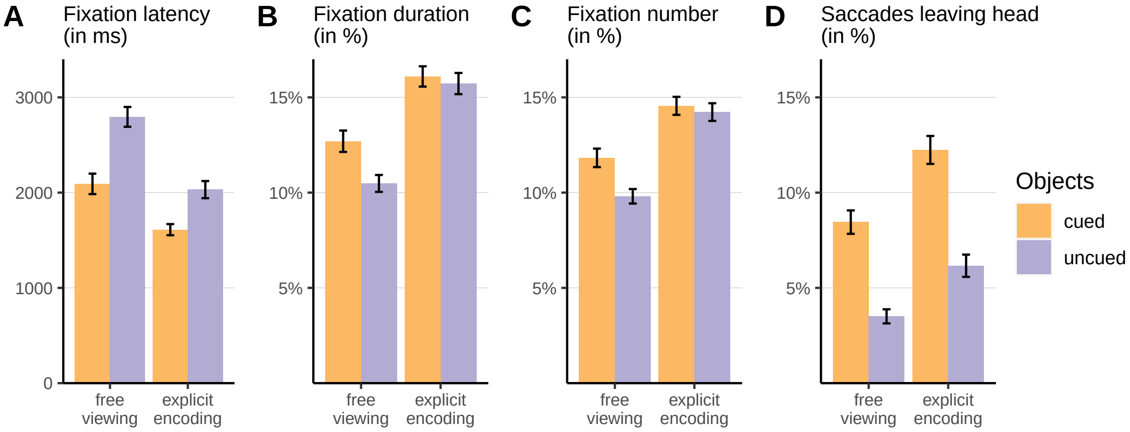

Figure 5.2: Bar plots of the different prioritization measures for the attentional orienting towards the cued and uncued objects as a function of instruction group. Note that data were aggregated across time bins for fixation duration and numbers. Error bars represent standard errors of the mean.

Comparable effects were obtained for saccades leaving the head which were more likely to land on the cued compared to the uncued object as confirmed by a significant main effect of object role, ($F(1, 91) = 43.17$, $p < .001$, $\hat{\eta}^2_G = .160$; $M_{sac,cued} = 10.35\%$, $M_{sac,uncued} = 4.83\%$). The main effect for group showed that saccades of participants in the explicit encoding group were more often directed towards any of the objects as compared to the free viewing group ($F(1, 91) = 24.71$, $p < .001$, $\hat{\eta}^2_G = .140$; $M_{sac,free} = 5.98\%$, $M_{sac,mem} = 9.20\%$). Again, the interaction effect of instruction group and object role failed to reach statistical significance ($F(1, 91) = 1.14$, $p = .288$, $\hat{\eta}^2_G = .005$; see Figure 5.2 B).

The time course of fixation durations and numbers on all four ROIs is depicted in Figure 5.3. In the corresponding ANOVA focusing on the temporal progression of gaze cueing effects, we obtained significant main effects of object role, indicating that participants fixated the cued object longer ($F(1, 91) = 5.60$, $p = .020$, $\hat{\eta}^2_G = .008$; $M_{dur,cued} = 14.40\%$, $M_{dur,uncued} = 13.10\%$), and more often ($F(1, 91) = 8.42$, $p = .005$, $\hat{\eta}^2_G = .009$; $M_{num,cued} = 13.19\%$, $M_{num,uncued} = 12.02\%$), than the uncued object. Explicit instructions also led to longer ($F(1, 91) = 19.31$, $p < .001$, $\hat{\eta}^2_G = .083$; $M_{dur,mem} = 15.91\%$, $M_{dur,free} = 11.59\%$), and more fixations ($F(1, 91) = 18.36$, $p < .001$, $\hat{\eta}^2_G = .082$; $M_{num,cued} = 14.39\%$, $M_{num,free} = 10.82\%$), on the objects as compared to the free viewing condition (see Figure 5.2 B & C).

The interaction effect of instruction group and object role was only statistically significant for the number of fixations ($F(1, 91) = 4.37$, $p = .039$, $\hat{\eta}^2_G = .005$), but failed statistical significance for the duration of fixations ($F(1, 91) = 2.84$, $p = .096$, $\hat{\eta}^2_G = .004$). However, contrasts of the estimated marginal means for both fixations measures revealed a statistically significant difference for object role only in the free viewing group (duration: , number:), with more and longer fixations on the cued object. In the explicit encoding group, contrasts of object role did not reach statistical significance (both $p > .5$).

The two way interactions between object role and time points were statistically significant for both measures (duration: $F(3.21, 291.82) = 23.93$, $\epsilon = 0.8$, $p < .001$, $\hat{\eta}^2_G = .051$, number: $F(3.11, 283.34) = 26.65$, $\epsilon = 0.78$, $p < .001$, $\hat{\eta}^2_G = .055$). Pairwise contrasts on estimated marginal means revealed that the interaction was mainly driven by more fixations on the cued object than on the uncued object during first and last time point (first: duration: $t(91) = 7.59$, $p < .001$, number: $t(91) = 7.90$, $p < .001$, last time point: duration: $t(91) = 2.99$, $p = .004$, number: $t(91) = 3.43$, $p = .001$, with all other bins $p > .2$, see Figure 5.3).

Additionally, time dependent group differences were revealed by significant interactions between instruction group and time points for both measures (duration: $F(2.91, 265.26) = 4.56$,

$\epsilon = 0.73$, $p = .004$, $\hat{\eta}^2_G = .012$, number: $F(2.90, 264.03) = 2.92$, $\epsilon = 0.73$, $p = .036$, $\hat{\eta}^2_G = .008$). Pairwise contrasts on estimated marginal means indicate significant differences with longer duration and more fixations on both objects for the explicit encoding group during all except the second time point (first: duration: $t(91) = -6.34$, $p < .001$, numbers: $t(91) = -6.04$, $p < .001$, third: duration: $t(91) = -3.20$, $p = .002$, numbers: $t(91) = -3.08$, $p = .003$, fourth: duration: $t(91) = -4.13$, $p < .001$, numbers: $t(91) = -3.74$, $p < .001$, and last time point: duration: $t(91) = -3.75$, $p < .001$, numbers: $t(91) = -3.87$, $p < .001$).

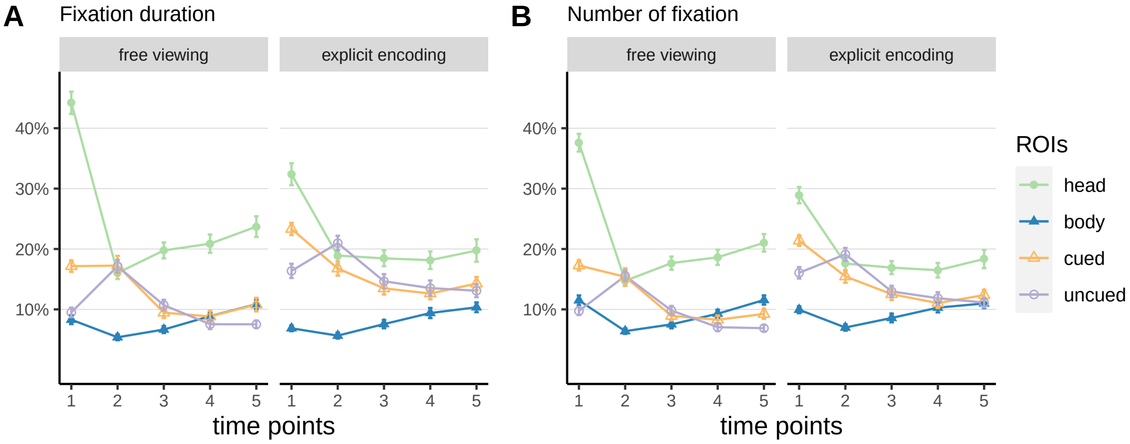

Figure 5.3: Time course of fixation durations and numbers as a function of the region of interest (ROI). Error bars represent standard errors of the mean.

The three-way interaction of instruction group, object role and time point failed to reach statistical significance for fixations durations ($F(3.21, 291.82) = 1.21$, $\epsilon = 0.8$, $p = .308$, $\hat{\eta}^2_G = .003$) and numbers ($F(3.11, 283.34) = 1.37$, $\epsilon = 0.78$, $p = .253$, $\hat{\eta}^2_G = .003$).

5.3.2 Memory for objects

An analysis of the recall data showed, that participants in the explicit encoding group remembered more items than participants from the free viewing group ($F(1, 92) = 33.23$, $p < .001$, $\hat{\eta}^2_G = .234$; $M_{recall,free} = 11.23$, $M_{recall,mem} = 18.72$). Neither the main effect of object role ($F(1, 92) = 0.59$, $p = .444$, $\hat{\eta}^2_G = .001$) nor the interaction effect were statistically significant ($F(1, 92) = 0.02$, $p = .878$, $\hat{\eta}^2_G = .000$; see Figure 5.4).

In order to examine the influence of visual exploration on recall performance, we used a GLMM approach starting with a first model where only group assignment was entered.

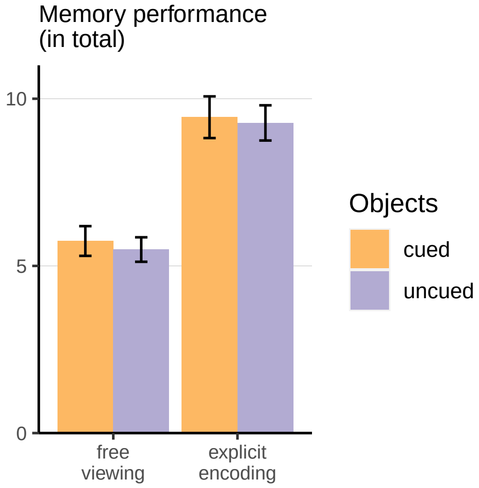

Figure 5.4: Bar plot of the memory performance for the cued and uncued objects as a function for instruction group. Error bars represent standard errors of the mean.

Corresponding to the ANOVA results discussed above, this model revealed a significant effect for group (see Table 5.1 for model parameters and model selection criteria). Next, we built two extended models incorporating measures of visual attention: Model 2a included the main effect of (z-standardized) relative fixation duration and its interaction with group. To Model 2b we added (z-standardized) relative number of fixations and its interaction with group. Surprisingly, model 2a including the main effect of fixation duration and its interaction with group did not yield a better prediction of recall performance in comparison to model 1 ($p = .168$). By contrast, model 2b including the main effect of number of fixations and its interaction with group improved the prediction of recalled stimuli ($\chi^2() = 9.67$, $p < .01$) with a significant weight for the number of fixations. As a last step, we tested whether object role further improves the prediction of recall performance in comparison to model 2b, which was not the case ($p = .924$).

5.3.3 Social prioritization

Fixation latencies differed remarkably between the head and the body (see Figure 5.5 A). Consequently, the ANOVA yielded a significant main effect of ROI, with earlier fixations of the head compared to the body ($F(1, 91) = 216.00$, $p < .001$, $\hat{\eta}^2_G = .585$; $M_{lat,head} = 722ms$, $M_{lat,body} = 2609ms$). There was neither a statistically significant main effect of instruction group ($F(1, 91) = 0.69$, $p = .407$, $\hat{\eta}^2_G = .003$) nor an interaction of both factors

Table 5.1: Parameters and model selection criteria of general linear mixed models predicting object recall from group, number/duration of fixation and object role.

Coefficient	Est.	SE	z	p	AIC	BIC	DIC	df
1					4930	4956	4922	4492
INTERCEPT	-1.46	0.16	-9.16	$< .001$				
GROUP	0.79	0.14	5.45	$< .001$				
2					4931	4969	4919	4490
INTERCEPT	-1.44	0.16	-9.02	$< .001$				
GROUP	0.77	0.14	5.3	$< .001$				
DURATION	0.12	0.06	1.9	.057				
GROUP x DURATION	-0.11	0.08	-1.39	.165				
3					4925	4963	4913	4490
INTERCEPT	-1.43	0.16	-8.94	$< .001$				
GROUP	0.75	0.14	5.19	$< .001$				
NUMBER	0.19	0.06	2.96	$< .01$				
GROUP x NUMBER	-0.12	0.08	-1.6	.11				
4					4932	4996	4912	4486
INTERCEPT	-1.41	0.17	-8.37	$< .001$				
GROUP	0.74	0.16	4.62	$< .001$				
NUMBER	0.19	0.08	2.21	$< .05$				
OBJECT ROLE	-0.04	0.11	-0.41	.683				
GROUP x NUMBER	-0.16	0.1	-1.51	.13				
GROUP x OBJECT ROLE	0.01	0.15	0.07	.943				
NUMBER x OBJECT ROLE	0	0.12	-0.01	.994				
GROUP x NUMBER x OBJECT ROLE	0.08	0.15	0.51	.613				

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; DIC = Deviance information criterion; df = Residual degrees of freedom.

$(F(1, 91) = 1.43$, $p = .235$, $\hat{\eta}^2_G = .009)$.

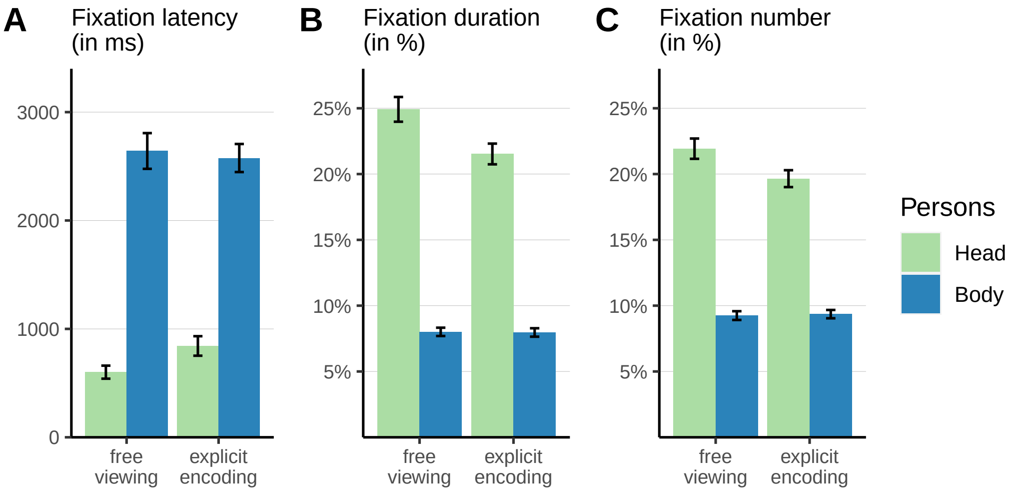

Figure 5.5: Bar plot of the different prioritization measures for attentional orienting towards and visual exploration of the depicted person's head and body as a function of instruction group. Note that data were aggregated across time bins for fixation duration and numbers. Error bars represent standard errors of the mean.

In the detailed ANOVA including time bin (see Figure 5.3), fixation duration and numbers showed a very similar pattern with longer, $(F(1, 91) = 292.51$, $p < .001$, $\hat{\eta}^2_G = .473$; $M_{dur,head} = 23.22\%$, $M_{dur,body} = 7.99\%)$, as well as more fixations $(F(1, 91) = 238.78$, $p < .001$, $\hat{\eta}^2_G = .413$; $M_{num,head} = 20.79\%$, $M_{num,body} = 9.30\%)$, on the head than the body. Remarkably, the instruction group did not exhibit a statistically significant main effect, neither for the fixation duration $(F(1, 91) = 3.25$, $p = .075$, $\hat{\eta}^2_G = .011)$, nor for the number of fixations, $(F(1, 91) = 1.76$, $p = .189$, $\hat{\eta}^2_G = .006)$. Furthermore, the interaction effects of instruction group and ROI failed to reach statistical significance for fixation duration $(F(1, 91) = 3.52$, $p = .064$, $\hat{\eta}^2_G = .011)$ and fixation numbers, $(F(1, 91) = 2.58$, $p = .112$, $\hat{\eta}^2_G = .008$; see Figure 5.5 B & C).

However, the ANOVA yielded a three-way interaction of instruction group, ROI and time point for fixation durations $(F(2.89, 263.12) = 8.26$, $\epsilon = 0.72$, $p < .001$, $\hat{\eta}^2_G = .016$, as well as for numbers of fixations $(F(2.93, 266.90) = 6.11$, $\epsilon = 0.73$, $p = .001$, $\hat{\eta}^2_G = .012)$. To follow-up on this result, we performed separate ANOVAs for each time point including instruction group and ROI as factors. Interestingly, for both measures, we observed a statistically significant interaction between instruction group and ROI only for the first time point (duration: $F(1, 91) = 12.36$, $p = .001$, $\hat{\eta}^2_G = .074$, number: $F(1, 91) = 8.07$,

$p = .006$, $\hat{\eta}^2_G = .053$). Pairwise contrasts of estimated marginal means for this interval revealed a significant difference between both groups for the head region for fixation duration ($t(91) = 4.57$, $p < .001$) as well as numbers of fixations ($t(91) = 4.35$, $p < .001$) with more and longer fixations in the free viewing group (see Figure 5.3). The fixation duration and fixation number for the body region did not differ between groups during the first time point (both $p > .1$). For all other time points follow-up ANOVAs did not yield significant interactions between instruction group and object role, neither for fixation duration nor fixation number (all interactions $p > .19$, for details see the online supplement, Tables S8 - S12).

5.4 Discussion

By using naturalistic scenes with rich detail, this study aimed at conceptually replicating previous findings of a general prioritization of social cues (i.e., heads and bodies, Birmingham et al., 2008a; End and Gamer, 2017; Flechsenhar and Gamer, 2017) as well as previously reported gaze cueing effects elicited by a person being directed towards a specific object in the scene (Zwickel and Võ, 2010). Both effects were replicated.

In detail, heads of persons in the scene were fixated earlier and explored more extensively as compared to body regions (and also more than the cued or uncued objects[1]). Additionally, in line with Zwickel and Võ (2010), cued objects were preferred over uncued ones. They were fixated remarkably earlier, longer and more often. Thus, gaze following effects did not only occur with respect to a more thorough processing overall but were also evident in an early allocation of attentional resources after stimulus onset. Additional support for such early prioritization was revealed in the temporal analysis of attentional exploration. Fixation durations and numbers differed most between the cued and uncued object during the first 2 s of the 10 s viewing period (see Figure 5.3).

Moreover, the prioritization of the head and the preference for the cued object indirectly suggest a link between these two regions. To investigate this relationship in more detail,

---

[1] A direct comparison of all ROIs, e.g., head with cued object, can be found in the supplementary material, Tables S13 – S15 and Figures S1 – S3.

we examined saccades leaving the head towards the cued and uncued object, respectively. Saccades leaving the head were significantly more likely to end on the cued than on the uncued object, directly linking fixations of the head and the cued object. Thereby, current results fully replicate the findings of Zwickel and Võ (2010) with a more naturalistic set of stimuli. As often, by using more naturalistic material, experimental control is reduced (cf. Chapter 9 on the the study of gaze cueing in VR). We tried to minimize unsystematic effects by producing the stimuli in the same way as Zwickel and Võ (2010), but using real as compared to 3D rendered scenes. In particular, each scene was photographed four times with gaze direction and object placement being fully counterbalanced. Since four individual photographs of each scene were taken in the current study, we could not fully control all stimulus aspects. However, the full replication of the effects previously obtained with a different set of virtual scenes indicates that these effects generalize to naturalistic conditions and are stable against small variations in scene layout and presentation.

Besides conceptually replicating previous findings, this study also aimed at extending the line of research by testing the robustness of gaze following against top-down modulations. This was achieved by instructing half of the participants to memorize as many details of the presented scenes as possible. Since the depicted human being was not relevant to this task, we expected a generally reduced attention towards head and body regions as well as a more systematic exploration pattern, potentially reducing gaze cueing effects in fixations on and saccades towards cued objects. Unsurprisingly, the memory task that was accomplished after the eye-tracking experiment showed that participants, who knew about the free recall task in advance performed better in recalling items. Interestingly, the hypothesized enhanced attentional preference for the uncued object in the explicit encoding group was only found for fixation numbers. Against our hypothesis, the effect did not reach statistical significance for fixation latencies and durations (while being descriptively in the hypothesized direction, see Figure
5.2 A - D).

