LOOK AT ME AND I WILL FEEL YOU: EYE CONTACT AND SOCIAL UNDERSTANDING

SCHAU MICH AN UND ICH SEHE DICH: BLICKKONTAKT UND SOZIALVERSTEHEN

Doctoral thesis for a doctoral degree

at the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Neuroscience

Submitted by

Christina Breil

from

Neuss

Würzburg, 2021

Submitted on:

……………………………………………………………..……..

Office stamp

Members of the Thesis Committee

Chairperson:

……………………………………………………………......

Primary Supervisor: Anne Böckler-Raettig

Supervisor (Second): Lynn Huestegge

Supervisor (Third): Marcel Romanos

Supervisor (Fourth): Matthias Gamer

Date of Public Defence: …………………………………………….…………

Date of Receipt of Certificates:

……………………………………………….

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to all my supervisors for their continuous and reliable support: Anne, Lynn, Marcel and Matthias. Your expertise and feedback have been invaluable to me for the preparation and formulation of this dissertation. Most notably, I wish to thank Anne for the great encouragement, the endless patience and the trustful guidance in all situations.

I also want to thank all of my colleagues: Roxana, Rainer, Tim, Micha, Lisa, Barbara, Eva, Mareike, Julian and Jens from Würzburg as well as Basil and Leticia from Hanover. It has been pure joy and motivation to be part of such a great team!

Furthermore, I would like to say a special thank you to all of our research assistants who supported me so greatly and whom I could always rely on: Sophia, Kristin, Laura, Anne, Caro, Leila, Beke, Sara, Kristina and Nimra.

Last, but certainly not least, I wish to thank my dear family, for your limitless love and support and for giving me so much faith and courage. Without you, I might never have started this dissertation, let alone finished it.

This work was supported by the Emmy Noether Programme of the German Research Foundation (Deutsche Forschungsgesellschaft (DFG); grant number BO 4962/1-1.).

Für Opa Heinz.

ABSTRACT

One of the features that defines humans as extraordinarily social beings is their striking susceptibility to the gaze of others. The research reported in this dissertation was undertaken to advance our understanding of the role of gaze cues in low-level attentional and higher-order cognitive processes. In particular, effects of gaze were examined with regard to three aspects of human cognition: (1) social attention, (2) social interaction and (3) social understanding. Chapter 1 consists of three manuscripts that investigate the boundary conditions of attention capture by direct gaze and how gaze direction is integrated with facial context information. Manuscript 1 and 2 suggest two necessary requirements for attention capture by direct gaze: a meaningful holistic facial context and sharp foveal vision, respectively. Manuscript 3 shows approach/avoidance-congruency effects between gaze direction and emotion expression on attention. Chapter 2 of this dissertation explores the role of gaze in more naturalistic social scenarios. Manuscript 4 demonstrates that gaze behavior during a conversation shapes our perception of another person. Manuscript 5 builds on these findings by showing that these perceptions define our willingness to act in a prosocial way towards our interaction partner. Finally, chapter 3 adopts a broader perspective on social cognition research with a special focus on methodological aspects. Manuscript 6 is a review highlighting the significance of methodological aspects in social cognition research and stressing the importance of sophisticated decisions on task and stimulus materials. Manuscript 7 introduces a new instrument for the assessment of social understanding in adolescents. Initial application in a young sample group indicates that an understanding of another person's mental states is a capacity that is still developing throughout adolescence. Both manuscripts of this final chapter include eye tracking data that suggest a relationship between gaze behavior and social understanding, a finding that further emphasizes the complex and multifaceted nature of social cognition. I conclude from the findings of this dissertation that

research can benefit from adopting a broad view in terms of methodological as well as temporal aspects in order to capture human social cognition in its entirety.

ZUSAMMENFASSUNG

Die herausragend soziale Natur des Menschen zeigt sich insbesondere in der sensiblen Reaktion auf die Blicke anderer. Ziel der in dieser Dissertation berichteten Forschung ist ein umfassendes Verständnis der Rolle von Blickreizen auf kognitive Prozesse niederer und höherer Verarbeitungsstufen. Im Einzelnen wurden Blickeffekte im Hinblick auf drei Aspekte menschlicher Kognition untersucht: (1) Soziale Aufmerksamkeit, (2) soziale Interaktion und (3) Sozialverstehen. In Kapitel 1 werden drei Studien beschrieben, die sich mit den Grenzbedingungen von Aufmerksamkeitsanziehung durch direkten Blickkontakt beschäftigen und die untersuchen, wie Effekte der Blickrichtung mit anderen Reizen interagieren. Manuskript 1 und 2 deuten auf zwei notwendige Voraussetzungen für den direkten Blickeffekt hin: ein holistisch bedeutsamer Gesichtskontext sowie scharfe, foveale Wahrnehmung. Manuskript 3 findet aufmerksamkeitsbezogene Annäherungs-/Vermeidungskongruenzeffekte zwischen Blickrichtung und emotionalem Gesichtsausdruck. Kapitel 2 dieser Dissertation untersucht die Rolle von Blicken in naturalistischeren sozialen Situationen. Manuskript 4 demonstriert, dass Blickverhalten in Gesprächen unsere Wahrnehmung anderer Personen beeinflusst. Manuskript 5 erweitert diesen Befund, indem es verdeutlicht, dass diese Eindrücke unsere Bereitschaft zu prosozialem Verhalten gegenüber unseren Interaktionspartner*innen bestimmen. Schließlich wird im 3. Kapitel eine breitere Sicht auf sozialkognitive Forschung eingenommen. Ein besonderer Fokus liegt dabei auf methodischen Aspekten. Manuskript 6 ist ein Review, das die Tragweite methodischer Aspekte in sozialkognitiven Untersuchungen herausarbeitet und auf die Bedeutung gut informierter und durchdachter Entscheidungen bezüglich der verwendeten

Versuchsmaterialien hinweist. In Manuskript 7 wird ein neues Instrument zur Erfassung sozialen Verstehens in jugendlichen Stichproben beschrieben. Eine erste Anwendung dieser neuen Methode deutet darauf hin, dass sich das Verständnis der mentalen Zustände anderer Menschen im Jugendalter noch in der Entwicklung befindet. Beide Manuskripte dieses letzten Kapitels enthalten Eye-Trackingdaten, die auf einen Zusammenhang zwischen Blickbewegungen und Sozialverstehen hindeuten. Dieser Befund verdeutlicht, dass soziale Kognition ein komplexes und breitgefächertes Konstrukt ist. Ich schließe aus den Ergebnissen dieser Dissertation, dass die Wissenschaft sowohl im Hinblick auf methodische als auch auf zeitliche Aspekte von einer umfassenden Sichtweise auf soziale Kognition profitieren könnte, da nur diese es ermöglicht, das Konstrukt in Gänze zu erfassen.

INTRODUCTION

In Richard Matheson's novel "I Am Legend", Robert Neville is the last-known survivor of a pandemic that killed most of the human population and turned the rest into vampires (Matheson, 2007). Neville is utterly alone and seeks escape in his memories, loud music and alcohol. Witnessing the protagonist day to day without any social contact, empathizing with his depression and unbearable loneliness, the reader gets the idea that this fate is even worse than becoming a vampire. The reason for this very natural feeling is that humans are inherently social beings. We grow up in groups, we learn from and cooperate with each other and the individual characteristics and traits that define each of us as a unique person are influenced by the people we surround ourselves with (Boyd & Richerson, 2009; Caspi & Shiner, 2006; Dockett & Perry, 1996; Tomasello & Vaish, 2013). One of the basic human desires is the need for love, affection and belonging, and the unfulfillment of these cravings increases the risks of morbidity and mortality (Bzdok & Dunbar, 2020; Cacioppo & Cacioppo, 2014; Maslow, 1943). Accordingly, every person has a hunger for affectionate relations and a secure place in the group. To achieve and maintain these social desires, each individual has to pay attention to other people in order to understand their intentions and emotions and to respond to them in an appropriate way.

In this sense, it is not surprising that the investigation of social attention and cognition has become one of the key research topics in psychology and behavioral and social neurosciences in the last decades. This dissertation will focus on three key branches of this line of research: on the human susceptibility to gaze cues (chapter 1), the role of gaze behavior in human interaction (chapter 2) and on the investigation of social understanding, that is, the capability or process to share other people's emotional and cognitive states, and the relation of these processes to gaze behavior (chapter 3).

GOAL AND ORGANIZATION OF THIS DISSERTATION

This dissertation is intended to advance the current literature on the mechanisms that underlie successful human social interaction. Three themed sections deal with three different aspects of social cognition. In each chapter, I will cite and summarize different research manuscripts and discuss their results together in an interim discussion. At the end of this dissertation, I will integrate all findings of my research in a general discussion. The complete manuscripts can be found in the appendix at the end of this dissertation.

In chapter 1, the basic effects of direct gaze on human attention are illuminated. I summarize three studies that deal with the boundary conditions of attention capture by direct gaze and the integration of this effect with the impact of other social and nonsocial cues. Manuscript 1 (published) consists of six experiments that systematically assess whether and to what degree the (sudden) direct gaze effect relies on social context information. In Manuscript 2 (under review), presentation eccentricity is parametrically manipulated in order to investigate how it shapes the effects of direct gaze and motion onset on attention. The last study of this section (Manuscript 3, submitted) tests the integration of gaze direction with facial information by concurrently presenting gaze cues and emotion expressions.

Chapter 2 includes two studies that explore the influence of gaze behavior in more naturalistic social interactions. The first study (Manuscript 4, published) investigates how gaze behavior during neutral or negative online video conversations shapes a person's impression of a conversation partner. The second study (Manuscript 5, in preparation) builds on these findings by testing whether these impressions transfer onto prosocial behavior and precisely how the effects are interrelated.

Chapter 3 focuses on methodological aspects of research on social understanding. Manuscript 6 (published) is a review, emphasizing the importance of targeted research on the relationship between task specifications and outcomes. This article includes data from a pilot study that probed the relationship between gaze behavior and social understanding. Finally, Manuscript 7 (published) introduces a new paradigm for the assessment of social cognition in adolescent samples. Non-published data of gaze behavior in experiment 2 of Manuscript 7 is added as supplementary material exclusively to this dissertation.

The general discussion at the end of this dissertation serves to integrate all findings and to deduce their implications for future research on social attention and cognition.

MATERIALS AND METHODS

The precise procedures employed, materials applied and analyses performed are described in detail in the methods section of each manuscript. For unpublished or supplementary data, a comprehensive description of materials and methods is included in the respective chapter of this dissertation.

CHAPTER 1

Studies in this chapter employ a basic target detection paradigm to assess attention capture by direct gaze and sudden onset motion. Each manuscript contains additional manipulations, for example the degree to which stimuli are socially relevant and naturalistic (Manuscript 1), how close to the central fixation cross stimuli are presented (Manuscript 2) or the emotion expression of the facial stimuli (Manuscript 3). Response times (RTs) are of particular interest, but error rates were also analyzed. In addition to that, Manuscript 3 includes eye tracking data. The analyses of all three studies are mainly focused on repeated measures or mixed effects analyses of variance (ANOVAs).

CHAPTER 2

The two studies in this chapter present short video-based interactions in which the gaze behavior of a listener and the emotional valence of a speaker's narrations are manipulated within-subjects. In Manuscript 4, the alleged relationship between the interaction partners is manipulated between three groups of participants. Dependent variables in Manuscript 4 are ratings of perceived empathy, perspective taking and trustworthiness as well as the emotional closeness between the conversation partners. Manuscript 5 additionally investigates trust and prosocial behavior by means of economic games. Analyses entail repeated measures and mixed effects ANOVA as well as mediation analyses.

CHAPTER 3

This chapter starts with a narrative review that reflects on the intricate relationship between task choice characteristics and findings in Theory of Mind (ToM) research (Manuscript 6). This article includes data of a pilot study that employed the EmpaToM (Kanske et al., 2015b) with eye tracking. In Manuscript 7, the EmpaToM is adjusted for application in adolescents to assess empathic responding and ToM performance. Dependent variables for the EmpaToM and its adaptation are ratings and performance data (RTs and error rates). Besides these behavioral measures, the analyzed variables include physiological data (pupillometry and electrodermal activity). Nonpublished eye tracking data that was collected in the context of the final experiment of this dissertation (experiment 2 of Manuscript 7) is added to chapter 3. The analysis of gaze behavior is focused on the number of fixations to and the time spent on the eye regions of the narrators.

OPEN SCIENCE STATEMENT

Within each manuscript and each experiment of this dissertation, I report all manipulations, all data and participant exclusions as well as all measures that were employed (Simmons et al., 2012). The sample size in each study was determined with G*Power (Faul et al., 2007b) and a majority of the data is publicly available within the Open Science Framework (osf.io). Experiment 2 of Manuscript 5 was preregistered.

CHAPTER 1: BASIC EFFECTS OF EYE GAZE

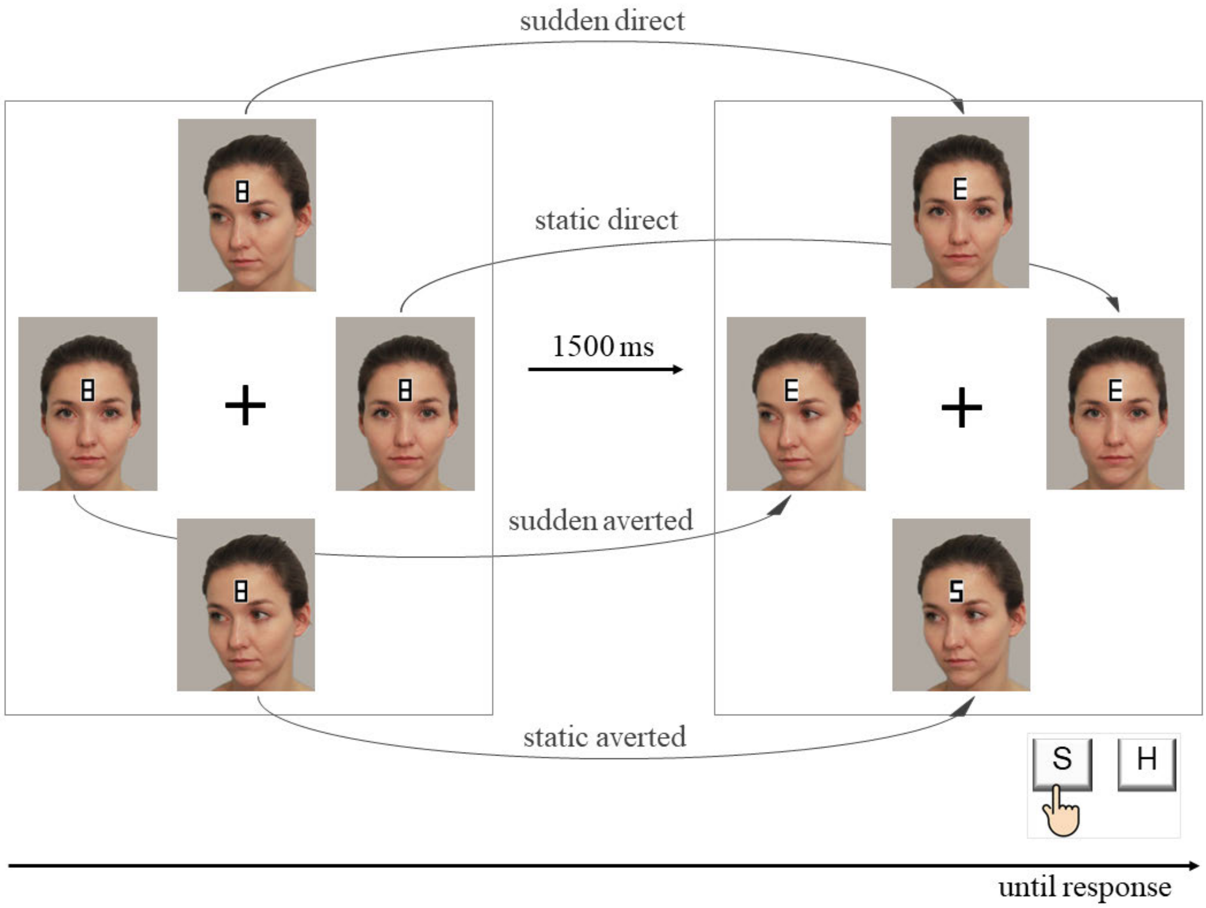
From childhood on, one of the first and most frequently fixated facial regions are the eyes (Arizpe et al., 2017). For humans, in contrast to many other species, eye contact has manifold and largely positive effects that provide the basis for successful communication and social interaction (Emery, 2000). Perceived eye contact modulates subsequent attention and cognition by activating large parts of the social brain network, a compilation of cortical and subcortical regions specialized in the processing of social information such as human action or the face (Senju & Johnson, 2009b). The effects of direct gaze are far-reaching and become apparent in an abundance of tasks that are typically employed to study human attention, perception and cognition (see Schilbach, 2015 for an overview).

A classic example is the Posner paradigm, a popular tool to investigate attentional mechanisms that are involved in human awareness about the environment (Posner et al., 1980). In the basic spatial cueing paradigm, participants are instructed to fixate on a centrally presented cross and respond by key press to a target stimulus that is presented in the left or right periphery. Prior to target onset, the presentation location is cued by the sudden onset of either a peripheral cue or a centrally presented directional cue such as an arrow. Responses are typically faster and/or more accurate if the target appears at the location that has been correctly predicted by the cue (as compared to distractor trials in which incorrect cues are presented), indicating that attention has been shifted to the cued location. This type of reflexive orienting occurs even when the cue is unpredictive (i.e. it predicts the correct location only in 50% of the trials) or when participants were explicitly instructed to ignore it (Remington et al., 1992; see Frischen et al., 2007 for a review). In modified versions, the Posner paradigm has proven invaluable to study attentional mechanisms in response to gaze direction. Even nonpredictive gaze cues lead to reflexive shifts of attention to peripheral locations (see Frischen et al., 2007 for a review). Not only do human participants follow the gaze of other's, the objects that are attended are also processed with extraordinary rapidity and precision (Driver et al., 1999). Even though, on a behavioral level, the same effects can be observed with non-social directional cues (e.g. arrows), the neural mechanisms that underlie attention shifts in response to biologically relevant stimuli, such as gaze cues, are fundamentally different (J. K. Hietanen, Leppänen, Nummenmaa, et al., 2008; Ristic et al., 2002).