The temporal analysis of attentional allocation furthermore indicates that effects of the instruction were most pronounced during early periods of picture viewing. In the first 2 seconds, fixations on the head differed clearly between instruction groups, with less social prioritization by participants in the explicit encoding group. In the same interval, however,

both groups showed the largest difference in attentional exploration of cued as compared to uncued objects. Overall, the explicit encoding group fixated longer and more often on both objects than the free viewing group but cueing effects were largely unaffected by the explicit task with the only exception of fixation numbers being slightly less biased towards the cued object in the explicit encoding group. Although the time course of attentional exploration (see Figure
5.3) seems to indicate that the encoding instruction induced a more systematic exploration of the objects particularly at early time points, the three-way interaction failed to reach statistical significance in both ANOVAs.

These findings indicate that the prioritization of social information is largely unaffected by a manipulation of goal-driven attention, although early fixations on the head were slightly inhibited in the current study. The attentional guidance of gaze was effective especially in the early phase of stimulus presentation, even when participants investigated the scenes with an explicit (non-social) task goal. In general, this early attentional preference for cued locations provides support for the automaticity and reflexivity of social attentional processes and is in line with previous studies on gaze cueing within highly controlled setups (e.g., Hayward et al., 2017; Ristic and Kingstone, 2005), more naturalistic laboratory studies (e.g., Castelhano et al., 2007; Zwickel and Võ, 2010) and real-life social situations (e.g., Hayward et al., 2017; Richardson et al., 2007). Moreover, the current results are consistent with recent findings of an early attentional bias towards social information in general (End and Gamer, 2017; Rösler et al., 2017) that seems to be relatively resistant against specific task instructions (Flechsenhar and Gamer, 2017).

As expected, participants with specific recall instructions performed better in the subsequent memory task. However, the contribution of the automatic attentional processes to memory encoding remains unclear. In particular, although cued objects were prioritized in the attentional exploration, only the general number of fixations irrespective of object role predicted stimulus recall across both groups (see Table 5.1). Fixation duration did not add incremental value. This is partially in line with studies on eye movements (e.g., Hollingworth and Henderson, 2002) and (non-social) cueing (Belopolsky et al., 2008; Schmidt et al., 2002), which showed that increased attention results in better memory performance. Originally, we additionally expected the cued object to be better recalled than the uncued one (Belopolsky et al., 2008; Schmidt et al., 2002). However, another study showed that if certain scene details have a special meaning (e.g., by being central to the content of a picture story), attention does no longer predict memory for these details (Kim et al., 2013). With respect to the current study, these findings may indicate that both objects that were placed within reaching distance of the depicted person conveyed such meaning and were therefore remembered with equal probability. Since we only tested for early memory effects, it would be very interesting to delay the memory test by at least 24h to examine whether memory consolidation differs between cued and uncued objects (Squire, 1993). Another explanation for the currently observed effects might be that exploration time was sufficient to process both objects equally well. It would thus be very interesting for future studies to manipulate viewing durations and examine the effect of such manipulations on memory performance.

Although the current study has several strengths including the systematic generation of novel stimulus material and the large sample size, it also has some limitations. First, although this study shows that humans follow other persons' gaze implicitly in unconstrained situations, this was shown for situations without real interactions between humans. Research shows that fixation patterns differ remarkably when a real interaction between persons is possible (e.g., Hayward et al., 2017; Laidlaw et al., 2011; for an overview see: Risko et al., 2016). However, our findings add evidence to classic highly controlled laboratory approaches to social attention, yet at the same time better approximates ecological research (Risko et al., 2012). Second, one might criticize that we did not control for directional information from the depicted person's body in contrast to the head. Earlier studies show that body orientation is relevant for cueing (Hietanen, 1999; Lawson and Calder, 2016) and the influence of body orientation on the cueing effects (e.g., through peripheral vision) cannot be dissociated by our study design. However, our results indicate a direct link between the head and the cued object, as do the results of Zwickel and Võ (2010). In fact, overall the first fixation of the body occurred about 800 ms after first fixation on the cued object. Third, we used a rather long viewing time of 10 s. This time allows for a very detailed exploration of the depicted scene and our analyses of the time courses of attentional measures showed that effects of top-down instructions seemed to be more pronounced during the first few seconds and quickly vanished afterwards. Future studies should therefore either use tasks that are cognitively more demanding or focus on a systematic variation of viewing durations to further examine the automaticity of social attention and gaze following.

Overall, the current results provide additional support for previous findings that attention is shifted reflexively to locations where other persons are looking at (e.g., Hayward et al., 2017; Ristic and Kingstone, 2005). This evidence, which was previously extended to free viewing of more complex static scenes by Zwickel and Võ (2010), was shown to be valid in more naturalistic scenes and relatively robust against top-down modulation. Even when explicitly directing attention away from depicted individuals by making objects task-relevant, social and joint attention still occurred, and were even largely comparable to the unbiased free viewing condition. These results indicate that the mere presence of other human beings as well as their gaze orientation have a strong impact on attentional exploration.

5.5 Acknowledgements

The authors thank Michael Strunz for his help with data collection. This work was supported by the European Research Council (ERC-2013-StG #336305).

Chapter 6

Study 3

Reality in a sphere: a direct comparison of social attention in the laboratory and the real world

This chapter also appeared largely in Großekathöfer et al. (in press).

6.1 Brief summary

Humans often show reduced social attention in real situations, a finding rarely replicated in controlled laboratory studies. Virtual reality is supposed to allow for ecologically valid and at the same time highly controlled experiments. This study aimed to provide initial insights into the reliability and validity of using spherical videos viewed via a head-mounted display (HMD) to assess social attention. We chose five public places in the city of Würzburg and measured eye movements of 44 participants for 30 seconds at each location twice: Once in a real environment with mobile eye-tracking glasses and once in a virtual environment playing a spherical video of the location in an HMD with an integrated eye-tracker. As hypothesized, participants demonstrated reduced social attention with less exploration of passengers in the real environment as compared to the virtual one. This is in line with earlier studies

showing social avoidance in interactive situations. Furthermore, we only observed consistent gaze proportions on passengers across locations in virtual environments. These findings highlight that the potential for social interactions and an adherence to social norms are essential modulators of viewing behavior in social situations and cannot be easily simulated in laboratory contexts. However, spherical videos might be helpful for supplementing the range of methods in social cognition research and other fields. Data and analysis scripts are available at `https://osf.io/hktdu/`.

6.2 Methods

Hypotheses, sample size, design specifications, and analysis steps were preregistered before data collection on Aspredicted.org (available at: `https://aspredicted.org/p7a83.pdf`). In our study, we used a fully nested within-subjects design with the factors environment (virtual environment vs. real environment) and region of interest (ROI, person vs. object, see below for further details).

6.2.1 Participants

The sample consisted of 44 participants (33 female; age: M = 22.10 years; SD = 6.00 years) who were recruited via the online participant pool of the University of Würzburg. Students participated for course credit. All participants had normal or corrected-to-normal vision by means of contact lenses. Sample size planning was done using PANGEA (Westfall, 2016) before collecting any data[1]. The planned sample size allows for detecting the anticipated interaction of interest with a medium effect (Cohen's $d = 0.3$) at a conventional level of α = .05 and an adequate power of $1 - \beta = 0.87$.

6.2.2 Stimuli and apparatus

The eye-tracking data were collected for five different locations in Würzburg, Germany. The participants experienced the selected locations in two environments: in the virtual

---

[1] See the preregistration for details.

environment (VE) through watching spherical videos in an HMD and in the real environment (RE) by visiting the location in reality (see Figure 6.1).

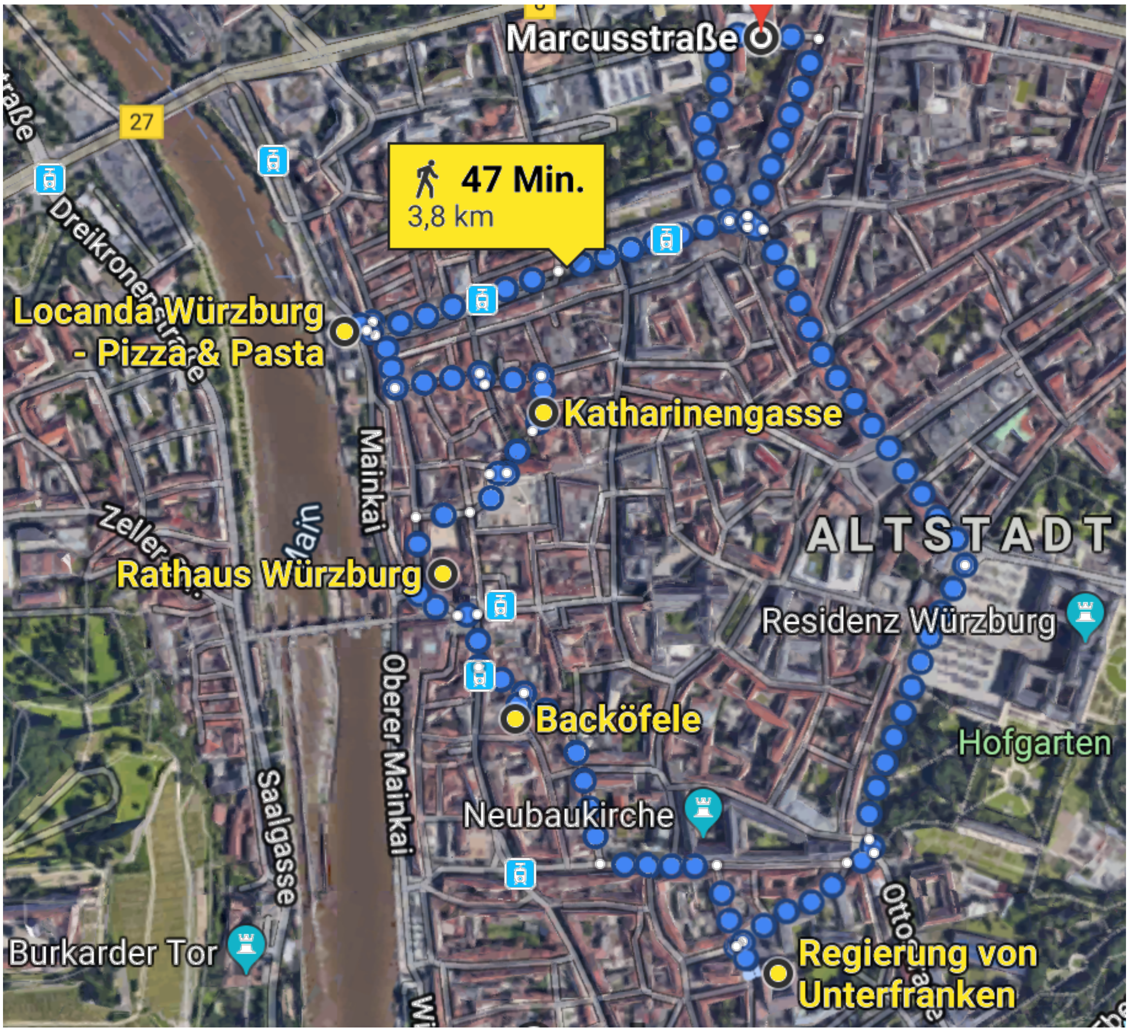

Figure 6.1: Google Maps route the experimenter and participants walked to the five measurement locations (highlighted with yellow dots and corresponding descriptive labels) in the real environment. Spherical videos were recorded at the identical locations for the virtual environment.

Locations

The five locations in the city of Würzburg included places located in rather quite side streets as well as more crowded spots. On average, the number of pedestrians was comparable between the VE ($M = 10.60$, $SD = 6.23$) and the RE ($M = 8.18$, $SD = 6.15$).

Locations were visited in RE at the shortest route to keep walking time minimal. During the experiment, the route was used in two different directions, counterbalanced across participants. The order of locations in both environments was kept identical for each participant, resulting in only two sequences of spherical videos in the VE.

Virtual environment

The stimuli for the VE were spherical videos recorded at the five locations with a GoPro Omni camera mount of six *GoPro HERO4 black* cameras. The six resulting videos were then stitched together into a single spherical video for each location using *Kolor Autopano Pro* (Version 4.2). The final video had a total resolution of 3840 × 1920 pixels with 50 frames per second (FPS) and a duration of 15 seconds[2]. We added an additional seven seconds of black screen (5 s at the start and 2 s at the end) and the audio track of one camera to each video.

Two videos from each location were used for each participant resulting in 30s of spherical videos per location. The videos were projected on a virtual sphere rendered by the 3D game engine Unity (Version 2018.2.18f1) onto an HTC Vive. We used the HMD with HTC Vive's default internal rendering resolution of 3024 × 1680 pixels (or 1512 × 1680 pixels per eye and display) at a refresh rate of 90 FPS. The HTC Vive provides a FOV of approximately 110° × 110° of visual angle at a typical distance of 10 mm from the eyes to the internal displays. Eye-tracking data relative to the FOV were collected with an integrated SensoMotoric Instruments (SMI) binocular eye-tracker and the SMI plug-in for Unity at a sampling rate of 250 Hz.

Real environment

Eye-tracking in the RE was conducted using SMI Eye-Tracking Glasses 2.1 with the iViewETG software at a sampling rate of 60 Hz. The integrated camera recorded the participants' FOV at 30 Hz with a resolution of 1280 × 960 pixels. The FOV amounts to approximately 60° × 46° of visual angle.

6.2.3 Procedure

Upon arrival, participants were welcomed and provided written informed consent. To conceal the aim of the current study and ensure that participants are not concentrating on their own

---

[2] Examples of the spherical videos from the different locations can be watched at `https://www.youtube.com/playlist?list=PLFO679j3PTWpcFPRZ4i75_usOH7UwrQ4b`.

eye movements, the experimenter provided erroneous information that we were interested in examining the suitability of the current devices for measuring pupil width in different environments. Following the general introduction, the participants started with one of the environments. The starting environment was counterbalanced between participants as well as the specific route they walked or the sequence of the spherical videos they watched, respectively.[3] Consequentially, measurements in the RE were conducted directly at the five locations in Würzburg and in the VE, measurements took place in a laboratory of the University of Würzburg.

Virtual environment

For the virtual environment, we first equipped and positioned participants with the HMD and headphones in the laboratory. Before we started the sequence of spherical videos, we asked participants to accomplish the numerical validation as provided by the manufacturer SMI as well as an external three-point validation (the average distance between validation marks and the recorded gaze points amounted to $M = 1.91°$, $SD = 1.42°$). Afterwards, participants started watching the spherical videos while being able to actively explore the environment with unconstrained head and eye movements. Furthermore, participants were allowed to move their body (e.g., to turn around) but they were instructed not to walk. After all spherical videos were played, we repeated the initial validation procedure, to ensure that the device was still properly calibrated (deviation between validation marks and gaze coordinates: $M = 1.28°$, $SD = 0.66°$). Directly after the exposure we assessed presence, i.e., the feeling of being there in a VR using the Igroup Presence Questionnaire (Schubert, 2003). Participants indicated a moderate feeling of presence ($M = 3.89$, $SD = 0.89$) on a scale 0 to 6. Simulator sickness was assessed using the Simulator Sickness Questionnaire (SSQ, Kennedy et al., 1993). Participants reported absence of most sickness symptoms and reached a total score of $M = 27.97$ ($SD = 21.36$) on a scale ranging from 0 to 235.62.

---

[3] Incorporating the environment and the location that participants started with into the analyses did not reveal statistically significant effects of these factors. Thus, order effects do not seem to constitute a source of error and were therefore neglected in the final set of analyses.

Real environment

For the real environments, we equipped participants with mobile eye-tracking glasses. Additionally, we asked participants to wear a baseball cap to reduce the influence of direct sunlight. Before walking to the first location in the real environment, the eye-tracker was calibrated and validated in the laboratory using a three-point validation procedure (average distance between validation marks and the recorded gaze points amounted to $M = 2.65°$, $SD = 2.93°$). Then the experimenter walked with the participant to the first location of one of the two predetermined routes. At every location, the eye-tracker was again calibrated using three predetermined landmarks. After calibration, participants received the instruction to hold a notebook for about 10s in front of their face and thus cover the camera of the eye-tracker. This was required to further align recording conditions between virtual and real environments: It simulated a sudden trial onset and reduced the influence of prior contextual information similar to the VE. On top, it was also used as an objective starting point for data analysis (see Image data processing below). Participants were further told that the experimenter would move out of their sight and were shown the direction of the hide-out. After answering potential questions of the participant, the experimenter asked them to bring the notebook in position and moved away. Participants were given about 2 minutes to freely explore the environment before the experimenter reentered the FOV and ended the trial. Since the experimenter had no further control over the behavior of the participant when waiting in the hide-out (e.g., about the precise time point when exploration of the surrounding started), we deliberately chose a longer viewing time than in the VE to ensure a sufficient amount of usable data. Note, that during active exploration of the environment, participants were not allowed to walk to keep the situation as similar to the VE as possible. During the recording, the experimenter tried to overview the location from her hide-out and estimated the number of pedestrians around the participant. For crowded places where the experimenter lost track of the total number of pedestrians, we set an upper limit of 20 pedestrians. Afterwards, the experimenter accompanied the participant to the next location, and the procedure was repeated. After the last location, the experimenter and the participant returned to the laboratory where the eye-tracker was calibrated and validated once more to ensure that proper recording quality could still be achieved (deviation between validation marks and gaze coordinates: $M = 2.16°$, $SD = 1.79°$).

6.2.4 Questionnaires

After finishing measurements in both environments, we asked participants to complete a brief demographic questionnaire, the Social Interaction Anxiety Scale (SIAS, Stangier et al., 1999, $M = 20.30$, $SD = 8.57$, *Range* = 8 to 41) and the Autism-Spectrum-Quotient short version (AQ-k, Freitag et al., 2007, $M = 6.50$, $SD = 3.32$, *Range* = 1 to 17). Upon completion of the questionnaires, we disclosed the actual aim of the study and explained that we measured gaze positions instead of pupil responses. We then offered the opportunity to delete the participants' data upon request but no participant made use of this possibility.

6.2.5 Image data processing

To analyze participants' viewing behavior in both environments, we manually scored what participants were looking at in their FOV. We tried to align data processing and analysis for both environments as closely as possible to permit a direct comparison of gaze patterns between viewing conditions. For that purpose, we first extracted individual video frames from both environments. For image data processing, we used Python (Version 3.7, Van Rossum & Drake Jr, 1995) with OpenCV (Heinisuo, 2019), NumPy (Oliphant, 2019) and Pandas (The PyData Development Team, 2019). For the VE, frames were extracted directly using Unity's screenshot function at 5 FPS at half of the internal monocular rendering resolution of the HMD (i.e., 756 × 840 pixels) resulting in a total of 150 frames for each location. For the RE, we first exported a video of the FOV as well as the log file for each participant and location via the software BeGaze (Version 3.7, SMI, 2017). Next, we extracted all frames at 5 FPS from the videos using OpenCV (Heinisuo, 2019). We kept 150 frames from trial onset to cover a comparable time window as for the VE. Trial onset was exactly 15 frames, i.e., 3 seconds after the notebook vanished from the first of all extracted frames. The delay of 3 seconds was necessary for the camera to adapt to the sudden change in lighting conditions due to the removal of the notebook. The resulting frames had a resolution of 1280 × 960 pixels. To prepare the extracted video frames for manual scoring, we added a gaze point at the respective gaze coordinates in form of a circle with a size of 1° to the video frames using

OpenCV (Heinisuo, 2019). Accordingly, the frame of reference for the gaze coordinates can be classified as head-centered (for more details and a discussion on the use and terminology of frame of references in eye-tracking research see Hessels et al., 2018). These gaze coordinates resembled the binocular gaze from a hypothetical cyclopic eye, as internally processed by the SMI software. Subsequently, two raters categorized the gaze points on a total of 66,000 video frames (44 participants × 2 conditions × 5 locations × 150 frames). Each rater scored one half of the stimulus set using the following scheme: First, raters categorized whether other persons were present in a given frame. For frames with persons present, raters additionally scored the gaze point as falling on one of three ROIs (*person*, *object*, *background*). Scoring followed a hierarchical assignment. If any part of the gaze point overlapped with any part of a person, the ROI for the frame was scored as *person.* If the gaze point was not scored as *person* but overlapped with an object that could be directly interacted with (e.g., car, bike, sign, baby carriage), it was scored as *object.* If the ROI was neither scored *person* nor *object* the gaze point was scored as *background* (e.g., sky, crosswalk, buildings). Frames missing a gaze point (e.g., due to blinks or recording difficulties) were categorized as *missing gaze.* For the analyses, we excluded all frames with *missing gaze* and frames in which no persons were present. To ensure that raters were consistent in their scoring, a subset of 5 participants (i.e., 7,500 frames) was scored by both raters, and we assessed their interrater reliability. Cohen's $\kappa = .87$ indicated a good agreement between both raters.