A special case of gaze cues are eyes that are targeted directly at us. Direct gaze is particularly powerful at attracting and holding human attention: when direct gaze cues are employed in the gaze cueing paradigm, responses to peripheral targets are delayed unless a temporal gap is inserted between cue and target presentation (Senju & Hasegawa, 2005). Another paradigm that allows to study direct gaze effects is visual search, a standard task to investigate attentional mechanisms and the hierarchy of information-processing operations. In this task, participants are required to judge whether or not a target (such as a human face with direct gaze) is present among a set of distractors (the same faces with averted gaze). Typically, this is achieved much faster than detecting a face with averted gaze among direct gaze distractors (Conty et al., 2006; Senju et al., 2008). Perhaps you can relate to this "stare in the crowd effect" from your own experience, for example when looking around in a crowded bar and, despite the mass of people chatting and dancing, immediately catching the eye of someone else.

Another phenomenon most people can relate to is the strong attraction of sudden onset direct gaze. When another person raises their head to look at us, we are drawn to make eye contact. What is so special about this situation is that two powerful cues that capture attention coincide in time and space: direct gaze and the sudden onset of motion. Böckler et al. (2015; 2014) designed a neat paradigm to disentangle the effects of these two cues and found that they exert their influences independently and in parallel. In this task, four faces are presented around a central fixation cross with small number-8 figures on their foreheads (*Figure 1*).

Figure 1. *Paradigm to assess attention capture by direct gaze and sudden onset motion.*



Note. Adapted from Böckler et al. (2014). At the beginning of each trial, number-8-figures were presented at the forehead of all four faces. After 1500 ms, they were replaced by one target and three distractor letters while, at the same time, one direct stimulus changed to averted and one averted stimulus changed to direct. The other two stimuli remained unchanged, resulting in the four experimental conditions sudden direct, static direct, sudden averted, static averted. Participants were required to react to the target letter by pressing the corresponding button on the keyboard as fast and accurate as possible.

Two of these faces directly gaze at the participant and two of them look into the periphery. After a fixed period of 1500 ms, one direct gaze stimulus changes to averted while one averted gaze stimulus changes to direct, inducing apparent motion. At the same time, the four number-8 cues are replaced by three distractor letters ("E" or "U") and one target letter ("S" or "H") to which the participant is required to respond to as quickly and accurately as possible by pressing the corresponding button on a standard keyboard. Typically, participants respond faster when the target is presented on a face that depicts direct gaze (direct gaze effect) or on faces that had apparently moved prior to target presentation (motion effect). Böckler (2015; 2014) found that these two effects work independent and in parallel.

Chapter 1 of this dissertation will summarize three studies that employ this paradigm in order to systematically assess the boundary conditions of attention capture by ostensive cues (Manuscript 1), to test how attention capture by direct gaze is shaped by the characteristics of the visual display (Manuscript 2) and how it relates to effects of facial expression (Manuscript 3).

MANUSCRIPT 1

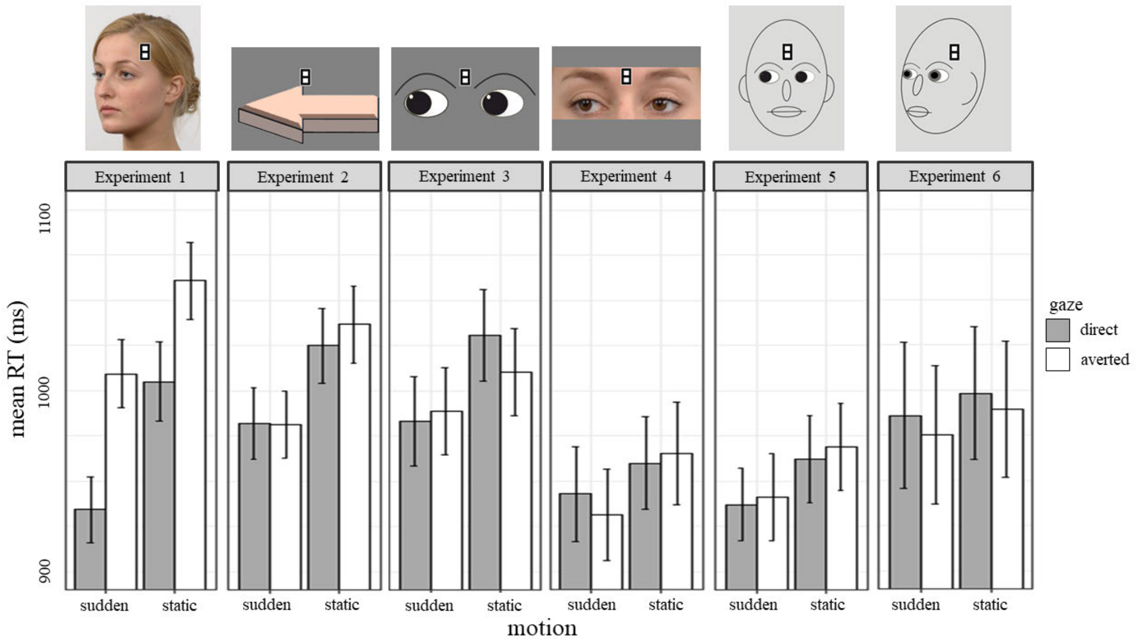
Breil, C., Huestegge, L. & Böckler, A. (2021). From eye to arrow: Attention capture by direct gaze requires more than just the eyes. *Attention, Perception & Psychophysics.* DOI: 10.3758/s13414-021-02382-2

As highlighted in the introduction of this chapter, (sudden) direct gaze is a powerful cue to capture human attention. However, eye gaze does usually not occur in isolation, but is embedded in a social context, such as the face. In Manuscript 1, we systematically assessed the necessity of naturalistic and holistic social information for the (sudden) direct gaze effect.

To this end, we employed the paradigm of Böckler et al. (2014) to six consecutive experiments with different stimuli that varied in their degree of holistic and realistic context information. In particular, we replicated previous studies with holistic photographic cues (experiment 1, photographic faces; Böckler et al., 2014, 2015; Boyer & Wang, 2018) and subsequently tested the effects of gaze and motion with ostensive symbolic cues (arrows: experiment 2), non-holistic cues (isolated eyes: photographic in experiment 3 and schematic in experiment 4) and holistic non-photographic cues (schematic faces without and with apparent head-turn movement: experiments 5 and 6, respectively).

We found a largely stable motion effect across all experiments except for one (experiment 6; *Figure 2*). Most notably, we replicated the direct gaze effect for photographs of human faces and demonstrated its break-down for stimuli with decreasing amount of social information, namely schematic faces with and without head-turn, photographic eyes, schematic eyes and arrows.

Figure 2. *Mean RTs for all conditions for experiments 1-6 of Manuscript 1.*



Note. Error bars represent standard errors. Pictures above the charts depict stimuli of the averted gaze-condition of each experiment.

Hence, our findings show that neither symbolic ostensive cues like arrows nor isolated eye cues or schematic facial configurations are powerful enough to capture human attention. In this way, our results highlight the importance of a naturalistic and socially meaningful face context for attention capture by direct gaze.

MANUSCRIPT 2

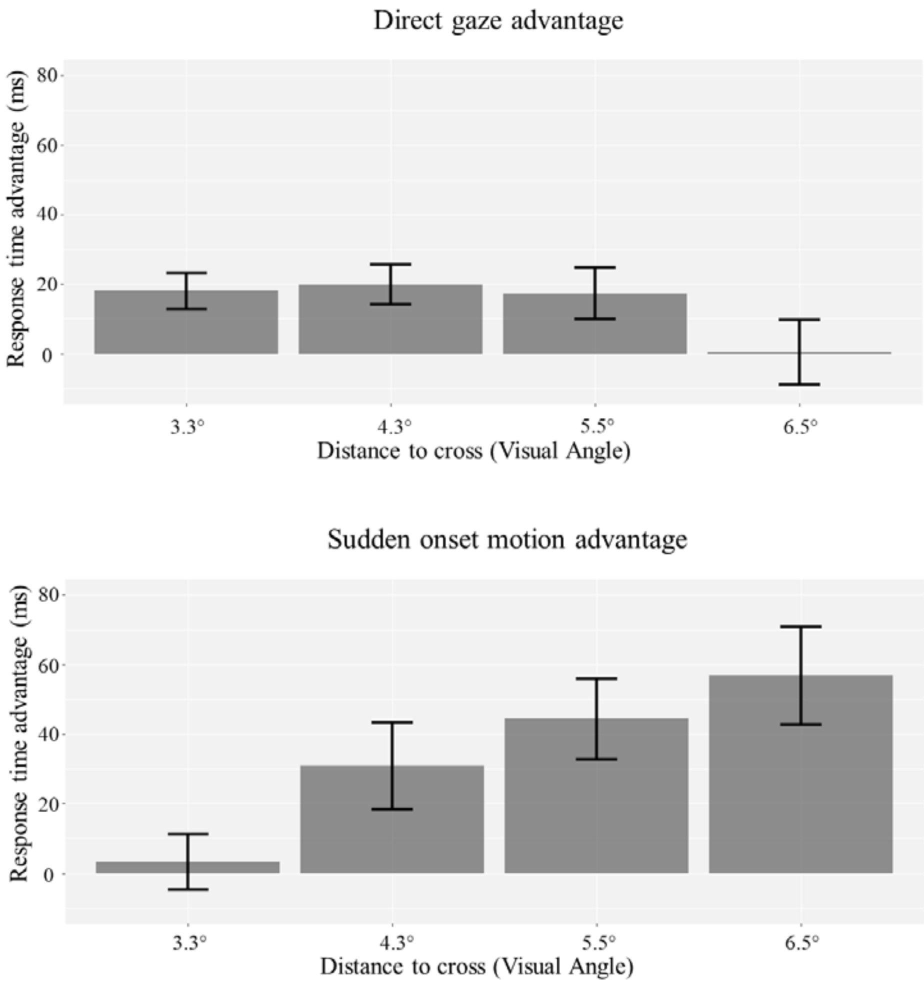
Breil, C., Huestegge, L., Pittig, R. & Böckler, A. (under review). How eccentricity modulates attention capture by direct gaze and facial motion. *Journal of Vision.*