In VE, 97% of all frames included a valid gaze points, from which 73% of frames were further analyzed based on the presence of persons. In RE in contrast, valid measures of gaze points were only present in 58% of frames, from which again 73% included persons. Thus, the final analyses were based on 71% and 42% of data from the VE and RE, respectively.

6.2.6 Data analysis

We used the programming language R (version 3.6, R Core Team, 2020) for statistical analyses and numerical data processing. Specifically, we relied on the functionality provided by the *tidyverse* packages (Wickham, 2017) for data processing. To conduct linear mixed models, we used the *afex* package (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2019) as an interface for *lme4* functions (Bates, Mächler, Bolker, & Walker, 2015). Degrees of freedom to calculate p-values from the according t-distribution for the linear mixed model were obtained using the Sattertwhaite approximation (with *afex* via *lmerTest* package, Kuznetsova, Brockhoff, & Christensen, 2017). To calculate and plot the models' estimated marginal means, we used the *emmeans* package (Lenth, 2020). We used the conventional threshold of $\alpha = .05$ for determining statistical significance. All analysis scripts and data are available at `https://osf.io/hktdu/`.

Confirmatory analysis

To test the main hypothesis that social attention differs between VE and RE, we calculated the average gaze proportion on each ROI as a function of the environment for each participant and conducted a linear mixed model on these proportions using the fixed effects environment and ROI (ROI: persons or objects). Please note that the background ROI was dropped since all proportions sum up to 1 and thus the background information is redundant. The random effect structure for this Model 1 included random intercepts for participant ID and followed the preregistered *a-priori restricted model*. Although it would have been possible to also include a random intercept for location, we decided to rather rely on a parsimonious account and kept the preregistered model simple but suitable to address our research question. This approach seemed adequate given the small number of locations (Judd et al., 2012) and it followed conventions used in the field (i.e., 2 × 2 ANOVA designs on data aggregated across trials) as well as considerations that the variance-covariance matrices could be estimated precisely enough to avoid singularity (Matuschek et al., 2017).

Additional models were built upon the preregistered model but now included a maximum random effect structure with respect to the newly added predictors as recommended by Barr et al. (2013). First, we added predictors for *social anxiety* (Model 2) and *autism spectrum* traits (Model 3) to Model 1. For both new models, we included all additional 2-way interactions as well as the 3-way interaction of all factors. We used sum-to-zero contrasts for categorical factors in all models. To test the performance of the resulting models, we compared the log-Likelihood of Models 2 and 3 to the preregistered Model 1.

To analyze the consistency of viewing behavior across the five locations in each environment, we calculated Cronbach's α of gaze proportions using the *psych* package (Revelle, 2019). The generalizability across both environments was assessed by correlating average viewing preferences between VE and RE. Finally, to estimate the stability of viewing patterns between identical locations viewed in VE and RE, we calculated correlations between gaze proportions at each location and pooled them using Fisher z-transformations. All these analyses were accomplished separately for visual exploration of persons and objects, respectively, and the whole pattern of correlations was visualized using a correlation matrix including all pairwise Pearson correlation coefficients r for gaze proportion at each location in each environment for each ROI.

Exploratory analysis

For exploratory purposes, we conducted an additional linear mixed model (Model 4) including the number of pedestrians at the locations as a continuous fixed effect ($min = 1$, $max = 20$, standardized to $M = 0$ and $SD = 1$) and location as an additional random effect. Again, we initially specified the full random effects structure as in Models 2 and 3. As the full model did not converge, we pruned the model stepwise which resulted in a restricted model that included only uncorrelated random slopes for locations.

Finally, in order to elucidate general differences in visual exploration behavior between RE and VE, for example regarding the center bias relative to the FOV (Tatler, 2007), we plotted a smoothed density map (Gaussian kernel with a standard deviation of 1° of visual angle) of gaze positions relative to the FOV for a central viewing region spanning 60° × 46° for both environments across all participants.

6.3 Results

6.3.1 Comparison between real and virtual environment

To test our main hypothesis, we conducted the preregistred linear mixed model on gaze proportions with environment (RE vs. VE) and ROI (person vs. object) as fixed effects and participant ID as random effect[4]. This analysis revealed significant main effects for

---

[4] Since the residuals of the linear mixed model were not normally distributed, we ensured the validity of the present analysis by additionally calculating a robust linear mixed model (Koller, 2016). This robust model provided almost identical parameter estimates (see Table S1 in the supplementary material) and thus

Table 6.1: Estimated coefficients for the preregistered Model 1 with environment and ROI as fixed and participant ID as random effects for the prediction of gaze proportions.

	Estimate	SE	Df	t	p
Intercept	0.20	0.01	43	31.00	$< .001$
Environment (RE)	-0.10	0.01	129	-18.01	$< .001$
ROI (object)	0.01	0.01	129	2.15	.033
Environment (RE) × ROI (object)	0.02	0.01	129	3.55	.001

Note. The linear mixed model is based on sum-to-zero contrasts. RE: real environment, ROI: region of interest.

environment and ROI that were qualified by significant interaction of both factors (see Table 6.1). Overall, participants tended to look more on objects than on persons, but the significant interaction effect indicates that this was only true for the RE ($M_{\mathrm{RE,object}} = 0.13$, $SD_{\mathrm{RE,object}} = 0.05$, $M_{\mathrm{RE,person}} = 0.07$, $SD_{\mathrm{RE,person}} = 0.06$, $t(129) = 4.03$, $p = .001$) but not for the VE ($M_{\mathrm{VE,object}} = 0.28$, $SD_{\mathrm{VE,object}} = 0.05$, $M_{\mathrm{VE,person}} = 0.30$, $SD_{\mathrm{VE,person}} = 0.11$, $t(129) = -0.99$, $p = .758$). These findings confirm our primary hypothesis that social attention is reduced in the real world. Furthermore, the main effect of environment describes a general tendency of fewer gazes on persons and objects - and thus an increased amount of background exploration - in the RE as compared to the VE (see Figure 6.2A).

6.3.2 Consistency of viewing behavior within and across environments

In a first step, we assessed the consistency of gaze proportions on persons and objects, respectively, within each environment. Figure 6.3A illustrates that gaze on persons was more stable across locations in the VE (lower left triangle) as compared to the RE (upper right triangle). This difference was also evident in measures of internal consistency, which were substantially higher for the VE (Cronbach's $\alpha = .75$, 95% CI [.64, .86]) compared to RE (Cronbach's $\alpha = .38$, 95% CI [.32, .44]). By contrast, no such consistency was evident in gaze on objects (see Figure 6.3B) and we obtained low values of Cronbach's α in both, the VE (Cronbach's $\alpha = .29$, 95% CI [.24, .34]) and the RE (Cronbach's $\alpha = -.03$, 95% CI [-

---

supports our interpretations.

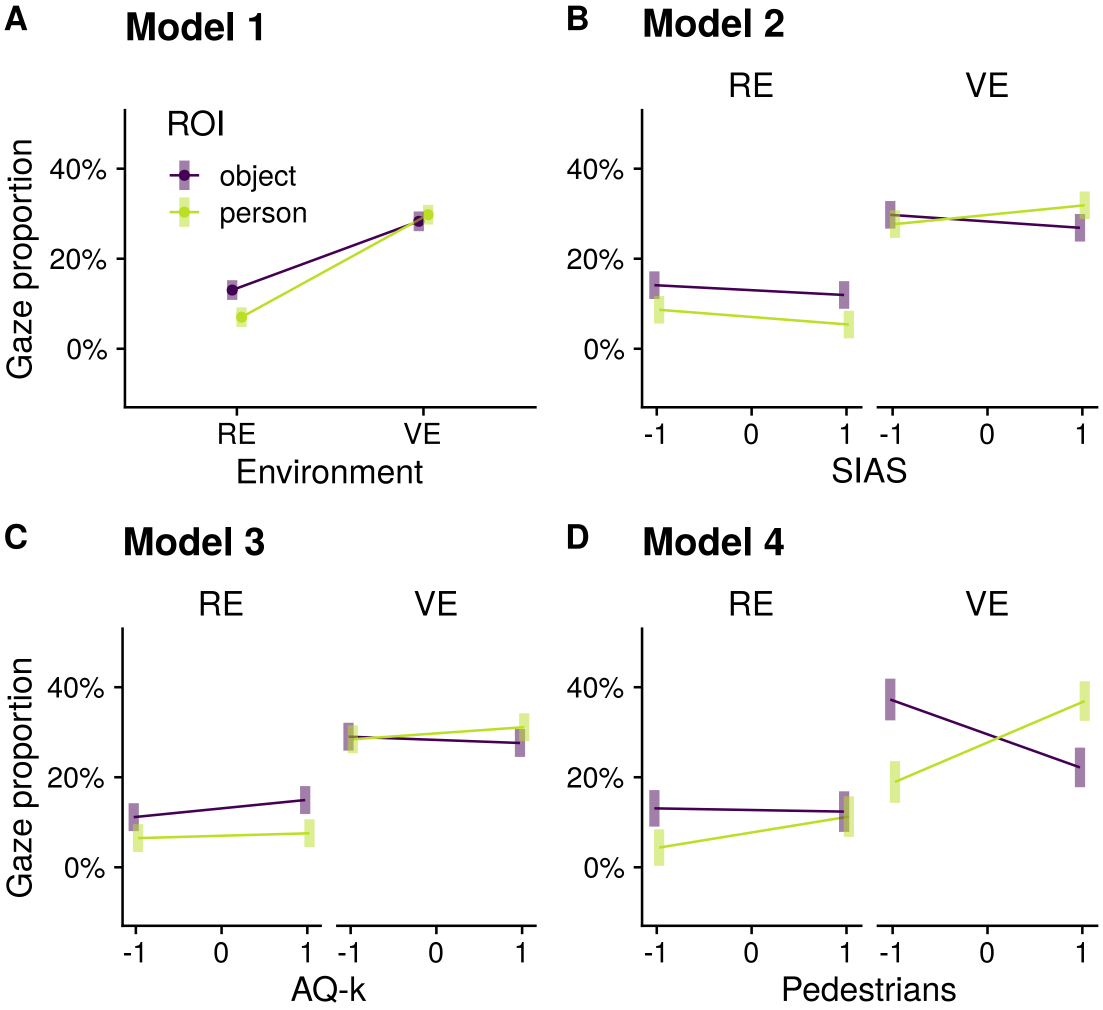

Figure 6.2: Interaction plots of all linear mixed models showing estimated marginal means as a function of the included predictors. Error bars depict 95% confidence intervals of predicted values. (A) The preregistered Model 1 included environment (real environment, RE vs. virtual environment, VE) and ROI (person vs. object) as predictors. Models 2 (B) and 3 (C) additionally included standardized scores of the Social Interaction Anxiety Scale (SIAS, depicted range from 12 (-1 *SD*) to 29 (+1 *SD*)), and the Autism-Spectrum-Quotient short version (AQ-k, depicted range from 3 (-1 *SD*) to 10 (+1 *SD*)), respectively. (D) Compared to Model 1, Model 4 additionally included the standardized number of pedestrians in the environment at each location (depicted range from 3 (-1 *SD*) to 15 (+1 *SD*)).

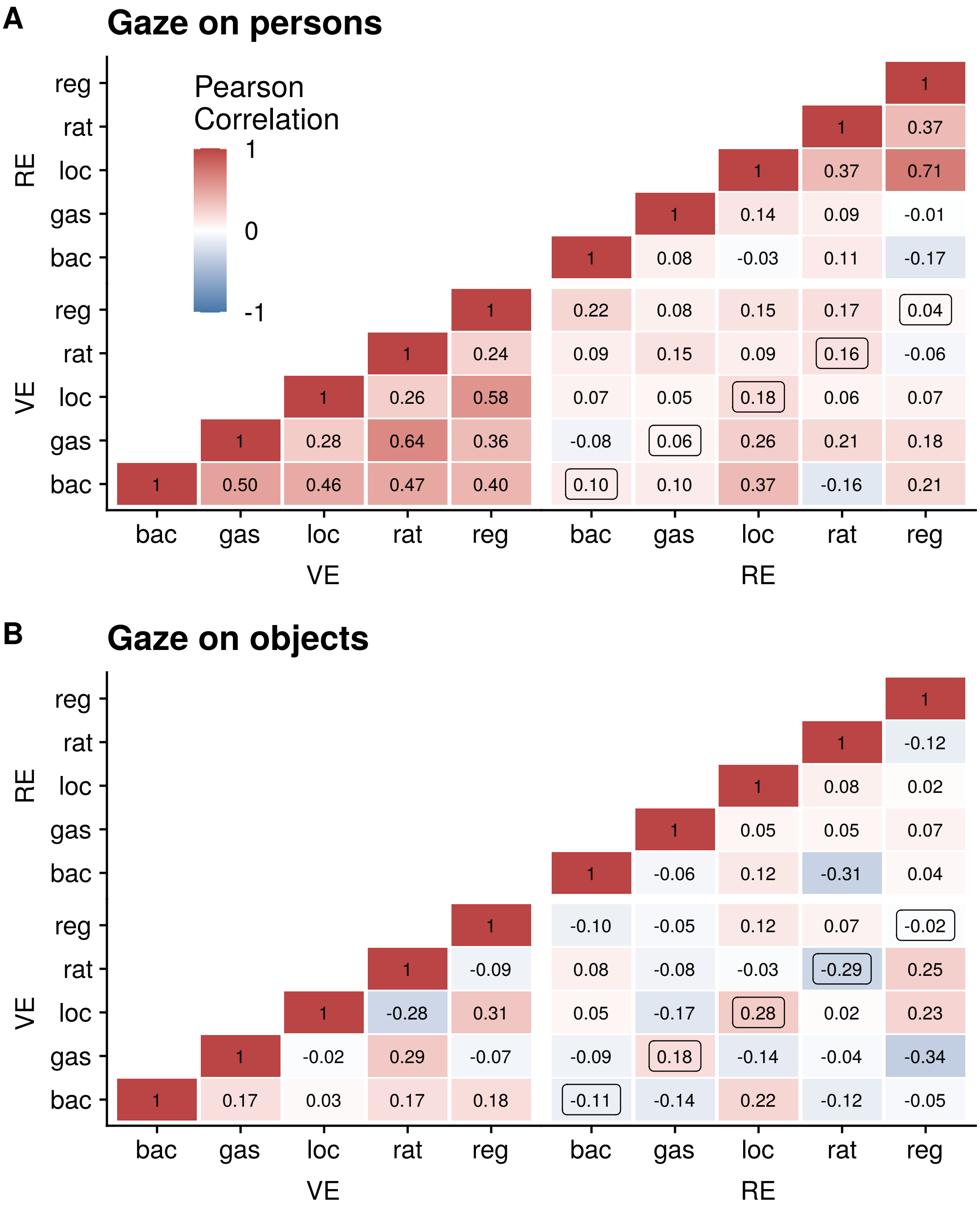

Figure 6.3: Heat map of pairwise Pearson correlation coefficients r for gaze on persons (A) and objects (B) for all five locations in both environments. Framed coefficients highlight correlations at identical locations between both environments.

.08, .03]). In order to estimate the generalizability of viewing patterns across VE and RE, we first calculated the correlation between average gaze proportions across locations between both environments. Although the correlation was positive for gaze proportions on persons ($r = .22$, 95% CI [$-.08$, .48], $t(42) = 1.46$, $p = .153$) but close to 0 for objects ($r = .01$, 95% CI [$-.29$, .30], $t(42) = 0.04$, $p = .965$), both correlations were not statistically significant and 95% confidence intervals overlapped. In a second step, we only focused on the correlation of gaze proportions between identical locations in the VE and the RE (see the highlighted diagonal in the lower right of Figure 6.3A and B). Although the average correlation was again descriptively higher for gaze on persons ($r = .11$) than on objects ($r = .01$), values are generally low, which indicates that viewing behavior differed between environments.

The spatial distribution of gaze coordinates within the FOV also indicates strong differences between VE and RE (see Figure 6.4). Whereas gaze points mostly clustered below the horizon in the RE and showed a larger spread on the vertical axis, they were vertically more centered slightly above the horizon in the VE.

6.3.3 Influence of personality

To exploratively test the influence of relevant personality traits on viewing patterns, we separately extended our preregistered Model 1 with the standardized scores of the SIAS (Stangier et al., 1999) and the AQ-k (Freitag et al., 2007) as fixed factors. The linear mixed model conducted to examine the influence of social anxiety (Model 2) did supply only weak evidence that social anxiety influences gaze proportions. Specifically, the three-way interaction between ROI, environment, and social anxiety just failed statistical significance (see Table 6.2). Interestingly, the previously estimated coefficients were very robust and did not change substantially with the inclusion of SIAS scores (see Figure 6.5). The linear mixed model incorporating autistic traits (Model 3) showed the same weak influence on gaze proportions (see Table 6.3). Again, previously observed effects were very robust (Figure 6.5).