A recurring finding of studies from our group is the dissociation of gaze and motion effects on human attention. In the previous section, I summarized a study showing that a change in social features of the stimuli can affect the direct gaze effect while leaving the motion effect largely intact (Breil, Huestegge, et al., 2021). As another example, Böckler et al. (2014) showed that effects of direct-gaze cues persisted when a stimulus onset interval of 900 ms is introduced while motion cues are susceptible to the inhibition of return effect and reduce detection performance. Taken together, these studies indicate that the attentional channels underlying the processing of gaze and motion cues work independently and in parallel.

In all of these studies, stimuli were presented at a fixed distance just at the edges of foveal vision in order to enable reliable conclusions about the origin of effects. However, the notion that gaze and motion cues work on distinct processing channels raises the question whether they are affected in different ways by the properties of the human visual system. For example, it is a well-accepted finding that the visual periphery is particularly sensitive to motion cues while the foveal region is highly contrast-sensitive and most accurate in perception (Burnat, 2015; Kitterle, 1986; Yu et al., 2010). If you have read the previous sections of this chapter, you will remember that face (and particularly gaze) cues produce extraordinary effects on the human attentional system. With this in mind, you may not be surprised that faces are once more an exception to the rule that typically applies to peripheral vision: Even at relatively high eccentricities, faces, in contrast to many other objects, are processed quite accurately (Boucart et al., 2016; Hershler et al., 2010). In Manuscript 2 we parametrically manipulated presentation eccentricities in order to investigate how they shape the effects of direct gaze and motion onset on attention. To this end, we employed the attention capture paradigm by Böckler et al. (2015; 2014) at different horizontal foveal eccentricities in a between-subjects design, namely at 3.3°, 4.3°, 5.5° and 6.5° of visual angle (VA). We replicated the direct gaze and motion effect at the original distance of 4.3° VA and found that both effects vary differently across different eccentricities. In particular, the direct gaze effect was reliable only when the stimuli were presented close to the center while the motion effect profited from more peripheral presentation (*Figure 3*).

These findings substantiate the notion of independent processing channels for direct gaze and motion cues and suggest a functional specialization of the central and peripheral visual field. Earlier studies linked central vision to brain areas that are related to endogenous attention for fine detail perception while responses in areas involved in action and object allocation were enhanced by attention to peripheral objects (Bressler et al., 2013; Corbetta & Shulman, 2002; Roberts et al., 2007). The results of Manuscript 2 are in line with these observations by revealing that attentional effects of gaze and motion cues are differentially shaped by eccentricity.

Figure 3. *Direct gaze and sudden onset motion advantage effects across distances to the center in Manuscript 2*



Note. The direct gaze advantage is calculated as the mean RT for averted gaze minus the mean RT for direct gaze stimuli. The sudden onset motion advantage is calculated as the mean RT for static minus the mean RT for sudden stimuli. Hence, positive values indicate a direct gaze or a sudden onset motion advantage, respectively. Experiment 1 = 3.3°VA, experiment 2 = 4.3°VA, experiment 3 = 5.5°VA, experiment 4 = 6.5°VA. Error bars represent standard errors.

MANUSCRIPT 3

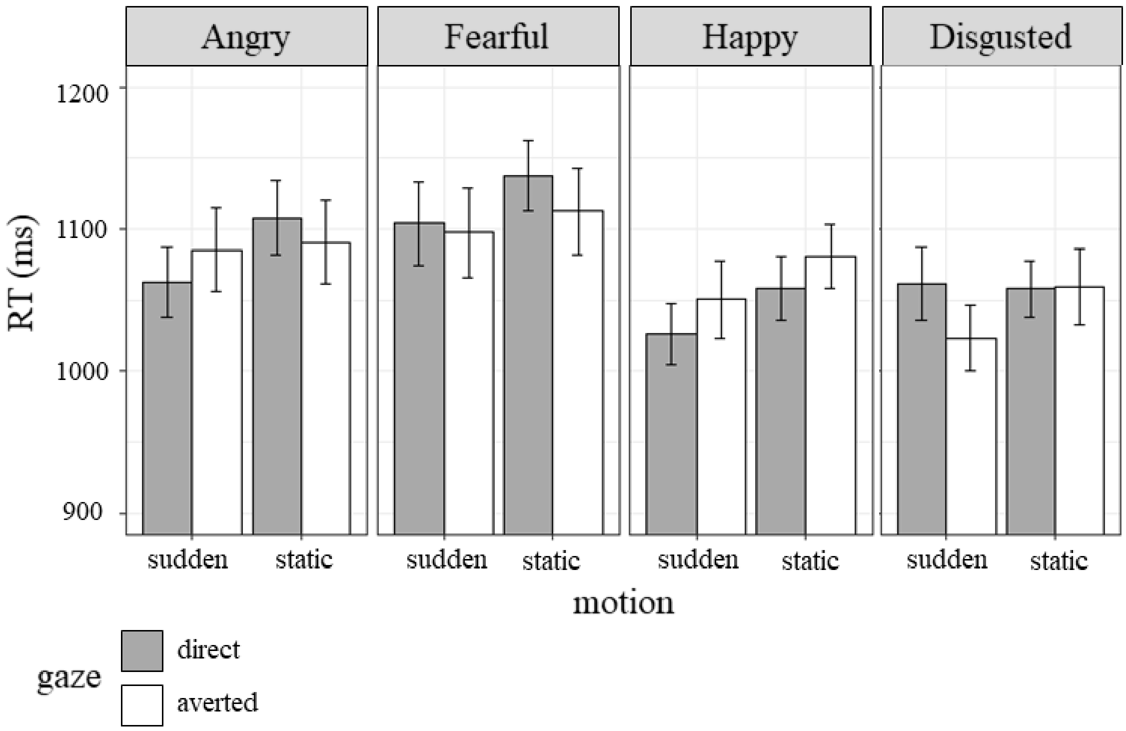
Breil, C., Raettig, T., Pittig, R., van der Wel, R., Welsh, T. & Böckler, A. (submitted). Don't look at me like that: integration of gaze direction and facial expression. *Journal of Experimental Psychology: Human Perception and Performance.*

As highlighted in the first section of this chapter (Manuscript 1), the effects of direct gaze on attention do not emerge in isolation but are profoundly shaped by facial context information. The idea that facial information, such as emotion expression, is integrated with gaze direction is not new and has received considerable attention in scientific debates in the past decades. A recurring notion in this line of research is that the processing of emotion expressions is facilitated when it is congruent to the expresser's approach-avoidance behavioral tendency (Adams & Kleck, 2003, 2005). As an example, fearful faces are processed more efficiently when they look into the periphery while angry faces are prioritized when they establish direct gaze (Bindemann et al., 2008; N'Diaye et al., 2009; Sander et al., 2007). However, the integration of gaze and face information is subject to a complex and emotion-specific interaction of top-down influences and stimulus-driven effects across the visual processing stream. The precise origin of these effects is still debated (e.g. Haxby & Gobbini, 2011; Itier & Neath-Tavares, 2017; Neath & Itier, 2015; Neath-Tavares & Itier, 2016; Schindler & Bublatzky, 2020; Schupp et al., 2004).

Manuscript 3 was designed to investigate the temporal dynamics of effects of gaze direction and of facial expression as well as their interplay on attention. To this end, we employed an adaptation of Böckler et al.'s (2014) paradigm in two consecutive experiments. The critical manipulation was the sudden introduction of an emotion expression at the exact same time that the target was presented. In particular, one approach-oriented emotion (experiment 1a: angry, experiment 1b: happy) and one avoidance-oriented emotion (experiment 1a: fearful, experiment 1b: disgusted) were presented in experiment 1, each in randomly selected 50% of trials. This setup allowed us to assess the time course of the integration of emotion expression and gaze direction because both cues were simultaneously presented in this study.

We found a break-down of the gaze effect for expressions of anger and fear (experiment 1a) but a modulation of the direct gaze effect by happy and disgusted expressions (experiment 1b; *Figure 4*). These results speak to the idea of a flexible and emotion-specific interaction of gaze and face information.