Ratio log-likelihood tests between our preregistered model and the additional models considering individual differences in social anxiety and autistic traits supported our impression that model performance did not benefit from including personality factors. The additional

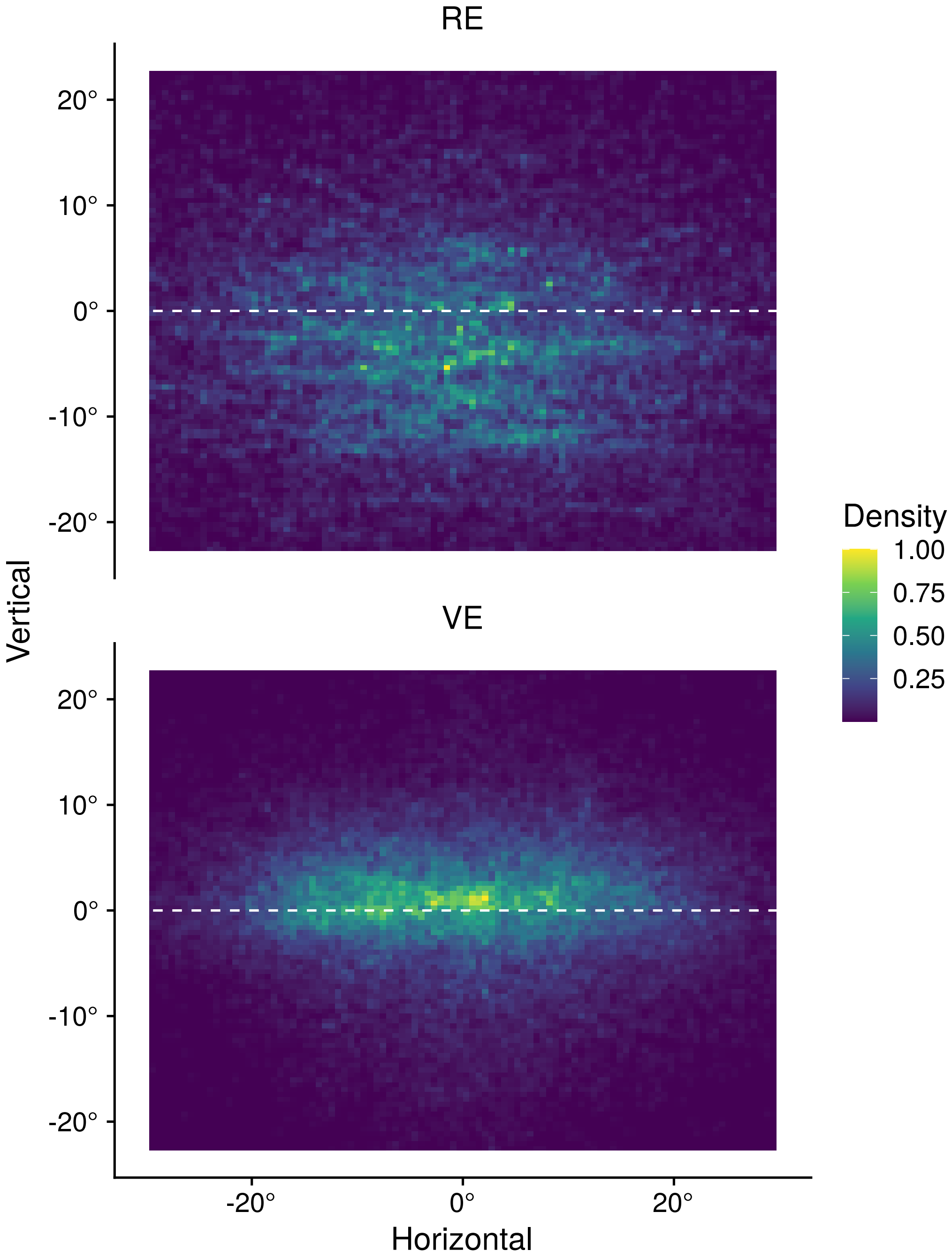

Figure 6.4: Smoothed density maps illustrating the distribution of gaze coordinate relative to the field of view in the real environment (RE) and the virtual environment (VE). Densities are depicted for a central area of 46° vertically and 60° horizontally (normalized to a range of 0 to 1).

Table 6.2: Estimated coefficients for the Model 2 with environment, ROI and SIAS as fixed and participant ID as random effects for the prediction of gaze proportions.

	Estimate	*SE*	*Df*	*t*	*p*
Intercept	0.20	0.01	42	30.88	$< .001$
Environment (RE)	-0.10	0.01	126	-18.40	$< .001$
ROI (object)	0.01	0.01	126	2.20	.030
SIAS	-0.01	0.01	42	-0.82	.418
Environment (RE) × ROI (object)	0.02	0.01	126	3.62	$< .001$
Environment (RE) × SIAS	-0.01	0.01	126	-1.44	.153
ROI (object) × SIAS	-0.01	0.01	126	-1.63	.105
Environment (RE) × ROI (object) × SIAS	0.01	0.01	126	1.97	.052

Note. The linear mixed model is based on sum-to-zero contrasts. RE: real environment, ROI: region of interest, SIAS: standardized sum score of the Social Interaction Anxiety Scale.

Table 6.3: Estimated coefficients for the Model 3 with environment, ROI and AQ-k as fixed and participant ID as random effects for the prediction of gaze proportions.

	Estimate	*SE*	*Df*	*t*	*p*
Intercept	0.20	0.01	42	31.18	$< .001$
Environment (RE)	-0.10	0.01	126	-18.04	$< .001$
ROI (object)	0.01	0.01	126	2.16	.033
AQ-k	0.01	0.01	42	1.23	.226
Environment (RE) × ROI (object)	0.02	0.01	126	3.55	.001
Environment (RE) × AQ-k	0.00	0.01	126	-0.32	.748
ROI (object) × AQ-k	0.00	0.01	126	0.85	.398
Environment (RE) × ROI (object) × AQ-k	0.01	0.01	126	1.61	.110

Note. The linear mixed model is based on sum-to-zero contrasts. RE: real environment, ROI: region of interest, AQ-k: standardized sum score of the Autism-Spectrum-Quotient short version.

Table 6.4: Model comparison of the preregistered model with the models including personality traits.

	Coef.	AIC	BIC	LogLik	deviance	chi^2	Df	p
Model 1	6	-412.83	-393.81	212.42	-424.83			
Model 1 - Model 2	10	-414.23	-382.52	217.11	-434.23	9.4	4	.052
Model 1 - Model 3	10	-409.91	-378.21	214.96	-429.91	5.08	4	.279

Note. AIC: Akaike information criterion; BIC: Bayesian information criterion; Df: Residual degrees of freedom.

parameters in the two latter models did not significantly increase model performance (see Table 6.4).

6.3.4 Relevance of the number of persons present in a scene

In general, the average number of pedestrians was comparable between the VE ($M = 10.60$, $SD = 6.23$) and the RE ($M = 8.18$, $SD = 6.15$), but there was substantial variability between locations, both in the RE ($Min = 4$, $Max = 20$) as well as the VE ($Min = 4.40$, $Max = 19.07$). In an additional exploratory analysis, we examined to what degree this number of persons who were present at a given location affects viewing behavior. We therefore added the number of pedestrians to the previously specified Model 1. This value was constant for every video shown in the VE but was estimated individually by the experimenter in RE. The newly specified linear mixed Model 4 included locations as an additional random effect and the number of pedestrians in the environment as an additional fixed effect, plus all interaction terms. The maximum model, including random intercepts for location and random slopes for the number of pedestrians at each location, did not converge. Therefore, we estimated the model suppressing the correlations between the random intercepts for location and random slopes for pedestrians[5]. Most interestingly, the two-way interaction between environment and ROI was substantially reduced in this model and did not remain statistically significant (see Table 6.5 and Figure 6.5). This was probably due to the strong three-way interaction between ROI, environment, and the number of pedestrians. Figure 6.2B shows that in the VE, a high number of pedestrians was associated with enhanced gaze on persons as compared

---

[5] See supplementary material for further details.

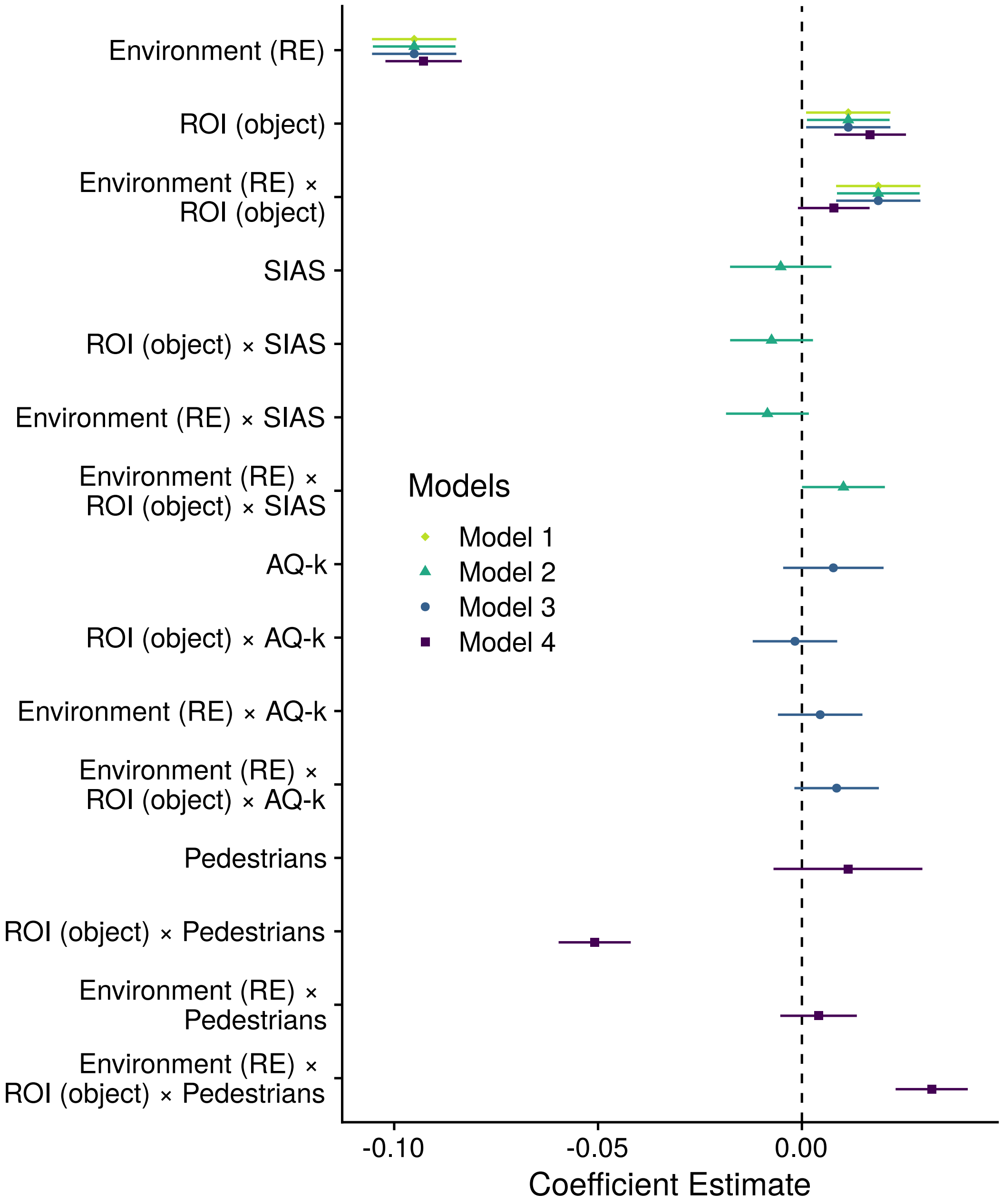

Figure 6.5: Dotwhisker plot of estimated model coefficients for Model 1 to 4 (dots) with 95% confidence intervals shown as whiskers. Environment (RE): Environment reference *Real environments*, ROI (object): ROI reference *Objects*, SIAS: standardized sum score of the Social Interaction Anxiety Scale, AQ-k: standardized sum score of the Autism-Spectrum-Quotient short version. Pedestrians: standardized number of pedestrians.

Table 6.5: Estimated coefficients for the Model 4 with environment, ROI and number of pedestrians as fixed and participant ID and location as random effects for the prediction of gaze proportions.

	Estimate	*SE*	*Df*	*t*	*p*
Intercept	0.20	0.01	4.84	14.08	$< .001$
Environment (RE)	-0.09	0.00	129.67	-19.43	$< .001$
ROI (object)	0.02	0.00	811.74	3.73	$< .001$
Pedestrians	0.01	0.01	2.26	1.21	.337
Environment (RE) × ROI (object)	0.01	0.00	811.74	1.74	.082
Environment (RE) × Pedestrians	-0.05	0.00	811.74	-11.26	$< .001$
ROI (object) × Pedestrians	0.00	0.00	545.23	0.85	.394
Environment (RE) × ROI (object) × Pedestrians	0.03	0.00	811.74	7.05	$< .001$

Note. Model 4: The linear mixed model is based on sum-to-zero contrasts. RE: real environment, ROI: region of interest. The number of Pedestrians was included as a standardized value with $M = 0$ and $SD = 1$. Location was included as additional random effect including uncorrelated random intercepts and random slopes by Pedestrians.

to objects, but this pattern flipped when only a few people were around. Qualitatively, such a pattern was also evident in the RE, but it was much less pronounced, and gaze proportions on persons never exceeded gaze proportions on objects.

6.4 Discussion

In the current study, we directly compared viewing behavior in real and virtual environments with a specific focus on social attention using spherical videos as a novel stimulation technique. In general, our results support previous findings (Foulsham et al., 2011; Laidlaw et al., 2011; Rubo et al., 2020) of a reduced attention towards conspecifics in the real as compared to the virtual environment. Extending previous studies, these results were obtained even when closely matching the laboratory environment to reality by using spherical videos recorded at the same locations that were also visited in the real world. These conditions allowed participants to freely explore and actively experience naturalistic stimuli in the laboratory while being contextually embedded in the environment. Since we observed reduced social attention in the real environment even in such closely matched conditions and a low correlation of gaze proportions on persons between both environments, our results indicate that the possibility to socially interact with other persons is the main driver of these differences between conditions. It thus seems sensible to assume that a real confrontation with conspecifics enhances the activation of social norms (e.g., not staring at others) and thus results in a reduced overt visual exploration of other persons in real life. This hypothesis is also supported by the observed modulation of this effect by the number of pedestrians in the surroundings. Whereas gaze on other individuals increased strongly with the number of pedestrians in the virtual environment, this effect was substantially weaker in the real world. Collectively, these findings indicate that it is not sufficient to focus on aspects of the viewing situation (e.g., active exploration, contextual embedding) to enhance the generalizability of laboratory findings on social attention to the real world (see Chapter 7 for a discussion of ecological validity in VR studies). The main aspect that modulates attention towards conspecifics seems to be the actual presence of other persons and the associated possibility for an interaction (cf. Risko et al., 2016; Zaki and Ochsner, 2009). These findings call for an enhanced focus on social interactions in social cognition research (Jaegher et al., 2010).

In addition to these variations of social attention between real and virtual environments, we also observed more general differences in viewing behavior between contexts. Interestingly, attention towards conspecifics seems to be more stable across locations in the virtual than the real environment and measures correlated only weakly between conditions. This could indicate that attentional preferences that were recently described for several semantic features and visual properties (de Haas et al., 2019; Linka and de Haas, 2020; Rubo and Gamer, 2018) are more robust in laboratory than in real life conditions and do not necessarily generalize from the laboratory to field contexts. Regarding gaze on objects, we neither found a stability of gaze proportions within each environment nor between conditions but this finding might also be attributed to the rather broad categorization of objects that neglected specific object classes (e.g., cars, symbols, text) or dimensions (e.g., static vs. moving or artificial vs. natural objects).

We also observed general differences in the spatial distribution of gaze coordinates within the FOV between virtual and real environments (see Figure 6.4). In both cases, a center bias (Tatler, 2007) was evident which is consistent with previous research using mobile eye-tracking in the field (Foulsham et al., 2011; Ioannidou et al., 2016) and stationary eye-tracking during video viewing (e.g., Tseng et al., 2009). However, this center bias was much more pronounced in the virtual environment where participants showed a substantially reduced spread of gaze points along the vertical axis. The reasons for this discrepancy remain elusive. On the one hand, it might be related to the HMD itself since wearing such device was novel to most participants (only 7% of the current sample indicated some previous experience with virtual reality). On the other hand, it could also result from an interaction between head and eye movements (Einhäuser et al., 2007) since participants were free to move their head in both environments. Unfortunately, tracking head movements could not be accomplished with the currently used eye-tracking glasses which precludes a detailed analysis of differences between conditions. Thus, it remains unclear whether participants more strongly relied on head movements to visually explore their surroundings in the virtual environment or whether the observed enhanced center bias in this condition indeed reflects less exploration. Furthermore, in the real environment, gaze was more concentrated below a relative horizon. Interestingly, this is compatible with results from studies with walking participants (e.g., Foulsham et al., 2011; or Matthis et al., 2018) even though participants were not allowed to walk in the current study. Although speculative, this could indicate that the real environment primed participants to engage in a more active mode of visual exploration that includes planning for potential walking movements. Taken together, these general differences between viewing conditions highlight the need for future studies to elucidate these aspects in more detail before uncritically translating experimental paradigms to VR environments and assuming comparability to field conditions.

Regarding the influence of personality traits on gaze proportions, we neither observed significant effects of social anxiety nor of autism spectrum traits. This contrasts with previous studies that documented reduced attention towards faces or eyes of conspecifics in individuals with high autism spectrum (Hessels et al., 2017; Laidlaw et al., 2011) or social anxiety traits (Howell et al., 2015; Rubo et al., 2020), respectively. Note however, that some studies did not observe general effects of such traits but rather only for specific situations, e.g., an effect of social anxiety on gaze at people in the vicinity of the observer (Rubo et al., 2020). Moreover, other studies failed to observe effects of social anxiety or autism spectrum traits on measures of social attention in real environments (e.g., Horn et al., 2021; Rösler et al., 2021; Vabalas and Freeth, 2015). The current findings might therefore be attributed to a genuine absence or a very small effect of personality traits on viewing patterns, that could not be reliably detected with the limited sample size of the current experiment. Alternatively, such effects might only surface in more heterogeneous samples that also include participants with clinically relevant autism spectrum or social anxiety symptoms.

Although our study has several strengths including a close matching of laboratory and field conditions regarding data acquisition and analysis, it also comes with some limitations. First, scene presentation in the laboratory was somewhat restricted by technical limitations of the HMD (for a more general discussion see Chapter 8). For example, the display resolution degraded the degree of detail of objects and pedestrians in the distance. However, we do not believe that these limitations had a major impact on the results of this study since the videos were short, novel and interesting and therefore effectively captured participants' attention. No participant complained about the presentation quality or spontaneously mentioned problems with the HMD. We believe that these technical limitations will also become weaker as this technology matures. Second, most of the participants were not experienced with VR and this novelty might lead to certain viewing biases. However, as the current results are comparable with previous findings obtained in other settings (Foulsham et al., 2011; Rubo et al., 2020) and since we observed more consistent instead of more variable viewing patterns in the virtual environment, we suspect these biases to be rather small. Third, our research design involved walking to the locations in the real environment and consequently, participants had prior information about the location before the actual trial began. This difference to the VE could hardly be eliminated but we tried to reduce its impact by choosing well-known locations in the city of Würzburg, Germany, that should be familiar to most participants. Moreover, to align recordings conditions between virtual and real environment, we required participants to use a notebook to cover their sight before starting measurements in the RE. This procedure was implemented to reduce the influence of contextual information and to simulate a sudden trial onset similar to the VE. Fourth, although we tried to match presentation conditions in virtual and real environments as closely as possible, some environmental factors were beyond experimental control. Apart from weather conditions and daytime, this mainly applied to the number and behavior of pedestrians at the different locations. However, the average number of pedestrians was comparable between both environments and we explicitly considered the variability across locations in an exploratory statistical analysis that also revealed a crucial influence of this factor on measures of social attention. Fith, reality is multimodal. Within our setup, we tried to account for this by including visual and auditory stimulation in the virtual environment (Zaki and Ochsner, 2009). Although we suggest that these two modalities are most important for generating a sense of presence, it seems interesting for future research to stimulate additional senses (e.g., olfaction) and improve the audiovisual stimulation (e.g., by including 3D sound). A final limitation might be the lack of body representation in the virtual environment. Body representation seems to enhance presence in virtual reality (Sanchez-Vives and Slater, 2005) but because of technical limitations, participants could not see their own body within the currently used spherical videos. Although none of the participants articulated irritations regarding the missing body, it seems interesting but also ambitious for future research to include a rendering of the own body into the virtual environment. While such procedure might enhance a feeling of presence, it also certainly requires an additional experimental phase to familiarize participants with this new situation.

Besides these limitations and the differences between virtual and real environments that were observed in the current study, we see great potential in the use of spherical videos as stimuli for social cognition research. Compared to 3D virtual reality environments, spherical videos are comparatively cheap and easy to generate. These videos can be presented using HMDs to allow for natural head and body movements and permit the acquisition of eye-tracking data that is not deteriorated by quickly changing light or weather conditions that can be encountered in real life environments involving mobile eye-tracking (Niehorster et al., 2017). Since our results indicate that the possibility for social interaction seems important for modulating social attention, it might be an interesting approach for future research to script spherical videos in order to effectively simulate such interaction. Although such approach seems demanding since the observer's behavior is difficult to predict and would therefore require a precisely orchestrated scene, some basic aspects of social attention might well be simulated with such scripted videos. For example, a crucial aspect of social interaction is eye contact (Ellsworth et al., 1972; Wirth et al., 2010), which could be simulated by purposefully looking into the camera at defined time points during the recording of the spherical video. Furthermore, it has been shown that social status is relevant for gaze allocation (Foulsham et al., 2010) but in this study, participants watched a group discussion on a desktop monitor "as if they were in the room". Spherical videos could further enhance the external validity of such study designs. As another example to test the influence of norms, one can think of a setup similar to Risko and Kingstone (2011). They concealed the fact that they recorded eye movements by apparently switching off the eye-tracker. This manipulation resulted in a substantial change in viewing behavior, presumably caused by a shift in social norms. Similarly, Cañigueral et al. (2018) also showed that wearing an eye-tracker itself alters viewing behavior. Assuming compliance with ethical considerations, an HMD setup holds the opportunity to completely conceal eye-tracking. It is easy to implement with an HMD since the built-in eye-tracker is usually not recognizable by laypersons. All in all, we feel that we have only touched the surface of what is possible with the usage of spherical videos for social cognition research. At the same time, several limitations of (interactive) eye-tracking with unrestrained head movements are addressed (cf. Valtakari et al., 2021). We believe that this technique offers great potential for many research questions, especially since accessibility increases with the availability of spherical cameras and HMDs with included eye-tracking devices.

To sum up, this study examined the reliability and validity of spherical videos for examining social attention and it provided evidence for a reduction of gaze on other persons in real life as compared to laboratory conditions even when closely matching both environments. Viewing behavior was largely unaffected by social anxiety and autism spectrum traits but was modulated by the number of persons in the scene, especially when viewing spherical videos. In addition to these findings, we also observed general differences between virtual and real environments with respect to the stability of viewing patterns across locations and the spatial distribution of gaze proportions within the field of view. Despite these discrepancies, we believe that the use of HMDs and especially spherical videos holds great promise for social cognition research since they allow for a multimodal, contextually embedded, and dynamic stimulus presentation (Parsons et al., 2017; Risko et al., 2016; Zaki and Ochsner, 2009). However, the simulation of potential or actual social interactions in controlled laboratory research remains a challenging problem where, as discussed, spherical videos are only of limited help.

6.5 Acknowledgement

We thank Pia Linden for significant contributions to the experimental design (e.g., filming and processing of spherical videos, piloting the HMD eye-tracker and software) and Isabelle Lindner for her help during data acquisition and scoring.

Data and materials of this experiment are available at `https://osf.io/hktdu/` and the experiment was preregistered at `https://aspredicted.org/p7a83.pdf`.

Part III

General discussion

The present work demonstrates the demands for future social attention research. Furthermore, it offers novel evidence for virtual reality as compensation for shortcomings of traditional research paradigms. This is shown in three studies and by applying various methods to extend the understanding of social attention.

Several prior studies have noted the importance and relevance of gaze cueing as the *window to social cognition* (Shepherd, 2010). However, the body of evidence has never been accumulated in a systematic literature review and meta-analysis. As expected, the robust gaze cueing effects across studies was evident. However, the experimental factors included did not explain the surprisingly large variance in the published results. Thus, there seem to be further, not well-understood variables influencing these social processes.

Notably, the gaze cueing effect can still be observed in a wide variety of experimental designs. The second study provides evidence that gaze cueing is also elicited in a more ecologically valid research design. The conceptual replication of Zwickel and Võ (2010) shows that the gaze is followed when the cue (i.e., the human) is embedded within complex and contextual meaningful scenes. These findings corroborate results from the meta-analysis on robustness but for a different experimental setup.

A highly ecologically valid laboratory research design was used for the third study to compare social attention in natural and virtual environments. In a so-called free-viewing paradigm fixation patterns were measured, in the field with a mobile eye tracker and in the laboratory playing spherical videos in an HMD with an integrated eye tracker. Both environments were almost identical, but differences could still be observed between the environments. The study provides additional evidence of for influence of social interactions and norms on basic social attentional processes. Additionally, it addressed limitations of previous studies examining the difference between laboratory and field studies. This study demonstrates the new opportunities in social attention research. It is often claimed to use VR for such research questions. However, this is still rarely done (for an exception see Rubo and Gamer, 2021). Study 3 shows that VR can untie the antagonism between experimental control and ecological validity (Blascovich et al., 2002). However, the differences in viewing patterns show that VR is not just a simple drop-in solution for ecologically valid studies.

Overall, the present work examined the current state and future direction in social attention research. It highlights the robustness of gaze following processes and the need for methods capable of investigating the influence of ecologically valid on social cognitive processes.

The following part of the present work discusses in great detail the implications of the newly gathered evidence. First, ecological validity itself is critically assessed in the light of change coming from VR as a research method. Second, practical considerations are given, again with a focus on ecological validity. Third, several future research lines are sketched that seem to be a promising area of application for the VR.

Chapter 7

Rethinking ecological validity

The present findings highlight the need for ecological validity in social attention research. However, with new possibilities in research designs, ecological validity as a concept might also need to evolve to meet the new developments. The idea of ecological validity as it is used today is already heavily criticized for not being precisely defined (see Chapter 2, Holleman et al., 2020; Mook, 1983; Schmuckler, 2001). This criticism is presumably exacerbated in its justification by the spread of VR in psychological research. Different methods for creating virtual environments (i.e., recording vs. developing, see Section 2.2 and the following chapter) highlight that the concept of ecological validity as used today might not be sufficient to cover new research designs with adequate precision.

Ecological validity underwent a significant change in meaning in the history of psychology. The first record of the term can be pinned to Brunswik (1947). In a broader framework about *representative research designs*, he used the term for describing the predictive utility of a cue (the "proximal cue") about the state of an environment (the "distal cue"). According to Brunswick, a good (i.e., representative) research design should take the limited predictive utility of cues into account (Brunswik, 1947; Holleman et al., 2020). With that, he criticized learning studies at that time for having perfect correlations (+ / 1 or 0) between proximal and distal cues. These perfect correlations are most often unrealistic in natural environments. Thus, he argues, limit the generalizability of the results of such research designs (Brunswik, 1947).

As an illustration, a person judging how angry someone is might use proximal cues such as facial expression, facial redness, or body posture to infer the state of a distal cue, i.e., the level of anger. In reality, however, facial redness might also indicate shame. Consequently, the ecologic validity of facial redness for an angry person is not perfect (i.e., the correlation is below 1). A representative design incorporates such imperfect correlations. As such, ecological validity is part of a larger framework of "representative designs". These designs, according to Brunswick, lead to better research designs that ensure generalizability (Brunswik, 1947; Holleman et al., 2020). Accordingly, ecological validity needs a meaningful reference class to formulate and address specific limitations of experiments (Holleman et al., 2020; Schmuckler, 2001).

The current use of ecological validity (Shamay-Tsoory and Mendelsohn, 2019; Sonkusare et al., 2019, e.g., as in), however, is criticized for its lose definition (Holleman et al., 2020) . Although Brunswick's use of the term is rather different, Holleman et al. (2020) argue that the implications still hold for ecological validity in the commonly used sense. For example, it becomes clear that ecological validity quickly falls short in describing relevant differences between a recent VR study (Rubo and Gamer, 2021) and the current Study 3 (see Chapter 6). Rubo and Gamer (2021) investigated the influence of the experimental setting on the reactivity to social gaze. They measured eye movements towards a virtual agent in a virtual environment and on a computer screen. The virtual agent brought different objects from a house and placed them onto a table in the front yard and, while doing so, randomly smiled at the participant. Participants either stood at the driveway when observing the agent in the virtual environment or watched the agent in a video scene on a computer screen. Comparing the eye movements of the two groups of participants, the authors found a stronger reactivity towards the social gaze in the virtual environment. Hence, as Study 3 (see Chapter 6), the different gaze behavior in the groups provides additional evidence on the relevance of ecological validity in social attention research. In contrast to Study 3, however, Rubo and Gamer (2021) used a more reactive environment and computer-generated world. Additionally, they compared the eye movements in a between-subject design with eye movements recorded at watching a video on a computer screen (these between subject comparisons of eye movements from different experimental settings are discussed in detail in Section 2.3 and Section 6.4). Still, both studies claim high ecological validity, even though the studies differ remarkably with respect to what they aim to generalize. A closer examination reveals that both studies lack different features of the *real world*. In Rubo and Gamer (2021), the virtual environment is clearly a computer-generated world. Thus, it lacks some visual details of the depicted persons and objects that are available by a spherical video presentation. Study 3, however, falls short regarding interactions and reactivity of the depicted persons. Obviously, the term ecological validity is limited in describing the properties of those studies in a meaningful and constructive fashion.

Thus, the highlighted limitations of the current use proposed by Holleman et al. (2020) or Schmuckler (2001) might become more evident in virtual reality studies because they potentially approximate real-life in its entirety. With such possibilities, deviations from the full potential of VR, e.g., as in Study 3 that lacks interactions, must be explicitly stated for the research to be sound. In fact, most VR studies come with limited approximations of real life. These limitations might come from finite resources for developing complex VR environments (see the following chapter) and deliberate research design decisions. For example, Study 3 addresses specific flaws in more traditional research designs with videos played on a computer screen. Thus, spherical videos were deliberately chosen, including the introduced deficiencies of the technology. Finally, although empirical research is never generalizable in total (e.g., over centruies Cronbach, 1975), VR expands the domains to which a single experiment can be generalized. From these capabilities arises the responsibility to report aspects the given research aims to generalize in detail with novel requirements for the concept of ecological validity.

Chapter 8

Practical considerations for VR

VR is a powerful tool to study cognitive processes in naturalistic situations (Parsons et al., 2017). It provides very naturalistic environments and offers a high degree of experimental control. Additionally, participants are naturally embedded in a scene. On top, such research design allows for precisely tracking human behavior. It is possible, that this development changes empirical research as much as the introduction of the computer did around 50 years ago (Aaronson et al., 1976).

As outlined, social attention research faces challenges that are not easily addressed with traditional methods (Risko et al., 2016). Research designs should therefore take advantage of the new possibilities offered by VR technology (Blascovich et al., 2002; Parsons et al., 2017). By comparing two different VR studies, the previous chapter highlights that the concept of ecological validity as commonly used today might not be granular enough. In general, two options exist for creating a research design with virtual environments. Virtual environments can be realized with recorded spherical videos. Alternatively, they can be developed in game engines. These technologies have different consequences for the ecological validity, for the experimental design, and finally for the addressed research question. Here the two somewhat distinct approaches towards VR will be contrasted, highlighting the strengths and weaknesses of each approach.

Study 3 of the present work includes a study that uses a VR research design with spherical videos. In these virtual environments, participants are exposed to photorealistic representations that can be naturally explored by body and head movements. This method has two main advantages (for details, see Study 3): It is the fastest option to create a complex virtual environment, and it provides photorealistic scenes. Spherical videos are a cheap and quick method to create complex and naturalistic virtual environments that still allow natural exploration. The scenes are recorded with a spherical camera (i.e., a camera that films omnidirectional with multiple lenses) and projected onto a virtual sphere. The complexity of the scene does not increase the effort needed to create a virtual environment. Thus, this technique is advantageous in capturing highly complex scenes. By no means is the technique limited to recordings of public places, as used in the present study. An exciting opportunity also lies in recording orchestrated scenes for specific research questions. The second advantage is that such scenes are inevitably photorealistic. This is great, as the visual properties closely match what humans see. A downside of the method is that research designs in which participants interact with the environment (including other persons) can only be realized within extremely narrow boundaries. Although, the aforementioned orchestrated scenes might offer a partial solution. For example, something like eye contact can be achieved by recording persons focusing on a lens, providing an interaction that might be sufficient for particular research questions (see the next chapter for details). However, individual reactions of the participant can hardly be acted upon in such virtual environments. Note, however, that photorealism cannot be escaped with such videos, as such research designs are also bound to physical laws and real objects. A limitation that the computer-generated virtual worlds do not have.

Virtual environments can also be developed within a game engine. Rubo and Gamer (2021) created such a rendered virtual scene. This approach has the advantage that the scene is entirely in the developer's hands. All features can be changed during the developmental phase or even interactively changed within an experimental procedure itself. This technique is most often referred to, when discussing the tremendous power that VR offers to researchers (Parsons et al., 2017). In contrast to spherical videos, physical laws can be hurt in such environments, such as traveling back in time (Friedman et al., 2014) or it is also possible to meet oneself (Yee and Bailenson, 2007) or embodying someone else. With these almost endless possibilities, one should be aware that, unlike with spherical videos, the more complex a virtual scene becomes, the higher are the demands for the developer. Photorealism, for example, is hardly achieved without a multi-million dollar budget that film companies like Disney have at disposal. As a consequence, the look and feel of VR studies using this approach are often comparable with computer games. Whether this is problematic or not depends on the specific research questions. For gaze cueing, for example, the meta-analysis revealed that the degree of photorealism of the cue does obviously not alter covert attentional orientation processes (see Study 1), at least not in traditional research designs.

For interactions, it is also the case that increasing complexity results in increasing development demands. Reality-like interactions, including social interactions and interactions with physical objects, are challenging to achieve. However, with VR it is now possible to have reality-like research designs. This incorporates great opportunities to manipulate or isolated social interactions to every degree (for details, see the next chapter). Still, the introduction of interactive virtual characters either driven by a human (i.e., an avatar) or by a computer (i.e., an agent) enables researchers to systematically and independently test the effects of various social cues (Parsons et al., 2017; Risko et al., 2016).

For interactions, it is also the case that increasing complexity results in increasing development demands. Reality-like interactions, including social interactions and interactions with physical objects, are challenging to achieve. However, every degree is possible, so isolated social interactions might be feasible (for details, see the next chapter). Still, the introduction of interactive virtual characters either driven by a human (i.e., an avatar) or by a computer (i.e., an agent) is desired by the researcher to systematically and independently test the effects of various social cues (Parsons et al., 2017; Risko et al., 2016). However, reality-like VR worlds do not only come by an increase in development costs. Researchers should also be aware of the so-called *uncanny valley* (Mori, 1970 as cited in the official English translation Mori et al., 2012). Generally, it is assumed that acceptance and affinity rise the more human-like or real, for example, agents in VR are. However, the uncanny valley hypothesizes a specific range before approaching perfect human-likeness (i.e., human-likeness of being a human) where the general acceptance and affinity are reversed. And in fact, this effect could be shown in robotics (MacDorman and Ishiguro, 2006) and artificial faces (Seyama and Nagayama, 2007). An agent of almost perfect human-likeness thus might elicit rejection and repugnance and might not be worthwhile, depending on the research question.

Both approaches, computer-generated and spherical video scenes, are on par regarding natural exploration behavior. However, they still differ in several aspects regarding ecological validity. Spherical videos achieve high ecological validity concerning visual presentation.

In contrast, computer-generated virtual environments enable the full potential of VR with unlimited possibilities. However, interactions are a hard problem in spherical videos. In computer-generated environments, complexity might quickly outgrow the available resources. Still, the technology, in general, holds tremendous potential for social attention research. As it is a very young technology in psychological research, a lot of research is waiting to be done without the limitations mentioned above.

Chapter 9

The study of gaze cueing in VR

The final chapter explores the unification of two lines of research presented in this work. Namely, the combination of investigating gaze cueing in the laboratory alongside the full potential of VR studies. The present work sets out that these two approaches can be merged to investigate the underlying conditions of social attention. Evidence in this field can be further strengthened by adding VR research paradigms to current approaches that often use simple and static stimuli. This combination releases research of the properties of social attention that are particularly difficult to study in classical laboratory experiments. The difficulties arise from several sources. First, it is almost impossible in a classical laboratory experiment to investigate an unconstrained attentional flow of individuals (as in Study 3 from Chapter 6). An unconstrained attentional flow is by definition also accompanied by unrestricted head and body movements, something a classical laboratory setting can barely provide. Exceptions, for example Laidlaw et al. (2011), often face problems in distinguishing between (potentially strong) demand characteristics of a computer screen and social attentional prioritization. Additionally, when such limitations are addressed by research designs containing confederates, experimental and temporal control of the experiment is often limited. The examples outlined in the following are intended to further emphasize the versatility and usefulness of VR.

Humans provide a multitude of social cues, and a hierarchy for different social cues is assumed. This theory originated from neurophysiological evidence from macaques (Perrett et al., 1992) and is also supported by empirical (Hietanen, 2002, 1999) and meta-analytical evidence (see Study 1 from Chapter 4) on gaze cueing in humans. Besides eye and head direction, other social directional cues include body posture (Azarian et al., 2017), body gestures (Langton and Bruce, 2000), and human motion (Shi et al., 2010). According to the theory, these cues should be appropriately placed within the hierarchy. These studies have a specific methodological limitation in common, as they use highly controlled but not ecological valid stimuli. Shi et al. (2010), for example, used a point-light human for inducing human motion. VR can directly address these limitations, offering a controlled and ecologically valid introspection into the potentially diverging cognitive processes for the various, potentially competing intrapersonal social cues. Another limitation such classical computer experiments face is that they only display a miniature of an entire human body. This might be problematic when investigating the integration of several, if not all, directional social stimuli. For participants, the eyes might become indistinguishable small in such presentations, and for researchers, such small ROIs require high measurement precision. VR, by contrast, allows the exposure of life-sized persons to participants. Therefore it is particularly suitable for addressing this research question. Furthermore, it provides the possibility to integrate these mostly insular findings into a joint research design. Again, a point-light human, for example, misses a real head and with that all facial features. So, VR also allows investigating how various sources of social information are integrated. An early future study addressing these research questions might first investigate the temporal flow of attention towards the different sources of information. Computer-developed VR scenes might be the favored approach as they provide sufficient temporal control. In later work, ecological validity on the visual spectrum can be extended via spherical videos. This allows the investigation of very complex environments with numerous present humans and social cues, presumably at the expense of temporal resolution.

So far, empirical work on gaze cueing from groups included only field observations (Gallup et al., 2012; Milgram et al., 1969). For example, Gallup et al. (2012) filmed and counted how many pedestrians followed the gaze of different group sizes of confederates towards a building's rooftop. These studies consistently show that gaze following increases as a function of group size. However, such studies do not allow the investigation of cognitive processes, as the behavior of individuals is only rated subsequently, and other individual characteristics are not collected. This limitation could be addressed by filming such orchestrated behavior of a group with a spherical camera. Then participants' eye movements could be investigated individually in a laboratory VR setting. This research might provide important insights into the integration of competing interpersonal social information. Such research designs are *representative research designs* (Brunswik, 1947) or, in other words, should have high ecological validity (see Chapter 7 for more details). This is especially important, as social cues are rarely unambiguous, nor do they occur in the absence of other competing social cues. As mentioned before, only in VR such a scene can be investigated without directing the participants' attention. The life-size *stimuli* (i.e., depict persons) are additionally beneficial for a realistic availability of social cues (e.g., contrary to missing information from small heads).

These holistic investigations of social cues from humans are then very close to gaze cueing in authentic, interactive designs. A *real interaction* compared to the previously sketched designs might include initiation and meaningful intentions or consequences. For the initiating phase, it was shown that gaze contact has various influences on cognitive processes (for a review see Conty et al., 2016). For example, Bristow et al. (2007) and Dalmaso et al. (2020a) showed that attention-grabbing features of gaze contact facilitate gaze cueing compared to when the previous face showed an averted gaze.