Figure 4. *Mean RTs by condition for each emotion of experiment 1 of Manuscript 3*



Note. Experiment 1a: angry/fearful; experiment 1b: happy/disgusted.

To further investigate the origin of these effects and the time course of visual attention, we replicated experiment 1b (happy/disgusted), in which a modulation of the gaze effect by emotion expression was found, while tracking participant's eyes (experiment 2).

On the behavioral level, we (partly) replicated experiment 1b with a modulation of the gaze effect by facial expression, suggesting an integration before a response was made. These findings were largely mirrored in gaze behavior with an attentional benefit for happy faces when they depict direct gaze but for disgusted faces when they depict averted gaze. Most importantly, our findings suggest that an integration of gaze direction and facial expression took place at early processing stages, starting between 275 and 375 ms with a duration of 225 ms. The time windows of integration were overlapping but distinct for the two emotions.

Taken together, our findings suggest that gaze direction and emotion expression interact in a flexible fashion and that an early integration of both cues is possible. In particular, emotion expressions that convey an urgent need for action, namely anger and fear (Ortony & Turner, 1990; Schupp et al., 2004; Vuilleumier, 2002a), diminished the direct gaze advantage in our study. In contrast, expressions of happiness and disgust differentially modulated the gaze information, with a direct gaze advantage for happy (approach oriented) and an averted gaze advantage for disgusted (avoidance oriented) faces. In sum, our results speak to the idea of an approach/avoidance-congruency effect and an emotion-specific *early integration* of gaze direction and emotion expression (Adams & Kleck, 2003, 2005). In contrast, the motion effect was relatively stable across all conditions and experiments.

INTERIM SUMMARY

In this chapter, I summarized three studies that assess the boundary conditions of and influences on the effects of gaze and sudden onset motion on human attention. Findings from Manuscript 1 indicate that the direct gaze effect critically relies on a naturalistic and socially meaningful facial context while the motion effect remains stable across a variety of ostensive cues. This finding suggests that face perception occurs in a holistic fashion and that direct gaze and motion cues are based on separate processing channels. The idea of an independence of these mechanisms is further substantiated by findings of Manuscript 2: the attentional effects of gaze and motion cues are differentially shaped by eccentricity, exemplifying a functional specialization of the central and peripheral field. According to the findings of this study, the direct gaze effect is particularly strong near the foveal center while the motion effect profits from increasing peripheral presentation. The last study of this chapter, Manuscript 3, further illustrates the impact of social context on the direct gaze effect by showing an emotion-specific integration process of gaze direction and emotion expression. While expressions of anger and fear diminished the gaze effect, expressions of happiness and disgust modulated it. Visual attention patterns in experiment 2 suggest that the integration of gaze and emotion cues took place at early processing stages, starting at 300 ms after stimulus onset. These results speak to the idea of an early and emotion-specific approach/avoidance-congruency effect when processing gaze and face information.

Taken together, the findings of this chapter give insight into the mechanisms and time course of integration of gaze direction with social and non-social cues. Investigating the basic attentional mechanisms that are driven by gaze cues will allow a better grasp on social attention, face processing and person perception in order to ultimately gain a comprehensive understanding of interpersonal communication and decision making.

There are two overall conclusions from the studies of this chapter: First, direct gaze and motion cues rely on separate processing channels that work independently and in parallel. This notion is supported by the finding that the motion effect, in contrast to the direct gaze effect, remains relatively untouched by (social) context information and feature-related stimulus aspects. The implementation of different stimuli across the three studies speaks to the reliability and generalizability of this finding. Furthermore, attentional effects of both cues are shaped by the properties of our visual system in their very own ways. The second important conclusion is that gaze direction is not perceived in isolation but is embedded in the context of other facial cues and social information. This finding suggests an early interaction of bottom-up and top-down effects that are part of a sophisticated system specialized in the processing of gaze cues.

LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

The studies that I summarized in this chapter give important insights into the human attentional and perceptual system. Direct gaze produces a very prominent and stable effect on attention. This effect works independently from non-social cues such as the sudden onset of motion. However, it can be integrated with social context information such as facial emotion. Taken together, these findings suggest that the underlying mechanism is a fundamental and essential component of human cognition. At the same time, this notion highlights the importance of context variables when studying gaze and face processing in humans. Hence, in order to gain a comprehensive and ecologically valid understanding of interpersonal communication and decision making, it is important to depart from artificial laboratory conditions and include more naturalistic stimuli and tasks in social cognition research. In the next chapter, I will introduce two studies that were designed to meet this demand by employing dynamic video and sound stimuli in a naturalistic, yet controlled experimental setting.

CHAPTER 2: EFFECTS OF GAZE IN MORE COMPELX SOCIAL SCENARIOS

The effects of gaze cues that I described in the first chapter are unarguably powerful. However, their impact goes far beyond a modulation of human attention: gaze behavior serves important functions in social interactions that reach from information delivery to the expression of intimacy and attraction (see Kleinke, 1986).

During a conversation, gaze has two key functions: On the one hand, it can help to initiate, maintain and regulate a conversation (Kendon, 1967; for a review, see Cañigueral & Hamilton, 2019). As an example, gaze behavior can be used to control the conversation flow by looking away from the interaction partner during periods of speaking and looking back at them to signal that it is now their turn to talk (Ho et al., 2015). On the other hand, the eyes and the region around them can be used to express complex emotions and mental states (Baron-Cohen et al., 1997; Baron-Cohen & Cross, 1992a; Dimberg et al., 2000; Lee & Anderson, 2017).

Generally speaking, eye contact has positive effects on the evaluation of other people and has an impact on ratings of likeability, attractiveness, competence, intelligence and mental health, to name only a few (Kuzmanovic et al., 2009; Mason et al., 2005; Senju & Johnson, 2009; Wheeler et al., 1979; Kleinke, 1986). A potential mechanism behind this connection is that direct gaze is perceived as a sign of another person's interest in us and their motivation to approach us (Adams & Kleck, 2005; Cook, 1977; Frischen et al., 2007; J. K. Hietanen, Leppänen, Peltola, et al., 2008). Chapter 2 of this dissertation hypothesizes that gaze behavior during a conversation provides the basis to infer a bouquet of characteristics of the other person, such as their degree of empathic understanding and mental perspective taking. The paradigm employed in the two following studies was designed to be more akin to real-life scenarios by using naturalistic video stimuli and by explicitly considering context variables, particularly the emotional valence of the conversation. In Manuscript 4, we tested the effect of gaze behavior and emotional context on the perception of personal characteristics of the conversation partners and of the relationship between the interlocutors. Manuscript 5 builds on these initial findings by further assessing whether these effects translate onto prosocial behavior in economic games and how person perception and prosocial decision-making are interrelated.

MANUSCRIPT 4

Breil, C. & Böckler, A. (2021). Look away to listen: The interplay of emotional context and eye contact in video conversations. *Visual Cognition.* DOI: 10.1080/13506285.2021.1908470