Previous research found that intentions and goals alter the gaze cueing effect as well. A study by Perez-Osorio et al. (2015), for example, showed gaze cueing to be susceptible to the other's goals and intentions. In each trial they presented participants a context by showing an image of laundry or a bar. Additionally, a request to bring some softener, respectively, a drink was given. The follow-up scene included two horizontally placed bottles with glasses next to them. A person appeared, and after a few milliseconds of straight gaze, she looked towards a bottle of softener or orange juice. The person again looked straight ahead, and the target, a filled glass of softener or orange juice appeared. Participants reacted to which of the glasses was filled. Accordingly, the cues were context congruent when the context scene was congruent, e.g., the softener bottle was cued in concordance with the depicted image of a laundry. Cue validity, in contrast, was established when the depicted person cued the glass that was filled. This setup allowed the investigation of goals and intentions on gaze cueing. It represented also a relatively (for traditional research) ecologically valid research design. As in Study 2, the authors showed that gaze cueing takes place in settings with increased ecological validity. On top, they found a larger gaze cueing effect in the context of congruent conditions. VR can advance this and similar research designs (i.e., the one used in Study 2) in ecological validity. For example, VR can provide a more natural and omnipresent context. Instead of priming the context with a 2D picture, participants could find themselves in a laundry or a bar. With these modifications, the gaze cueing paradigm would more and more evolve into a naturalistic situation with real interactions that elicit joint attention.

In contrast to gaze following, joint attention describes an enduring state where two individuals initiate and maintain coordinated actions by creating a *perceptual common ground* (Sebanz et al., 2006). When target responses, as well as sustained interactions, are included in gaze cueing paradigms, they more and more enter the field of joint attention. The cueing reactions become meaningful and grounded in shared representations. These shared representations are important for efficient solutions in cooperative tasks (Clark and Krych, 2004) and for improving verbal communications (Özyürek, 2002). Such tasks require a recurring gaze following for the maintenance of joint attention. Arguably this is one, if not the one, key function of gaze following. It is essential in our daily lives but also for the understanding others intentions (Baron-Cohen, 1991; Perrett and Emery, 1994). Although not widely discussed in the present work, this field also lacks natural stimuli (Zaki and Ochsner, 2009). Again, VR can be a prosperous source for enhancing current research designs in the field. For example, Böckler et al. (2011) found that a gaze cue can be facilitated with a preceding gaze contact with a research design that is close to a classical gaze cueing paradigm (using both cartoon and photographs of faces). Thus the experiment has excellent experimental control at the expense of ecological validity. A close replication of the paradigm in VR could further strengthen the evidence by adding complex and dynamic stimuli. For the implementation of this paradigm, participants could find themselves at the table with two other humans. The depicted humans engage in gaze contact (or not) followed by a gaze cue to either a pear or apple on the table. This could provide more generalizable insights into the initiation of joint attention. For investigating maintenance of joint attention, a sustained or reoccurring task could then follow. In general, the benefits for joint attention are similar to what is reasoned for gaze cueing studies.

The last step towards reality in VR might be the agency of the depicted persons. In general, depicted persons are computer-controlled. However, the behavior of the human representation can also be controlled by humans. In VR terms, representations of humans are called agents when they are controlled by a computer program. When humans are controlled by another human, they are called avatars. For gaze cueing, first studies show that agency affects participants' social attention, with stronger gaze cueing effects from avatars (Wykowska et al., 2014). Additional evidence comes from studies showing differences in the neural processing of gaze cues from avatars vs. agents (Caruana et al., 2017; Pfeiffer et al., 2014). Again, such designs can significantly benefit from an adaptation for VR from the advantages already mentioned.

These various directions sketch how gaze cueing studies in VR can advance social attention research in general. In sum, gaze cueing research with VR allows almost continuous manipulations of ecological validity in research designs. Furthermore, it provides research designs with a holistic representation of humans.

Chapter 10

Outlook and concluding remarks

Virtual reality holds excellent opportunities for psychological research for multiple reasons. First, it offers naturalistic environments and, at the same time, tight experimental control. Second, natural behavior can be a fully integrated component as participants are completely embedded in such environments. Third, the technology inherits the precise tracking of behavior. These features make VR an attractive approach for investigating social attention. As the given work argues, especially the limited generalizability of social attention findings can be fruitfully addressed by using VR technology. Three studies are reported to further substantiate this often formulated claim.

The first two studies highlight the current state of social attention research. An empirical and a meta-analytic work focused on the current state and the generalizability of gaze cueing. These studies showed that gaze cueing is a robust phenomenon. Still, some processes are not well understood. From a meta-analytic perspective, a surprisingly large variety of gaze cueing effects that common moderators cannot explain were found in the published literature. On top, both studies provided evidence that gaze cueing effects occur regardless of the naturalistic properties within controlled laboratory studies. However, several lines of research show that findings do not generalize to real interactions with persons (e.g., Laidlaw et al., 2011).

These diverging results can be closely investigated with VR. The third empirical study in the present work demonstrates the utility of VR in social attention research. The experimental design included a close approximation of the real world in VR within the laboratory. However, differences in social attention between the laboratory and the real world were still evident.

These are in line with previous research, albeit providing a stricter test to the core hypothesis.

After all, explaining the origins of differences found in the social attention literature will continue to be the main challenge of future research. However, virtual reality offers excellent opportunities to open up *traditional paradigms* and examine critical moderators. With VR, experiments can include holistic interactions, and ecological validity can be manipulated continuously. A variety of insightful and promising experimental paradigms can follow from these two directions, a few of which are outlined above for gaze cueing. The concept of ecological validity, however, should be more sharply defined for future research. As it is used today, it will soon lose a functional meaning as it will not capture relevant properties of conducted VR studies satisfactorily. Nevertheless, with VR, the window into social cognition is open wider than ever.

References

Aaronson, D., Grupsmith, E., and Aaronson, M. (1976). The impact of computers on cognitive psychology. *Behavior Research Methods & Instrumentation*, 8(2):129–138.

Anderson, S. F. and Maxwell, S. E. (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, 21(1):1–12.

Argyle, M. and Cook, M. (1976). *Gaze and Mutual Gaze.* Cambridge U Press.

Argyle, M. and Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 28(3):289.

Argyle, M. and Ingham, R. (1972). Gaze, mutual gaze, and proximity. *Semiotica*, 6(1).

Azarian, B., Buzzell, G. A., Esser, E. G., Dornstauder, A., and Peterson, M. S. (2017). Averted body postures facilitate orienting of the eyes. *Acta Psychologica*, 175:28–32.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63(1):1–29.

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

Bakker, M., van Dijk, A., and Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science*, 7(6):543–554.

Baron-Cohen, S. (1991). Precursors to a theory of mind: Understandingattention in others. In Whiten, A., editor, *Natural theories of mind:Evolution, development and simulation of everyday mindreading*, chapter 16, pages 233–251. Oxford: Basil Blackwell.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Barresi, J. and Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, 19(1):107–122.

Bayless, S. J., Glover, M., Taylor, M. J., and Itier, R. J. (2011). Is it in the eyes? dissociating the role of emotion and perceptual features of emotionally expressive faces in modulating orienting to eye gaze. *Visual Cognition*, 19(4):483–510.

Bayliss, A. P., Frischen, A., Fenske, M. J., and Tipper, S. P. (2007). Affective Evaluations of Objects Are Influenced By Observed Gaze Direction and Emotional expression. *Cognition*, 104(3):644–653.

Bayliss, A. P., Murphy, E., Naughtin, C. K., Kritikos, A., Schilbach, L., and Becker, S. I. (2013). "gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General*, 142(1):76–92.

Bayliss, A. P., Paul, M. A., Cannon, P. R., and Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review*, 13(6):1061–1066.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., and Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, 92(2):179–190.

Belopolsky, A. V., Kramer, A. F., and Theeuwes, J. (2008). The role of awareness in processing of oculomotor capture: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 20(12):2285–2297.

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., and Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, 17(22):2537–2550.

Birmingham, E., Bischof, W. F., and Kingstone, A. (2008a). Gaze selection in complex social scenes. *Visual Cognition*, 16(2-3):341–355.

Birmingham, E., Bischof, W. F., and 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.

Birmingham, E., Bischof, W. F., and Kingstone, A. (2009). Saliency Does Not Account for Fixations To Eyes Within Social scenes. *Vision Research*, 49(24):2992–3000.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2):103–124.

Böckler, A., Knoblich, G., and Sebanz, N. (2011). Observing Shared Attention Modulates Gaze following. *Cognition*, 120(2):292–298.

Böckler, A., van der Wel, R. P. R. D., and Welsh, T. N. (2014). Catching eyes: Effects of social and nonsocial cues on attention capture. *Psychological Science*, 25(3):720–727.

Boll, S., Bartholomaeus, M., Peter, U., Lupke, U., and Gamer, M. (2016). Attentional Mechanisms of Social Perception Are Biased in Social phobia. *Journal of Anxiety*

Disorders, 40:83–93.

Boothroyd, L. G., Jones, B. C., Burt, D. M., DeBruine, L. M., and Perrett, D. I. (2008). Facial correlates of sociosexuality. *Evolution and Human Behavior*, 29(3):211–218.

Borji, A. and Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):185–207.

Bristow, D., Rees, G., and Frith, C. D. (2007). Social interaction modifies neural response to gaze shifts. *Social Cognitive and Affective Neuroscience*, 2(1):52–61.

Brooks, R. and Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6):535–543.

Brunswik, E. (1947). Systematic and representative design of psychological experiments. In *Proceedings of the Berkeley symposium on mathematical statistics and probability*, pages 143–202.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4):297–312.

Cañigueral, R., Hamilton, A., and Ward, J. A. (2018). Don't look at me, i'm wearing an eyetracker! In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers.* ACM.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13):1484–1525.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., and Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, 2(2):115–144.

Caruana, N., de Lissa, P., and McArthur, G. (2017). Beliefs About Human Agency Influence the Neural Processing of Gaze During Joint attention. *Social Neuroscience*, 12(2):194–206.

Castelhano, M. S., Wieth, M., and Henderson, J. M. (2007). I See What You See: Eye Movements in Real-World Scenes Are Affected By Perceived Direction of gaze. In *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint*, volume 3, pages 251–262. Springer Berlin Heidelberg.

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

Cerf, M., Harel, J., Einhäuser, W., and Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Advances in neural information processing systems*, 20:1–7.

Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 29(4):387–396.

Clark, H. H. and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1):62–81.

Cole, G. G., Smith, D. T., and Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception, & Psychophysics*, 77(4):1105–1115.

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716–aac4716.

Collie, A., Maruff, P., Yucel, M., Danckert, J., and Currie, J. (2000). Spatiotemporal distribution of facilitation and inhibition of return arising from the reflexive orienting of covert attention. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6):1733–1745.

Conty, L., George, N., and Hietanen, J. K. (2016). Watching Eyes effects: When others meet the self. *Consciousness and Cognition*, 45:184–197.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12(11):671–684.

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

Dalmaso, M., Alessi, G., Castelli, L., and Galfano, G. (2020a). Eye contact boosts the reflexive component of overt gaze following. *Scientific Reports*, 10(1).

Dalmaso, M., Castelli, L., and Galfano, G. (2020b). Social modulators of gaze-mediated orienting of attention: A review. *Psychonomic Bulletin & Review*, 27(5):833–855.

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., and Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences*, page 201820553.

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1):193–222.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial orienting. *Visual Cognition*, 6(5):509–540.

Eerland, A., Sherrill, A. M., Magliano, J. P., Zwaan, R. A., Arnal, J. D., Aucoin, P., Berger, S. A., Birt, A. R., Capezza, N., Carlucci, M., Crocker, C., Ferretti, T. R., Kibbe, M. R., Knepp, M. M., Kurby, C. A., Melcher, J. M., Michael, S. W., Poirier, C., and Prenoveau, J. M. (2016). Registered replication report. *Perspectives on Psychological Science*, 11(1):158–171.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109):629–634.

Eimer, M. and Holmes, A. (2002). An ERP study on the time course of emotional face processing. *NeuroReport*, 13(4):427–431.

Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., and König, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, 18(3):267–297.

Ellsworth, P. and Ross, L. (1975). Intimacy in response to direct gaze. *Journal of Experimental Social Psychology*, 11(6):592–613.

Ellsworth, P. C., Carlsmith, J. M., and Henson, A. (1972). The stare as a stimulus to flight in human subjects: A series of field experiments. *Journal of personality and social psychology*, 21(3):302.

Emery, N. J. (2000). The Eyes Have It: the Neuroethology, Function and Evolution of Social gaze. *Neuroscience and Biobehavioral Reviews*, 24(6):581–604.

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

Erp, S. V., Verhagen, J., Grasman, R. P. P. P., and Wagenmakers, E.-J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in psychological bulletin from 1990–2013. *Journal of Open Psychology Data*, 5(1):4.

Fichtenholtz, H. M., Hopfinger, J. B., Graham, R., Detwiler, J. M., and LaBar, K. S. (2007). Happy and fearful emotion in cues and targets modulate event-related potential indices of gaze-directed attentional orienting. *Social Cognitive and Affective Neuroscience*, 2(4):323–333.

Fischer, B. and Ramsperger, E. (1984). Human express saccades: Extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, 57(1):191–195.

Flechsenhar, A., Rösler, L., and Gamer, M. (2018). Attentional selection of social features persists despite restricted bottom-up information and affects temporal viewing dynamics. *Scientific Reports*, 8(1).

Flechsenhar, A. F. and Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PLOS ONE*, 12(8):e0183799.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., and Benson, V. (2008). Rapid Detection of Person Information in a Naturalistic scene. *Perception*, 37(4):571–583.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., and Kingstone, A. (2010). Gaze Allocation in a Dynamic Situation: Effects of Social Status and speaking. *Cognition*, 117(3):319–331.

Foulsham, T., Walker, E., and Kingstone, A. (2011). The Where, What and When of Gaze Allocation in the Lab and the Natural environment. *Vision Research*, 51(17):1920–1931.

Fox, E., Mathews, A., Calder, A. J., and Yiend, J. (2007). Anxiety and sensitivity to gaze direction in emotionally expressive faces. *Emotion*, 7(3):478–486.

Freedman, E. G. (2008). Coordination of the eyes and head during visual orienting. *Experimental Brain Research*, 190(4):369–387.

Freeth, M., Foulsham, T., and Kingstone, A. (2013). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLoS ONE*, 8(1):e53286.

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., and von Gontard, A. (2007). Evaluation Der Deutschen Version Des Autismus-Spektrum-Quotienten (AQ) - Die Kurzversion AQ-K. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36(4):280–289.

Friedman, D., Pizarro, R., Or-Berkers, K., Neyret, S., Pan, X., and Slater, M. (2014). A method for generating an illusion of backwards time travel using immersive virtual reality – an exploratory study. *Frontiers in Psychology*, 5.

Friesen, C. K. and Kingstone, A. (1998). The Eyes Have It! Reflexive Orienting Is Triggered by Nonpredictive Gaze. *Psychonomic Bulletin & Review*, 5(3):490–495.

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

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007a). Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual differences. *Psychological Bulletin*, 133(4):694–724.

Frischen, A., Smilek, D., Eastwood, J. D., and Tipper, S. P. (2007b). Inhibition of Return in Response To Gaze Cues: the Roles of Time Course and Fixation cue. *Visual Cognition*, 15(8):881–895.

Frischen, A. and Tipper, S. P. (2004). Orienting Attention Via Observed Gaze Shift Evokes Longer Term Inhibitory Effects: Implications for Social Interactions, Attention, and Memory. *Journal of Experimental Psychology: General*, 133(4):516–533.

Gallup, A. C., Hale, J. J., Sumpter, D. J. T., Garnier, S., Kacelnik, A., Krebs, J. R., and Couzin, I. D. (2012). Visual Attention and the Acquisition of Information in Human crowds. *Proceedings of the National Academy of Sciences*, 109(19):7245–7250.

Gayzur, N. D., Langley, L. K., Kelland, C., Wyman, S. V., Saville, A. L., Ciernia, A. T., and Padmanabhan, G. (2013). Reflexive orienting in response to short- and long-duration gaze cues in young, young-old, and old-old adults. *Attention, Perception, & Psychophysics*, 76(2):407–419.

Genovesio, A., Brasted, P. J., and Wise, S. P. (2006). Representation of future and previous spatial goals by separate neural populations in prefrontal cortex. *Journal of Neuroscience*, 26(27):7305–7316.

Gleser, L. J. and Olkin, I. (2009). Stochastically dependent effect sizes. In Cooper, H., Hedges, L. V., and Valentine, J. C., editors, *The handbook of research synthesis and meta-analysis*, chapter 19, pages 357–376. New York: Russell Sage Foundation, 2nd edition.

Gobel, M. S., Kim, H. S., and Richardson, D. C. (2015). The Dual Function of Social gaze. *Cognition*, 136:359–364.

Graham, R., Friesen, C. K., Fichtenholtz, H. M., and LaBar, K. S. (2010). Modulation of reflexive orienting to gaze direction by facial expressions. *Visual Cognition*, 18(3):331–368.

Green, J. J., Gamble, M. L., and Woldorff, M. G. (2013). Resolving conflicting views: Gaze and arrow cues do not trigger rapid reflexive shifts of attention. *Visual Cognition*, 21(1):61–71.

Greenhouse, S. W. and Geisser, S. (1959). *Psychometrika*, 24(2):95–112.

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35(1):96–111.

Großekathöfer, J. D., Seis, C., and Gamer, M. (in press). Reality in a sphere: A direct comparison of social attention in the laboratory and the real world. *Behavior Research Methods Instrumentation.* in press.

Großekathöfer, J. D., Suchotzki, K., and Gamer, M. (2020). Gaze cueing in naturalistic scenes under top-down modulation – effects on gaze behaviour and memory performance. *Visual Cognition*, 28(2):135–147.

Gurevitch, J., Koricheva, J., Nakagawa, S., and Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555(7695):175–182.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., Ridder, D. T. D. D., Dewitte, S., Elson, M., Evans, J. R., Fay, B. A., Fennis, B. M., Finley, A., Francis, Z., Heise, E., Hoemann, H., Inzlicht, M., Koole, S. L., Koppel, L., Kroese, F., Lange, F., Lau, K., Lynch, B. P., Martijn, C., Merckelbach, H., Mills, N. V., Michirev, A., Miyake, A., Mosser, A. E., Muise, M., Muller, D., Muzi, M., Nalis, D., Nurwanti, R., Otgaar, H., Philipp, M. C., Primoceri, P., Rentzsch, K., Ringos, L., Schlinkert, C., Schmeichel, B. J., Schoch, S. F., Schrama, M., Schütz, A., Stamos, A., Tinghög, G., Ullrich, J., vanDellen, M., Wimbarti, S., Wolff, W., Yusainy, C., Zerhouni, O., and Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4):546–573.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6):223–233.

Hayward, D. A., Voorhies, W., Morris, J. L., Capozzi, F., and Ristic, J. (2017). Staring Reality in the Face: a Comparison of Social Attention Across Laboratory and Real World Measures Suggests Little Common ground. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 71(3):212–225.

Hedge, C., Powell, G., and Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3):1166–1186.

Hedges, L. V., Tipton, E., and Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1):39–65.

Hedges, L. V. and Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods*, 3(4):486.

Hein, G. and Knight, R. T. (2008). Superior temporal sulcus—it's my area: Or is it? *Journal of Cognitive Neuroscience*, 20(12):2125–2136.

Helmholtz, H. v. (1896). *Handbuch Der Physiologischen Optik, Dritter Abschnitt.* Hamburg: Voss.

Heron, J. (1970). The phenomenology of social encounter: The gaze. *Philosophy and Phenomenological Research*, 31(2):243.

Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., and Heeger, D. J. (2010). When size matters: Attention affects performance by contrast or response gain. *Nature Neuroscience*, 13(12):1554–1559.

Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., and Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 71(3):226–242.

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., and Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? a survey among 124 researchers. *Royal Society Open Science*, 5(8):180502.

Hietanen, J. (2002). Social attention orienting integrates visual information from head and body orientation. *Psychological Research*, 66(3):174–179.