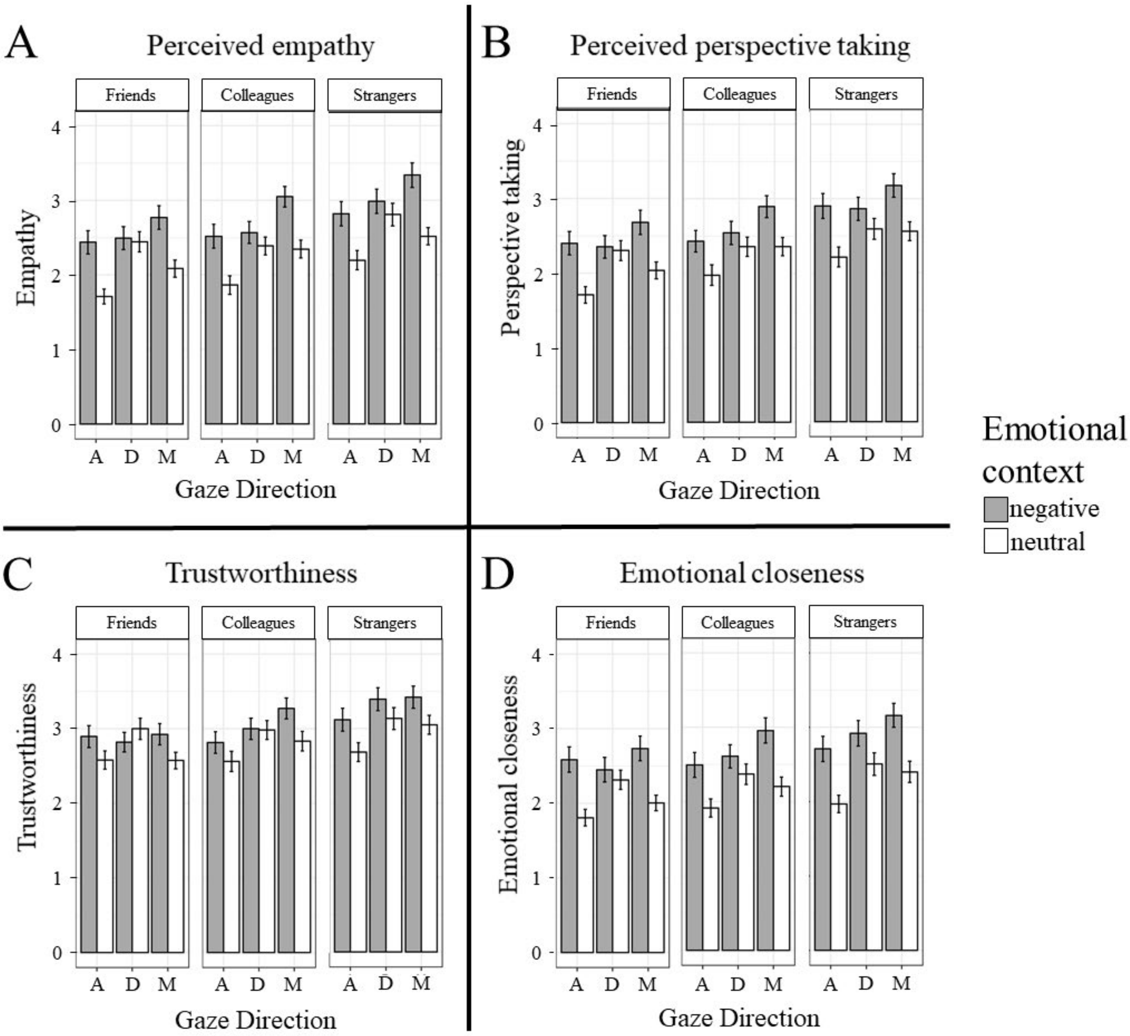
In this study, we investigated how gaze behavior shapes the perception of the conversation partner during a pre-recorded video conversation. We provided participants with short snippets of an allegedly longer video interaction and asked them to put themselves into the position of one of the conversation partners. In particular, the videos displayed the front view of one of two persons who were engaged in the conversation (*Figure 5*). This person ("the target") did not speak and remained mostly static throughout the video. Importantly, there were three different video conditions: constant direct, constant downwards-averted and mixed gaze behavior of the target person. The latter condition consisted of alternating direct and averted gaze. While participants watched the target person in the video, they listened to "the speaker" of the conversation, an unknown voice of a person that could not be seen during the video. This speaker recounted an autobiographical episode that was either neutral or negative in emotional valence. Based on previous studies with the same stimulus material, we assumed that stories of negative emotional valence would trigger an empathic response (e.g. Breil, Kanske, et al., 2021b; Kanske et al., 2015b). Participants were asked to empathize with the speaker while watching the video and to subsequently rate their perceptions of the target in terms of empathy, perspective taking and trustworthiness as well as to indicate the perceived closeness of the relationship between the two conversation partners.

Figure 5. *Screenshots of example videos from the direct and averted gaze conditions*

Note. The mixed gaze condition is formed by alternating direct and averted gaze direction.

Our findings clearly indicate that another person's gaze behavior as well as the emotional context of a conversation shape our perception of them (*Figure 6*).

Figure 6. *Mean ratings by condition for each relationship condition of Manuscript 4*



Note: Mean ratings per condition on a 6-point scale with separate graphs for each relationship condition and with gaze direction on the x-axis (A = averted gaze, D = direct gaze, M = mixture of averted and direct gaze). Panel A: Mean ratings of perceived empathy of the person in the video. Panel B: Mean ratings of perceived perspective taking of the person in the video. Panel C: Mean trustworthiness ratings of the person in the video. Panel D: Mean ratings of emotional closeness of the conversation partners.

Participants ascribed higher levels of empathy, perspective taking and trustworthiness to persons who engaged in direct gaze at least sometimes during a conversation (direct and mixed gaze). Furthermore, constant or partial direct gaze of the target person increased participants perception of the closeness of the relationship between the target person and the speaker. In contrast, constant gaze avoidance led to losses at all four rating scales. Finally, ratings of empathy and perspective-taking were higher when the target person was purportedly unknown to the speaker.

Most notably, the emotional context of the conversation shaped the effect of gaze behavior on person perception: while averted gaze during neutral narrations led to a break-in at all four ratings, gaze avoidance during emotionally negative stories was well accepted as evident by ratings that were largely similar to ratings after constant direct gaze. This interaction effect was even more pronounced when the target person switched between direct and averted gaze during an emotional conversation. Interestingly, with regard to ratings of trustworthiness, the described interaction effect was more pronounced when the conversation partners were allegedly less familiar with each other (i.e. colleagues or strangers).

These findings suggest that effects of gaze in social interactions are intricate and context dependent. Gaze behavior during social encounters is not interpreted in isolation but in relation to the matter of the conversation and the interpersonal history of the people involved. While establishing direct gaze has largely positive effects on the impression one can make on the interaction partner, certain situations allow for a deviation of the rule of making eye contact. One possible explanation for this finding might be that gaze avoidance during emotionally charged conversations serves to protect the private sphere of the speaker while allowing the listener to control their own emotional response (Kendon, 1967a). However, the exact source of these findings cannot be explained with the present data. Further studies, for example with an open answer format, would be necessary to investigate the precise nature of these effects.

MANUSCRIPT 5

Breil, C., Micheli, L.R. & Böckler, A. (unpublished). Golden gazes: Gaze direction and emotional context promote prosocial behavior by modulating perceptions of others' empathy and perspective-taking.

In the last section, I summarized a study showing that people who frequently engage in direct gaze during a conversation receive higher ratings of trustworthiness (Breil & Böckler, 2021). This is in line with previous studies reporting that faces with direct gaze are perceived as more trustworthy (Kaisler & Leder, 2016b) and that cooperation in economic exchanges increases when the players are allowed to look at each other (Behrens et al., 2020). Based on these findings, a straight-forward prediction would be that a person will also act in a more prosocial way towards other people who often establish eye contact. However, it has been noted before that self-report measures are susceptible to biases and can be unrelated to actual prosocial behavior (e.g. Böckler et al., 2016). Hence, it could be that gaze behavior during an observed conversation not necessarily translates into acts of prosociality. In Manuscript 5, we tested these questions by investigating the impact of gaze behavior and emotional valence on prosocial decision-making in economic games. These games can be seen as simple but coherent models of a variety of human encounters, hence providing an opportunity to measure actual behavior in an experimentally controlled setting (Murnighan & Wang, 2016).

In addition to the question whether frequent eye contact during a conversation increases prosociality on a behavioral level, we were interested in the connection of this effect to perceptions of empathy and perspective taking that we found in our previous study (see Manuscript 4). For example, if another person's gaze is related to increased trust behavior towards them, is this a direct connection? Or is the relationship between eye contact and trust modulated by the perception of the other person's social cognition? In other words, are people trusted more because they make eye contact or are they trusted more because they are perceived as being more socially competent? While a variety of previous studies report a connection between a person's individual level of empathy and their tendency to behave in a prosocial way (see Thielmann et al., 2020 for a meta-analysis), no study, to our knowledge, has yet assessed whether people who are perceived as more empathic are also treated with more trust and generosity by others. To answer these questions, we employed the paradigm of Manuscript 4 (Breil & Böckler, 2021) with an important twist: Instead of assessing trustworthiness of the target person on a rating scale, we provided participants with tokens that could be shared with the person in the video in a resource allocation game (experiment 1: Trust Game (Berg et al., 1995a); experiment 2: Dictator Game (Camerer et al., 2004)).

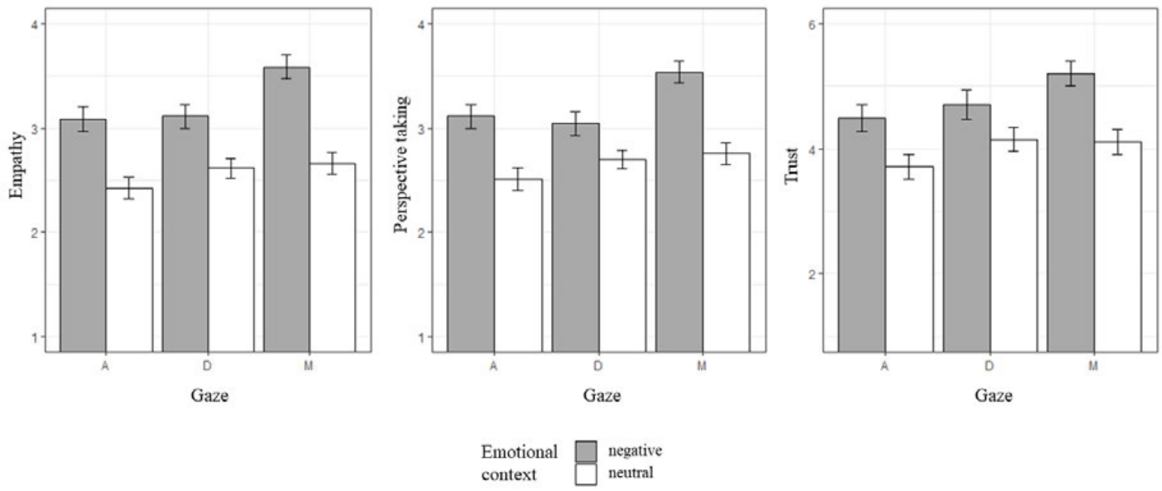
In experiment 1, participants in each trial were equipped with 10 tokens at a conversion rate of 1 token = 0,20€ and could decide freely how many tokens they wished to transfer to the person in the previously seen video. We told participants that the tokens they choose to invest would be tripled and transferred to the person they previously saw in the video. Participants believed that the trustee would now have a free choice to return a share of the tripled tokens back to them. Participant's payoff would be added to the remaining tokens (if any) that the participant kept and transferred as a bonus payment to the participant. Hence, in each of the six rounds, participants could decide whether they wanted to keep all tokens for themselves and fully receive the corresponding amount in euro or to invest (part of) their tokens in the person in the video. Participants were informed that at the end of the experiment, one round would be selected to determine their payment in the experiment. Participants were encouraged to take the decisions in each round seriously and independently of each other. Clearly, the investment choice in experiment 1 is a strategic one: transferring tokens to the person in the video could ultimately lead to a higher pay-off for the participant in case the trustee decides to return a generous amount of tokens as recompense. However, a cooperative choice of the participant could also lead to exploitation in case the receiver decides to leave this one-time interaction with full pockets. Because of this risk of losing money, the amount of tokens a participant sends to the person in the video in each round can be conceptualized as trust behavior (Berg et al., 1995a). Even though game theory predicts zero investments in one-shot Trust Games, a majority of participants send money and trust is generally reciprocated by the receivers (Stallen & Sanfey, 2013).