Hietanen, J. K. (1999). Does Your Gaze Direction and Head Orientation Shift My Visual attention? *NeuroReport*, 10(16):3443–3447.

Hietanen, J. K. and Leppänen, J. M. (2003). Does facial expression affect attention orienting by gaze direction cues? *Journal of Experimental Psychology: Human Perception and Performance*, 29(6):1228–1243.

Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., and Ruuhiala, H. J. (2008).

Seeing Direct and Averted Gaze Activates the Approach-Avoidance Motivational Brain systems. *Neuropsychologia*, 46(9):2423–2430.

Hietanen, J. K., Myllyneva, A., Helminen, T. M., and Lyyra, P. (2016). The Effects of Genuine Eye Contact on Visuospatial and Selective attention. *Journal of Experimental Psychology: General*, 145(9):1102–1106.

Holleman, G. A., Hooge, I. T. C., Kemner, C., and Hessels, R. S. (2020). The Reality of "real-life" Neuroscience: A Commentary on Shamay-Tsoory and Mendelsohn (2019). *Perspectives on Psychological Science*, page 174569162091735.

Hollingworth, A. and Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1):113.

Holmes, A., Mogg, K., Garcia, L. M., and Bradley, B. P. (2010). Neural activity associated with attention orienting triggered by gaze cues: A study of lateralized ERPs. *Social Neuroscience*, 5(3):285–295.

Hooker, C. and Park, S. (2005). You must be looking at me: The nature of gaze perception in schizophrenia patients. *Cognitive Neuropsychiatry*, 10(5):327–345.

Horn, A., Mergenthaler, L., and Gamer, M. (2021). Situational and personality determinants of social attention in a waiting room scenario. *Visual Cognition*, pages 1–14.

Howell, A. N., Zibulsky, D. A., Srivastav, A., and Weeks, J. W. (2015). Relations among social anxiety, eye contact avoidance, state anxiety, and perception of interaction performance during a live conversation. *Cognitive Behaviour Therapy*, 45(2):111–122.

Hungr, C. J. and Hunt, A. R. (2012). Rapid communication: Physical self-similarity enhances the gaze-cueing effect. *Quarterly Journal of Experimental Psychology*, 65(7):1250–1259.

Intriligator, J. and Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, 43(3):171–216.

Ioannidou, F., Hermens, F., and Hodgson, T. L. (2016). The central bias in day-to-day viewing. *Journal of Eye Movement Research.*

Itti, L. and Koch, C. (2001). Computational Modelling of Visual attention. *Nature reviews. Neuroscience*, 2(3):194–203.

Jaccard, J. and Jacoby, J. (2020). *Theory Construction and Model-building Skills: A*

Practical Guide for Social Scientists. Guilford Publications, 2 edition.

Jaegher, H. D., Paolo, E. D., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10):441–447.

James, W. (1890). *Principles of Psychology*, volume 2. Henry Holt and Company.

Jellema, T., Baker, C. I., Wicker, B., and Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain and Cognition*, 44(2):280–302.

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

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye. In Long, J. and Baddeley, A., editors, *Attention and performance*, chapter 11, pages 187–203. Hillsdale, N.J., Erlbaum, ix edition.

Judd, C. M., Westfall, J., and Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1):54–69.

Kalaian, H. A. and Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, 1(3):227–235.

Kawakami, K., Williams, A., Sidhu, D., Choma, B. L., Rodriguez-Bailón, R., Cañadas, E., Chung, D., and Hugenberg, K. (2014). An Eye for the I: Preferential Attention To the Eyes of Ingroup members. *Journal of Personality and Social Psychology*, 107(1):1–20.

Kawase, S. (2013). Gazing behavior and coordination during piano duo performance. *Attention, Perception, & Psychophysics*, 76(2):527–540.

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

Khayat, P. S. (2006). Attention lights up new object representations before the old ones fade away. *Journal of Neuroscience*, 26(1):138–142.

Kim, J. S.-C., Vossel, G., and Gamer, M. (2013). Effects of emotional context on memory for details: The role of attention. *PLoS ONE*, 8(10):e77405.

King, D., Rowe, A., and Leonards, U. (2011). I trust you; hence i like the things you look at: Gaze cueing and sender trustworthiness influence object evaluation. *Social Cognition*, 29(4):476–485.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1):78–100.

Knapp, G. and Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17):2693–2710.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual review of neuroscience*, 30(1):57–78.

Kobayashi, H. and Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, 387(6635):767–768.

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

Koster, E. H. W., Leyman, L., Raedt, R. D., and Crombez, G. (2006). Cueing of visual attention by emotional facial expressions: The influence of individual differences in anxiety and depression. *Personality and Individual Differences*, 41(2):329–339.

Kowler, E. (2011). Eye movements: The past 25years. *Vision Research*, 51(13):1457–1483.

Kühberger, A., Fritz, A., and Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE*, 9(9):e105825.

Kuzmanovic, B., Georgescu, A. L., Eickhoff, S. B., Shah, N. J., Bente, G., Fink, G. R., and Vogeley, K. (2009). Duration matters: Dissociating neural correlates of detection and evaluation of social gaze. *NeuroImage*, 46(4):1154–1163.

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., and Kingstone, A. (2011). Potential Social Interactions Are Important To Social attention. *Proceedings of the National Academy of Sciences*, 108(14):5548–5553.

Laidlaw, K. E. W., Risko, E. F., and 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(JUNE):1132–1143.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and Validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8):1377–1388.

Langton, S. R. H. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology: Section A*, 53(3):825–845.

Langton, S. R. H. and Bruce, V. (1999). Reflexive Visual Orienting in Response To the Social Attention of others. *Visual Cognition*, 6(5):541–567.

Langton, S. R. H. and Bruce, V. (2000). You Must See the Point: Automatic Processing of Cues To the Direction of Social attention. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2):747–757.

Langton, S. R. H., McIntyre, A. H., Hancock, P. J. B., and Leder, H. (2018). Saccades and smooth pursuit eye movements trigger equivalent gaze-cued orienting effects. *Quarterly Journal of Experimental Psychology*, 71(9):1860–1872.

Lassalle, A. and Itier, R. J. (2014). Autistic traits influence gaze-oriented attention to happy but not fearful faces. *Social Neuroscience*, 10(1):70–88.

Lassalle, A. and Itier, R. J. (2015). Emotional modulation of attention orienting by gaze varies with dynamic cue sequence. *Visual Cognition*, 23(6):720–735.

Lawson, R. P. and Calder, A. J. (2016). The "where" of social attention: Head and body direction aftereffects arise from representations specific to cue type and not direction alone. *Cognitive Neuroscience*, 7(1-4):103–113.

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

Little, A. C. and Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology*, 98(1):111–126.

Liuzza, M. T., Cazzato, V., Vecchione, M., Crostella, F., Caprara, G. V., and Aglioti, S. M. (2011). Follow my eyes: The gaze of politicians reflexively captures the gaze of ingroup voters. *PLoS ONE*, 6(9):e25117.

López-López, J. A., den Noortgate, W. V., Tanner-Smith, E. E., Wilson, S. J., and Lipsey, M. W. (2017). Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: A monte carlo simulation. *Research Synthesis Methods*, 8(4):435–450.

Lord, F., Novick, M. R., and Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley Pub. Co, Reading, Mass.

Lundqvist, D., Esteves, F., and Ohman, A. (1999). The face of wrath: Critical features for conveying facial threat. *Cognition & Emotion*, 13(6):691–711.

MacDorman, K. F. and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3):297–337.

Mares, I., Smith, M. L., Johnson, M. H., and Senju, A. (2016). Direct Gaze Facilitates Rapid Orienting To Faces: Evidence From Express Saccades and Saccadic potentials. *Biological Psychology*, 121:84–90.

Mathews, A., Fox, E., Yiend, J., and Calder, A. (2003). The face of fear: Effects of eye gaze and emotion on visual attention. *Visual Cognition*, 10(7):823–835.

Matthis, J. S., Yates, J. L., and Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, 28(8):1224–1233.e5.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language*, 94:305–315.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9(2):147–163.

Milgram, S., Bickman, L., and Berkowitz, L. (1969). Note on the Drawing Power of Crowds of Different size. *Journal of Personality and Social Psychology*, 13(2):79–82.

Moeyaert, M., Ugille, M., Beretvas, S. N., Ferron, J., Bunuan, R., and den Noortgate, W. V. (2016). Methods for dealing with multiple outcomes in meta-analysis:a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, 20(6):559–572.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., and Group, P. R. I. S. M. A.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1):1.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38(4):379–387.

Moore, T. and Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Annual Review of Psychology*, 68(1):47–72.

Mori, M. (1970). The uncanny valley. *Energy*, 7(4):33–35.

Mori, M., MacDorman, K., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100.

Müller, H. J. and Rabbitt, P. M. (1989). Reflexive and Voluntary Orienting of Visual

Attention: Time Course of Activation and Resistance To interruption. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):315–330.

Nakagawa, S., Johnson, P. C. D., and Schielzeth, H. (2017). The coefficient of determination r 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, 14(134):20170213.

Nasiopoulos, E., Risko, E. F., and Kingstone, A. (2015). Social attention, social presence, and the dual function of gaze. In *The Many Faces of Social Attention*, pages 129–155. Springer International Publishing.

Nation, K. and Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? is it automatic? is it social? *Development and Psychopathology*, 20(1):79–97.

Neath, K., Nilsen, E. S., Gittsovich, K., and Itier, R. J. (2013). Attention orienting by gaze and facial expressions across development. *Emotion*, 13(3):397–408.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., and Bateson, M. (2013). The watching eyes effect in the dictator game: It's not how much you give, it's being seen to give something. *Evolution and Human Behavior*, 34(1):35–40.

Neuberg, S. L. and Schaller, M. (2015). Evolutionary social cognition. In Mikulincer, M., Shaver, P. R., Borgida, E., and Bargh, J. A., editors, *APA handbook of personality and social psychology*, volume Volume 1: Attitudes and social cognition. American Psychological Association.

Newport, R. and Howarth, S. (2009). Short article: Social gaze cueing to auditory locations. *Quarterly Journal of Experimental Psychology*, 62(4):625–634.

Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., and Hessels, R. S. (2017). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, 50(1):213–227.

Noortgate, W. V. D. and Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, 63(5):765–790.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1):1–18.

Nummenmaa, L. and Calder, A. J. (2009). Neural Mechanisms of Social attention. *Trends in Cognitive Sciences*, 13(3):135–143.

Oertel, C., Włodarczak, M., Edlund, J., Wagner, P., and Gustafson, J. (2012). Gaze patterns in turn-taking. In *Thirteenth annual conference of the international speech communication association*.

Öhman, A., Lundqvist, D., and Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, 80(3):381–396.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? the effects of addressee location on representational gestures. *Journal of Memory and Language*, 46(4):688–704.

Parsons, T., Gaggioli, A., and Riva, G. (2017). Virtual Reality for Research in Social Neuroscience. *Brain Sciences 2017, Vol. 7, Page 42*, 7(4):42.

Pecchinenda, A., Pes, M., Ferlazzo, F., and Zoccolotti, P. (2008). The combined effect of gaze direction and facial expression on cueing spatial attention. *Emotion*, 8(5):628–634.

Perez-Osorio, J., Müller, H. J., Wiese, E., and Wykowska, A. (2015). Gaze following is modulated by expectations regarding others' action goals. *PLoS ONE*, 10(11):e0143614.

Perrett, D. I. and Emery, N. J. (1994). Understanding the intentions of others from visual signals: Neurophysiological evidence. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 13(5):683–694.

Perrett, D. I., Hietanen, J. K., Oram, M. W., Benson, P. J., and Rolls, E. T. (1992). Organization and Functions of Cells Responsive To Faces in the Temporal Cortex [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 335(1273):23–30.

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., and Vogeley, K. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, 101:124–137.

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., and Kastner, S. (2009). Neural representations of faces and body parts in macaque and human cortex: A comparative fMRI study. *Journal of Neurophysiology*, 101(5):2581–2600.

Pitcher, D. and Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences*, 25(2):100–110.

Ponari, M., Trojano, L., Grossi, D., and Conson, M. (2013). "avoiding or approaching eyes"? introversion/extraversion affects the gaze-cueing effect. *Cognitive Processing*, 14(3):293–299.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1):3–25.

Posner, M. I. and Cohen, Y. (1984). Components of visual orienting. In Bouma, H. and Bouwhuis, D., editors, *Attention and Performance*, chapter 32, pages 531–556. Lawrence Erlbaum Associates Ltd.

Posner, M. I., Rafal, R. D., Choate, L. S., and Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, 2(3):211–228.

Posner, M. I. and Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, 58(1):1–23.

Pourtois, G. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex*, 14(6):619–633.

Prazak, E. R. and Burgund, E. D. (2014). Keeping it real: Recognizing expressions in real compared to schematic faces. *Visual Cognition*, 22(5):737–750.

Raudenbush, S. W., Becker, B. J., and Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, 103(1):111–120.

Reutter, M., Hewig, J., Wieser, M. J., and Osinsky, R. (2019). Attentional bias modification in social anxiety: Effects on the n2pc component. *Behaviour Research and Therapy*, 120:103404.

Ricciardelli, P., Baylis, G., and Driver, J. (2000). The positive and negative of human expertise in gaze perception. *Cognition*, 77(1):B1–B14.

Richardson, D. C., Dale, R., and Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, 18(5):407–413.

Risko, E. F. and Kingstone, A. (2011). Eyes wide shut: Implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, 73(2):291–296.

Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., and 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.

Risko, E. F., Richardson, D. C., and Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science: Real-World Social Attention and the Dual Function of Gaze.

Current Directions in Psychological Science, 25(1):70–74.

Ristic, J. and Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition*, 94(3):B55–B65.

Roepstorff, A. and Frith, C. (2004). What's at the top in the top-down control of action? Script-sharing and 'top-top' control of action in cognitive experiments. *Psychological Research*, 68(2):189–198.

Rösler, L., End, A., and Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLOS ONE*, 12(7):e0182037.

Rösler, L., Göhring, S., Strunz, M., and Gamer, M. (2021). Social anxiety is associated with heart rate but not gaze behavior in a real social interaction. *Journal of Behavior Therapy and Experimental Psychiatry*, 70:101600.

Rothstein, H. R., Sutton, A. J., and Borenstein, M. (2005). *Publication Bias in Meta-analysis*. John Wiley Sons.

Rubo, M. and Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Scientific Reports*, 8(1).

Rubo, M. and Gamer, M. (2021). Stronger reactivity to social gaze in virtual reality compared to a classical laboratory environment. *British Journal of Psychology*, 112(1):301–314.

Rubo, M., Huestegge, L., and Gamer, M. (2020). Social anxiety modulates visual exploration in real life – but not in the laboratory. *British Journal of Psychology*, 111(2):233–245.

Samuel, A. G. and Kat, D. (2003). Inhibition of return: A graphical meta-analysis of its time course and an empirical test of its temporal and spatial properties. *Psychonomic Bulletin & Review*, 10(4):897–906.

Sanchez-Vives, M. V. and Slater, M. (2005). Opinion: From Presence To Consciousness Through Virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339.

Schmidt, B. K., Vogel, E. K., Woodman, G. F., and Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics*, 64(5):754–763.

Schmidt, F. L., Oh, I.-S., and Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62(1):97–128.

Schmuckler, M. A. (2001). What is ecological validity? a dimensional analysis. *Infancy*, 2(4):419–436.

Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, 19(7):595–605.

Schubert, T. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, 15:69–71.

Schultz, J. and Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, 194(3):465–475.

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2):70–76.

Seyama, J. and Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, 16(4):337–351.

Shamay-Tsoory, S. G. and Mendelsohn, A. (2019). Real-life neuroscience: An ecological approach to brain and behavior research. *Perspectives on Psychological Science*, 14(5):841–859.

Shepherd (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience.*

Shi, J., Weng, X., He, S., and Jiang, Y. (2010). Biological motion cues trigger reflexive attentional orienting. *Cognition*, 117(3):348–354.

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences*, 23(8):699–714.

Squire, L. (1993). The structure and organization of memory. *Annual Review of Psychology*, 44(1):453–495.

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., and Hoyer, J. (1999). Die Erfassung sozialer Phobie durch die Social Interaction Anxiety Scale (SIAS) und die Social Phobia Scale (SPS). *Zeitschrift für Klinische Psychologie und Psychotherapie*, 28(1):28–36.

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science*, 8(5):581–591.

Stanley, T. D., Carter, E. C., and Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, 144(12):1325–1346.

Stanley, T. D. and Doucouliagos, H. (2013). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1):60–78.

Stewart, E. E. M., Valsecchi, M., and Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, 20(12):2.

Strachan, J. W. A., Kirkham, A. J., Manssuer, L. R., Over, H., and Tipper, S. P. (2017). Incidental learning of trust from eye-gaze: Effects of race and facial trustworthiness. *Visual Cognition*, 25(7-8):802–814.

Strick, M., Holland, R. W., and van Knippenberg, A. (2008). Seductive eyes: Attractiveness and direct gaze increase desire for associated objects. *Cognition*, 106(3):1487–1496.

Takao, S., Yamani, Y., and Ariga, A. (2018). The gaze-cueing effect in the united states and japan: Influence of cultural differences in cognitive strategies on control of attention. *Frontiers in Psychology*, 8.

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.

Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S., and Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology*, 19(15):1274–1277.

Thönes, S. and Hecht, H. (2016). How long did you look at me? the influence of gaze direction on perceived duration and temporal sensitivity. *Perception*, 45(6):612–630.

Tipples, J. (2006). Fear and fearfulness potentiate automatic orienting to eye gaze. *Cognition & Emotion*, 20(2):309–320.

Tipples, J. (2007). Wide eyes and an open mouth enhance facial threat. *Cognition & Emotion*, 21(3):535–557.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5):675–691.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.

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

Tso, I. F., Mui, M. L., Taylor, S. F., and Deldin, P. J. (2012). Eye-contact perception in schizophrenia: Relationship with symptoms and socioemotional functioning. *Journal of Abnormal Psychology*, 121(3):616–627.

Vabalas, A. and Freeth, M. (2015). Brief report: Patterns of eye movements in face to face conversation are associated with autistic traits: Evidence from a student sample. *Journal of Autism and Developmental Disorders*, 46(1):305–314.

Valtakari, N. V., Hooge, I. T. C., Viktorsson, C., Nyström, P., Falck-Ytter, T., and Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods.*

Vuilleumier, P., George, N., Lister, V., Armony, J., and Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory. *Visual Cognition*, 12(1):85–101.

Wang, Z. and Klein, R. M. (2010). Searching for inhibition of return in visual search: A review. *Vision Research*, 50(2):220–228.

Watzlawick, P. (1967). *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes.* Norton, New York.

Westfall, J. (2016). Pangea: Power analysis for general anova designs.

Wiese, E., Wykowska, A., Zwickel, J., and Müller, H. J. (2012). I see what you mean: How attentional selection is shaped by ascribing intentions to others. *PLoS ONE*, 7(9):e45391.

Wieser, M. J., Pauli, P., Alpers, G. W., and Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. *Journal of anxiety disorders*, 23(1):93–103.

Wirth, J. H., Sacco, D. F., Hugenberg, K., and Williams, K. D. (2010). Eye gaze as relational evaluation: Averted eye gaze leads to feelings of ostracism and relational devaluation. *Personality and Social Psychology Bulletin*, 36(7):869–882.

Wu, D. W. L., Bischof, W. F., and Kingstone, A. (2013). Looking while eating: The importance of social context to social attention. *Scientific Reports*, 3(1).

Wundt, W. (1912). *An Introduction to Psychology.* MacMillan Co.

Wykowska, A., Wiese, E., Prosser, A., and Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS One*, 9(4):e94339.

Xu, B. and Tanaka, J. W. (2014). The cued recognition task: Dissociating the abrupt onset effect from the social and arrow cueing effect. *Attention, Perception, & Psychophysics*, 77(1):97–110.

Yarbus, A. L. (1967). *Eye Movements and Vision.* Springer US, New York.

Yee, N. and Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33(3):271–290.

Yeshurun, Y. and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396(6706):72–75.

Yeshurun, Y. and Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, 39(2):293–306.

Zaki, J. and Ochsner, K. (2009). The need for a cognitive neuroscience of naturalistic social cognition. *Annals of the New York Academy of Sciences*, 1167(1):16–30.

Zwickel, J. and Võ, M. L.-H. (2010). How the presence of persons biases eye movements. *Psychonomic Bulletin & Review*, 17(2):257–262.

Appendix

Appendix A

Supplementary material

Gaze cueing in naturalistic scenes under top-down modulation – Effects on gaze behavior and memory performance

A.1 Regions of Interest (ROIs)

List of all ROIs

In our study, each scene contained two central objects. Below is the full list of all objects presented in the scenes. Note that objects were counterbalanced regarding cueing (i.e., being cued or uncued) and image side (i.e., appearing on the left or right side of the scene).

stim_id	1	2
01	orange	melon
02	punches	roller stick
03	bike helmet	stethoscope
04	rack	ventilator
05	*blackroll*	hair dryer
06	fruit bowl	globe

stim_id	1	2
07	toaster	thermos flask
08	marmalade	watch
09	cup	water bottle
10	brush	glasses
11	chair	bucket
12	binoculars	camera
13	book	laptop
14	headphones	folder
15	water glass	towel
16	baggage	cap
17	strawberries	shoe
18	ukulele	hand brush
19	drum	hat
20	bowl	cushions
21	pencil	smartphone
22	feather	elephant
23	saw	hammer
24	flower	jug
25	coffee grinder	candle
26	bulb	flashlight

ROI size, position and distance

Each scene contained five regions: Background, Head, Body, Cued object, Uncued object. In the following tables, a complete overview is given regarding average ROI size in pixels, centimeters and degrees of visual angle, as well as ROI position and distances between ROIs on image.

Table A.2: Sizes of the regions of interest in pixels.

roi_id	m roi_size	sd roi_size
background	1,061,723.57	77,362.12
body	95,367.37	50,529.36
head	24,736.62	18,381.95
object_cued	23,390.59	23,912.69
object_uncued	23,581.86	24,833.02

Note. Mean [m] and standard deviation [sd].

Table A.3: Width and height of regions of interest in pixels.

roi_id	m roi_w_px	sd roi_w_px	m roi_h_px	sd roi_h_px
background	1,279.00	0.00	959.00	0.00
body	347.88	110.12	422.98	143.93
head	161.53	65.81	192.71	64.07
object_cued	199.86	114.17	175.52	97.51
object_uncued	194.25	106.76	174.04	95.95

Note. Mean [m] and standard deviation [sd] of the width [w] and height [h].

Table A.4: Width and height of regions of interest in centimeters on the computer screen.

roi_id	m roi_w_cm	sd roi_w_cm	m roi_h_cm	sd roi_h_cm
background	35.40	0.00	26.54	0.00
body	9.63	3.05	11.71	3.98
head	4.47	1.82	5.33	1.77
object_cued	5.53	3.16	4.86	2.70
object_uncued	5.38	2.95	4.82	2.66

Note. Mean [m] and its standard deviation [sd] of the width [w] and height [h].

Table A.5: Width and height of regions of interest in degrees of visual angle for the current viewing distance.

roi_id	m roi_w_dg	sd roi_w_dg	m roi_h_dg	sd roi_h_dg
background	36.84	0.00	28.04	0.00
body	10.34	3.25	12.55	4.23
head	4.82	1.96	5.74	1.91
object_cued	5.95	3.38	5.23	2.90
object_uncued	5.79	3.17	5.19	2.85

Note. Mean [m] and its standard deviation [sd] of the width [w] and height [h]

Table A.6: Positions (x- and y-coordinates) of the different regions of interest on the image in pixels relative to the top left corner.

roi_id	m roi_x	m roi_y
body	584.92	483.45
head	588.81	227.76
object_cued	582.57	669.46
object_uncued	581.86	671.58

Note. Mean [m] and standard deviation [sd].

Distances between the head, the cued and the uncued objects.

Table A.7: Distance between objects (cued & uncued) and the head of the depicted person in pixels, centimeters and degree of visual angle.

dist_id	m dist_px	sd dist_px	m dist_cm	sd dist_cm	m dist_dg	sd dist_dg
head_cued	549.00	150.49	15.19	4.16	16.25	4.39
head_uncued	544.68	156.02	15.07	4.32	16.12	4.55
cued_uncued	556.10	198.04	15.39	5.48	16.44	5.77

Note. Mean [m] and standard deviation [sd].

A.2 Follow-up ANOVAs on social attention

In our main analysis the three-way interaction of Group x ROI x Time was significant. As follow-up test separate ANOVAs for each time point were performed to reveal time specific patterns. Only the first ANOVA is reported in the manuscript with all relevant details due to the observed statistically significant effects including Group and ROI. Below are all ANOVA tables for each time point.

time point 1: 0 - 2 seconds

Table A.8: time point 1: 0 - 2 seconds

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	0.01	28.57	0.11	0.00	***
ROI	1.00	91.00	0.01	430.66	0.74	0.00	***
Group x ROI	1.00	91.00	0.01	12.36	0.07	0.00	***

Note. MSE: mean squared error; ges: generalized eta square; * : $p < .05$; ** : $p < .01$; *** : $p < .001$

time point 2: 2 - 4 seconds

Table A.9: time point 2: 2 - 4 seconds

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	0.00	2.42	0.02	0.12	
ROI	1.00	91.00	0.00	201.01	0.47	0.00	***
Group x ROI	1.00	91.00	0.00	2.39	0.01	0.13	

Note. MSE: mean squared error; ges: generalized eta square; * : $p < .05$; ** : $p < .01$; *** : $p < .001$

time point 3: 4 - 6 seconds

Table A.10: time point 3: 4 - 6 seconds

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	0.01	0.03	0.00	0.86	
ROI	1.00	91.00	0.00	171.70	0.43	0.00	***
Group x ROI	1.00	91.00	0.00	1.47	0.01	0.23	

Note. MSE: mean squared error; ges: generalized eta square; * : $p < .05$; ** : $p < .01$; *** : $p < .001$

time point 4: 6 - 8 seconds

Table A.11: time point 4: 6 - 8 seconds

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	0.01	0.74	0.00	0.39	
ROI	1.00	91.00	0.01	96.02	0.30	0.00	***
Group x ROI	1.00	91.00	0.01	2.35	0.01	0.13	

Note. MSE: mean squared error; ges: generalized eta square; * : $p < .05$; ** : $p < .01$; *** : $p < .001$

time point 5: 8 - 10 seconds

Table A.12: time point 5: 8 - 10 seconds

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	0.01	2.36	0.01	0.13	
ROI	1.00	91.00	0.01	71.75	0.26	0.00	***
Group x ROI	1.00	91.00	0.01	1.72	0.01	0.19	

Note. MSE: mean squared error; ges: generalized eta square; * : $p < .05$; ** : $p < .01$; *** : $p < .001$

A.3 ANOVA including all ROIs

According to our research question we only compared gaze following (cued object vs. uncued object) **or** social attention (head vs. body) in the accompanying manuscript. To compare prioritization across gaze following and social attention effects, an ANOVA with all ROIs can be calculated. This allows for a comparison between all ROIs. Below are the results of 2 x 4 ANOVAs including Group (explicit encoding vs. free viewing) and ROI (head, body, object cued, object uncued). Results indicate that the head ROI was fixated earlier, longer and more often than any of the other ROIs.

Fixation latency

Table A.13: Fixation latency

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	497,555.90	13.32	0.03	0.00	***
ROI	2.38	216.94	661,215.02	127.25	0.52	0.00	***
Group x ROI	2.38	216.94	661,215.02	8.74	0.07	0.00	***

Note. MSE: mean squared error; MSE: mean squared error; ges: generalized eta square; * : p < .05; ** : p < .01; *** : p < .001

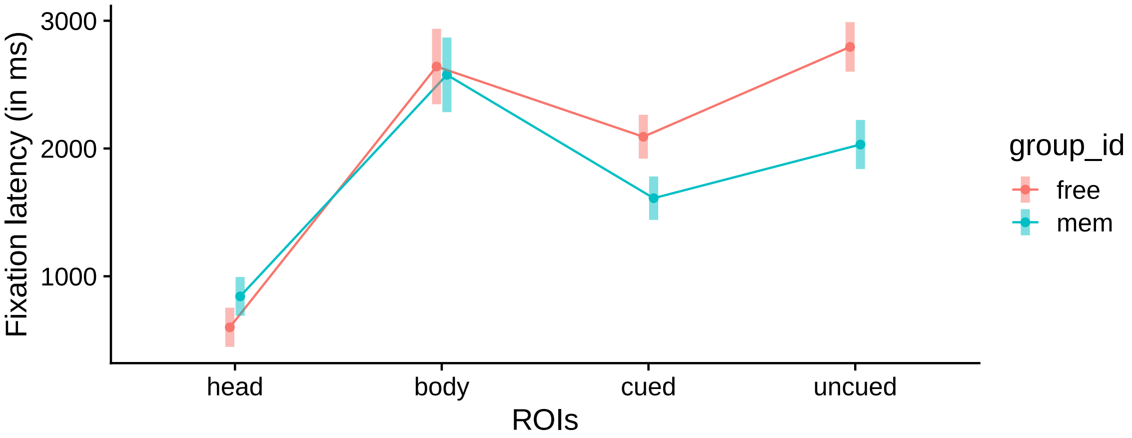

Figure A.1: Fixation latencies as a function of the region of interest and group. Error bars represent 95% confidence interval.

Fixation duration

Table A.14: Fixation duration

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	0.00	6.06	0.01	0.02	*
ROI	1.83	166.61	0.01	93.63	0.46	0.00	***
Group x ROI	1.83	166.61	0.01	8.55	0.07	0.00	***

Note. MSE: mean squared error; MSE: mean squared error; ges: generalized eta square; * : p < .05; ** : p < .01; *** : p < .001

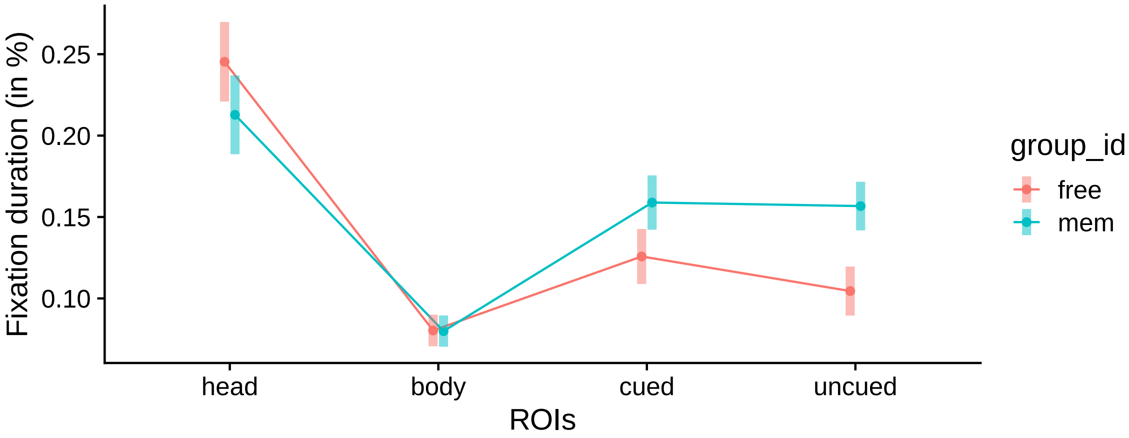

Figure A.2: Fixation duration as a function of the region of interest and group. Error bars represent 95% confidence interval.

Fixation number

Table A.15: Fixation number

effect	num Df	den Df	MSE	F	ges	p	
Group	1.00	91.00	0.00	7.35	0.02	0.01	**
ROI	1.92	174.60	0.00	92.03	0.44	0.00	***
Group x ROI	1.92	174.60	0.00	7.83	0.06	0.00	***

Note. MSE: mean squared error; ges: generalized eta square;
* : $p < .05$; ** : $p < .01$; *** : $p < .001$

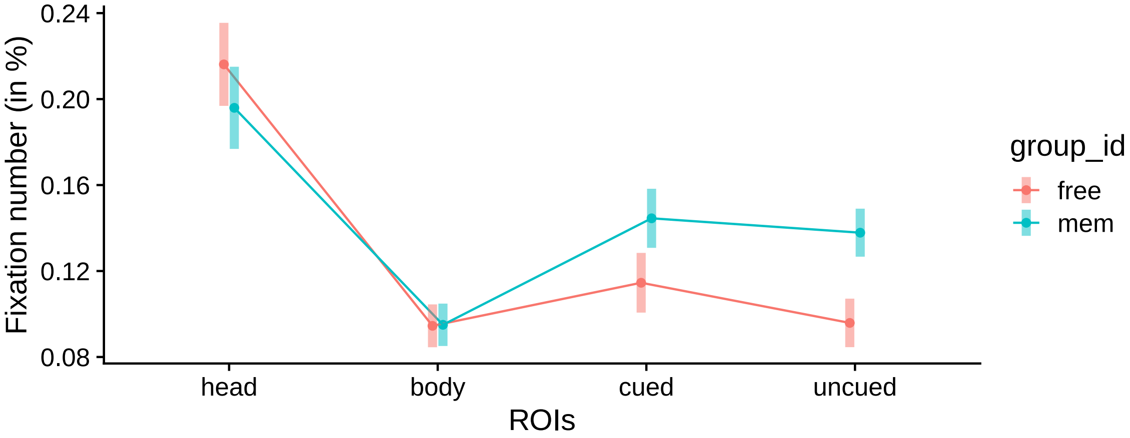

Figure A.3: Fixation numbers as a function of the region of interest and group. Error bars represent 95% confidence interval.

Appendix B

Supplementary material

Reality in a sphere: A direct comparison of social attention in the laboratory and the real world

B.1 Assumption check & robust LMM

A central assumption for the preregistered linear mixed models was not met by our data: the residuals were not normally distributed (see Figure S1, $W = 0.94$, $p < .001$).

Since this can have severe consequences for the statistical properties of the linear mixed model (Kenny and Judd 1986), we additionally calculated a robust linear mixed model following Koller (2016). These methods aim to provide robust estimates in conditions where assumptions for linear mixed models are violated. For the current study, the estimated coefficients of the robust model (see Table S1) closely matched the coefficients of the original linear mixed model (see Table 1 in the main article). As a result, the estimates, standard errors, and t-values are very similar and support our interpretation of the original findings in the main article.

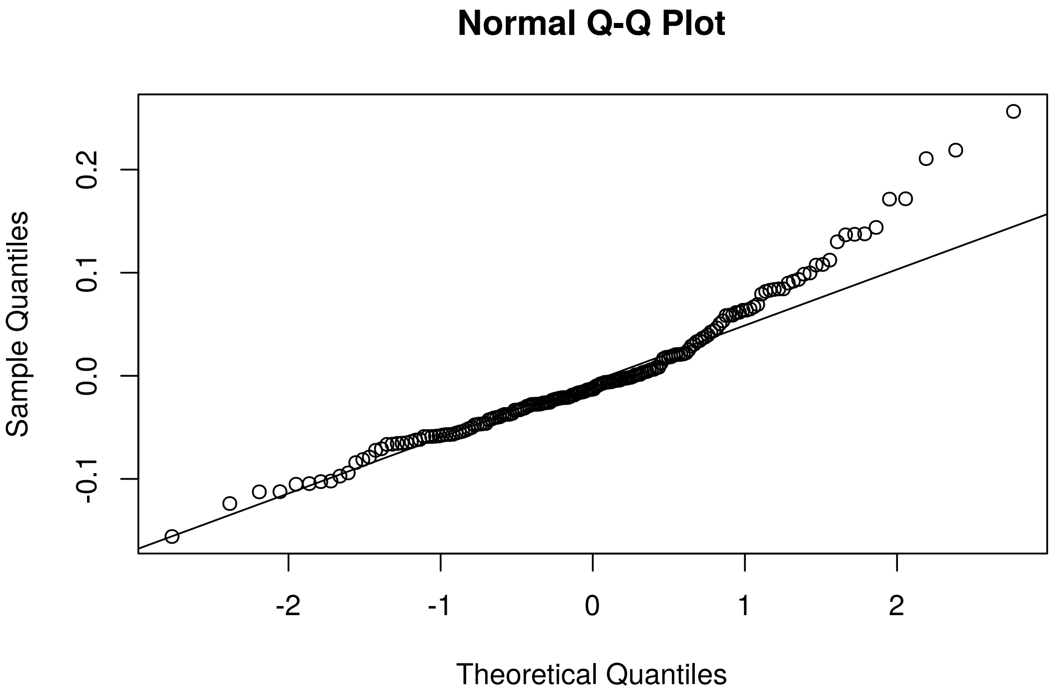

Figure B.1: QQ plot for the residuals of the originally preregistered model.

Table B.1: Estimated coefficients for a robust linear mixed model (Koller, 2016), specified exactly like the preregistered model 1 (see main article) with environment and ROI as fixed and participant ID as random effects for the prediction of gaze proportions.

	Estimate	*SE*	*t*
Intercept	0.19	0.00	38.03
Environment (RE)	-0.09	0.00	-18.63
ROI (object)	0.02	0.00	3.64
Environment (RE) × ROI (object)	0.02	0.00	3.09

Note. The robust linear mixed model is based on sum-to-zero contrasts. RE: real environment, ROI: region of interest.

Table B.2: Estimated standard deviations, and correlations between the random-effects.

Groups	Name	SD	r
Subject (sub)	Intercept	0.03	
Location (loc)	Intercept	0.02	
	Present persons (P)	0.01	-1
Residual		0.13	

B.2 Convergence issues in Model 4

Following the outlined model building path in the manuscript the complete model was initially specified[1] as:

$$fix \sim env * roi * P + (1|sub) + (1 + P|loc)$$

The correlation between location and present persons was estimated to be exactly -1 (see Table S2). Thus, the estimated correlation can be considered compromised, resulting in meaningless model output. In a subsequent step, we, therefore, suppressed the correlation between the random slope for present persons and locations, and specified the following pruned model (note the `{||}` indicating uncorrelated effects in the random term for `loc`):

$$fix \sim env * roi * P + (1|sub) + (1 + P||loc)$$

The model converged successfully and is reported in the manuscript (see Table 5).

---

[1] Using the lme4 notation (Bates et al. 2015).

Appendix C

Publications list

- **Großekathöfer, Jonas D.**, Suchotzki, K., & Gamer, M. (2020). Gaze cueing in naturalistic scenes under top-down modulation –effects on gaze behaviour and memory performance. Visual Cognition, 28(2), 135–147. `http://dx.doi.org/10.1080/13506285.2020.1742826`
- **Großekathöfer, Jonas D.**, Seis, C., & Gamer, M. (in press). Reality in a sphere: a direct comparison of social attention in the laboratory and the real world. Behavior Research Methods & Instrumentation, (), . `http://dx.doi.org/10.3758/s13428-021-01724-0`

Appendix D

Curriculum vitæ

OMITTED

OMITTED

Appendix E

Affidavit / Eidesstattliche Erklärung

I hereby confirm that my thesis entitled *Virtually Valid? On the Importance of Ecological Validity and Virtual Reality for Social Attention Research* 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

Hiermit erkläre ich an Eides statt, die Dissertation *Praktischerweise Valide? Über die Bedeutung von ökologischer Validität und virtueller Realität in der sozialen Aufmerksamkeitsforschung* eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters angefertigt zu haben 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 in einem anderen Prüfungsverfahren vorgelegt wurde.

Ort, Datum Unterschrift