The results of experiment 1 show an intricate interplay of gaze behavior, emotional context of the conversation, perceived social understanding and trust behavior. We confirmed our previous findings from Breil & Böckler (2021) that both negative narrations and increased levels of eye contact lead to higher perceptions of the target person's empathy and perspective taking (*Figure 7*). Furthermore, we extended previous findings with a behavioral measure by demonstrating significant differences between averted and mixed gaze on trust. Finally, as seen previously, the effect of gaze behavior on perceived perspective taking was modulated by the emotional context of the conversation. However, in contrast to earlier findings, this interaction effect was not significant for perceived empathy, and we did not find differences in trust behavior between direct and averted gaze.

Furthermore, we confirmed our hypothesis that direct gaze and negative narrations elicited observable effects at a behavioral level in an economic Trust Game. Participants in experiment 1 invested more tokens in persons who were listening to negative stories or who engaged more in mixed gaze. Mediation analyses revealed that this link was fully accounted for by increased levels of perceived empathy and perspective taking of the target person

(*Figure 8*). Mixed (compared to averted) gaze behavior as well as negative (compared to neutral) stories made the target person appear more empathic, which in turn led to higher levels of perceived perspective taking and, ultimately, increased trust behavior. Importantly, we showed that perceptions of empathy and perspective taking are distinct concepts that influence each other but exert independent influences on trust decisions.

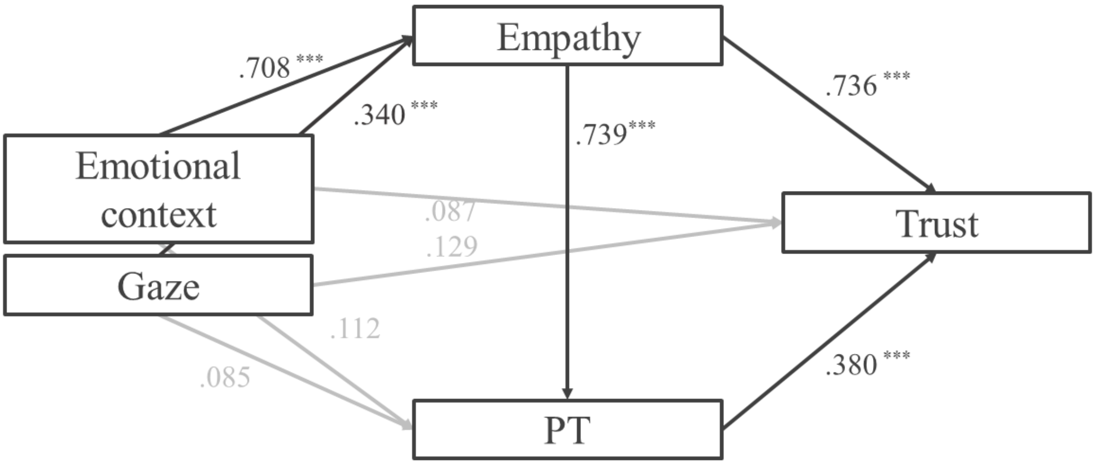
Figure 7. *Mean ratings and investment decisions per condition of Manuscript 5, experiment 1*



Note. Mean ratings of empathy and perspective taking on a 6-point scale; trust decisions on a 10-point scale. Gaze direction on the x-axis (A = averted gaze, D = direct gaze, M = mixture of averted and direct gaze). Panel A: Mean ratings of perceived empathy of the person in the video. Panel B: Mean ratings of perceived perspective taking of the person in the video. Panel C: Mean investment decisions in the Trust Game.

Despite the fact that the Trust Game is a simplified model of social decision-making, our findings are not necessarily unequivocal because people can make the same decisions for different reasons (Murnighan & Wang, 2016). One possible explanation for the findings of experiment 1 is that participants were generally more prosocial to the person in the video when they perceived this listener to be socially competent. In this scenario, the decisions in the Trust Game could be considered as generous or even altruistic. However, another possibility is that participants tried to stimulate reciprocity in people with (presumably) higher levels of social understanding. In this case, decisions in experiment 1 were more strategic (in contrast to generous).

Figure 8. *Results of the mediation analysis in experiment 1 of Manuscript 5*



Note. PT = perspective taking. Solid bars represent significant effects while grey bars represent not significant paths. Mixed vs. averted gaze and neutral vs. negative emotional context as independent variables; empathy and PT as mediators; trust as outcome measure. The effect of emotional context and gaze on trust was fully mediated by perceptions of empathy and PT.

To disentangle these possible explanations, we conducted a second experiment and tested whether the perception of other people's social cognition generalizes onto prosocial behavior when participants' earnings do not (allegedly) depend on the target person's decision. In a preregistered experiment, we repeated the procedure of experiment 1 but replaced the Trust Game with a Dictator Game (Camerer et al., 2004). The latter paradigm is an even simpler behavioral model because recipients are given no power to respond to the participant’s decision. Similar to experiment 1, participants in experiment 2 were equipped with 10 tokens in each round and could freely decide how many tokens they wished to transfer to the target person. Since the interactions ended at this point, tokens were not invested but anonymously donated in this game. As such, the situation mirrors one of charitable giving in real life.

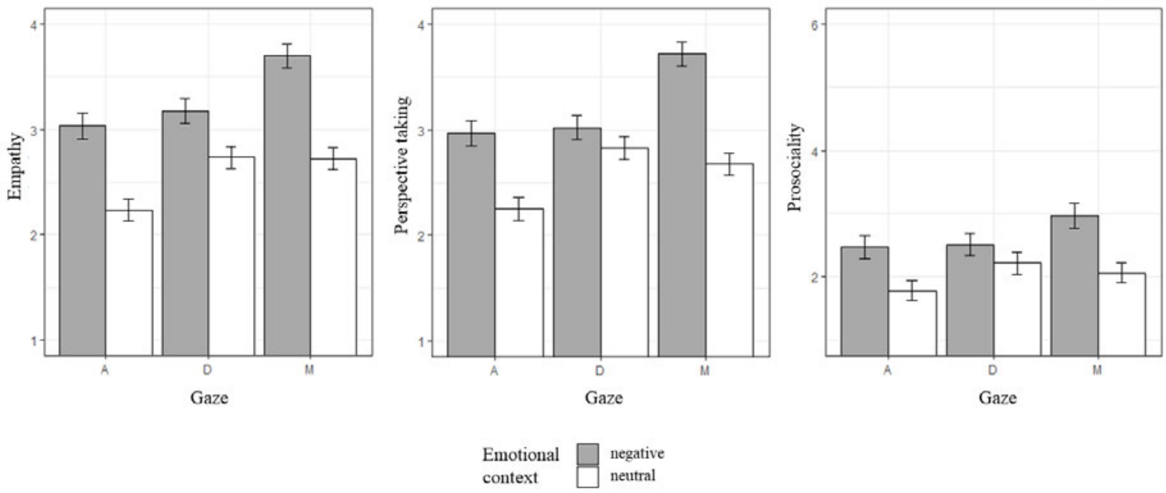
The results of experiment 2 match the effects that were observed in experiment 1, suggesting that decisions in both economic games (Trust Game and Dictator Game) were based on generosity instead of strategic thinking. Participants in experiment 1 ascribed higher level of social understanding to target persons who engaged in mixed gaze frequently (*Figure 9*).

Mirroring the findings from our seminal study (Breil & Böckler, 2021), the effect of gaze on perceived empathy and perspective taking was modulated by the emotional context of the conversation: gaze avoidance led to decreased perceptions of social understanding during neutral conversations but was accepted when the narrator’s story was emotionally negative.

We found corresponding effects for donation behavior in the Dictator Game. Consistent with findings of experiment 1, the effects of gaze and context on donations in the economic game were fully mediated by the perceived social skills of the target person: mixed (compared to averted) gaze and negative (compared to neutral) stories increased perceptions of empathy and perspective taking of the target person which ultimately led to more generous donations (*Figure 10*).

Although the interaction between gaze behavior and story valence became significant in the simple effects analyses, the interaction term did not significantly account for variations in giving behavior. Similar to experiment 1, perceived empathy and perspective taking were correlated but distinct concepts.

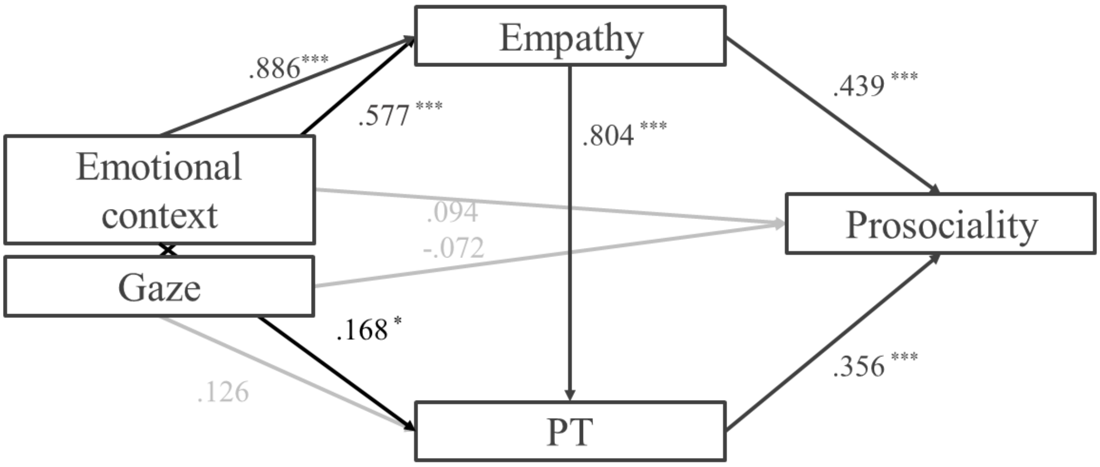
Figure 9. *Mean ratings and donation decisions per condition of Manuscript 5, experiment 2*



Note. Mean ratings of empathy and perspective taking on a 6-point scale; prosociality on a 10-point scale. Gaze direction on the x-axis (A = averted gaze, D = direct gaze, M = mixture of averted and direct gaze). Panel A: Mean ratings of perceived empathy of the person in the video. Panel B: Mean ratings of perceived perspective taking of the person in the video. Panel C: Mean donation decisions in the Dictator Game.

Taken together, these findings suggest that mixed (compared to averted) gaze during a conversation increases prosociality on a behavioral level. Importantly, this effect was fully accounted for by increased perceptions of the conversation partner's social cognition skills. Experiment 1 showed that mixed gaze and negative narrations increased perceptions of empathy and perspective taking of the conversation partner, which ultimately led participants to place more trust in this person. Experiment 2 replicated and extended these findings, suggesting that participant's choices in this study were not merely strategic but (also) driven by generosity or even altruism.

Figure 10. *Results of the mediation analysis of experiment 2 of Manuscript 5*



Note. PT = perspective taking. Solid bars represent significant effects while grey bars represent not significant paths. Mixed vs. averted gaze and neutral vs. negative emotional context as independent variables; empathy and PT as mediators; prosociality as outcome measure. The effect of gaze and emotional context on prosociality was fully mediated by perceptions of empathy and PT.

INTERIM SUMMARY

I summarized two studies that show that gaze behavior and emotional context can shape how a person is perceived by their social interaction partner and how this impression affects trust and generosity on a behavioral level.

Manuscript 4 demonstrated that, in online video conversations, effects of gaze behavior are shaped by the emotional context of the conversation. While gaze aversion led to reduced perceptions of empathy, perspective taking, trustworthiness and relationship closeness, (partly) averted gaze was well accepted when negative emotional story matter was told in a social interaction.

Manuscript 5 built on these findings, suggesting that people who are perceived as highly empathic also receive higher attributions of perspective taking and, as a result, are treated with more trust and generosity in economic games.

Taken together, these findings suggest that gaze cues exert a strong influence on cognition that goes beyond a modification of early attention allocation (see chapter 1 of this dissertation). Gaze behavior is not perceived in isolation but rather interpreted in the context of the conversation. As such, eye contact during a social interaction affects how we evaluate other people and how we behave towards them.

LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

One question that remains open is whether the perception of a socially skilled person increases generosity towards other people in general or whether the desire to act prosocial is limited to the conversation partner. To answer this, future studies should adopt our paradigm from Manuscript 5 to investigate giving behavior towards a stranger (i.e. a person that was not involved in the conversation and is unknown to the participant).

In sum, the findings of this chapter highlight the importance of studying the effects of gaze in rich and naturalistic contexts. Although a thorough and experimentally controlled investigation of the fundamental mechanisms of gaze perception are key to a profound understanding of social cognition, future research should explicitly consider the influence of context variables by adopting naturalistic settings. The integration of gaze behavior with other social (and) context information is a highly dynamic process. Hence, paradigms targeted at capturing effects of gaze in social interactions should be designed in a way that accounts for the complexity of the underlying mechanisms. Future research can benefit from adapting a second-person approach with live interactions or virtual reality in order to increase the ecological validity of findings (Schilbach et al., 2013).

CHAPTER 3: ASSESSMENT OF SOCIAL UNDERSTANDING AND EYE GAZE IN INTERPERSONAL INTERACTION

A crucial component of successful interaction is the ability to understand another person's affective and cognitive states, also referred to as empathy and ToM, respectively. Even though these terms have often been used synonymously and both capabilities often go hand in hand, evidence accumulates that they can be clearly distinguished on a behavioral, neuronal and developmental level (e.g. Kanske et al., 2015; Reiter et al., 2017; Schurz et al., 2020; Singer, 2006). Empathy is linked to, for example, limbic and para-limbic structures (de Vignemont & Singer, 2006; Oliver et al., 2018; Singer & Lamm, 2009) while ToM is strongly associated with activity in the temporo-parietal junction and the medial prefrontal cortex, among others (Frith & Frith, 2006; Schurz et al., 2014). Both capacities are subject to a lifelong development, with early evidence of social understanding already within the first years of life and independent maturation processes that extend well beyond adolescence (Decety & Michalska, 2010; Devine & Hughes, 2013; Dumontheil et al., 2010; Kanske et al., 2015b; Martin & Clark, 1982; Reiter et al., 2017; Wellman et al., 2001).

THE ROLE OF GAZE BEHAVIOR IN SOCIAL UNDERSTANDING

Research on social understanding has largely focused on neural substrates, developmental processes and alterations in clinical populations. Nevertheless, the connection to gaze behavior has received at least some attention in recent years.

Findings on the relationship between empathy and the establishment of eye contact have painted a clear picture: Highly empathic individuals spend more time looking at emotional eye regions (Cowan et al., 2014; Moutinho et al., 2021). This relationship is further strengthened by the finding that both the individual level of empathy and the individual time looking on emotional faces can be enhanced by oxytocin (OT) administration (Le et al., 2020). On a clinical level, callous-unemotional traits in children and psychopathic traits in adults, both characterized by a fundamental lack of empathy, are related to a decreased amount of eye contact during real life social interactions (Dadds et al., 2014; Gehrer et al., 2020).

In contrast, findings on the relationship between ToM and eye contact are less clear and it is still not known whether one directly leads to the other. However, there is a wide range of findings that indicate an (indirect) link between ToM and eye gaze. As an example, research on visual perspective taking, action coordination and shared attention, which are all linked to social competences, suggests that gaze behavior is fundamentally involved in these processes (Brennan et al., 2008; Clark & Krych, 2004; Symeonidou et al., 2016). Furthermore, studies from the developmental domain indicate that children's joint attention behavior, such as the number of gaze switches between looking at an adult and looking at a toy, can predict ToM skills two years later (Charman et al., 2000; Vaughan Van Hecke et al., 2007). Likewise, clinical research suggests that autistic individuals' responsiveness to joint attention at childhood is related to social skills in adulthood (Gillespie-Lynch et al., 2012). Notably, even though both distorted face processing and impaired ToM are cardinal symptoms of autism spectrum disorder (ASD), autistic individuals have not been found to fixate less on the eye region of another person, but rather to be impaired in holistic processing of faces (Baron-Cohen, 2000; Kirchner et al., 2011; Senju et al., 2008). On a neural level, research suggests a shared basis of ToM and joint action, action coordination as well as eye contact, respectively (Jiang et al., 2016; Newman-Norlund et al., 2008; J. H. G. Williams et al., 2005). However, whether and how the establishment of eye contact is directly linked to ToM is yet poorly understood.

THE IMPACT OF TASK CHOICE IN SOCIAL UNDERSTANDING

While it is a commonly accepted fact and largely established routine in psychological research that a psychometric task needs to be applied in the appropriate and intended age group, methodological requirements in ToM research go well beyond this demand. ToM is a multi-faceted construct that calls for a broad assessment with multiple-task batteries in order to be validly captured in all of its aspects. Manuscript 6 reviews recent findings in ToM research, highlights the striking link between task choice and outcome in this field and pleads for a more holistic assessment of social cognition. Importantly, this manuscript contains pilot data that suggests a relationship between gaze behavior and social understanding in healthy adults.

SOCIAL UNDERSTANDING IN ADOLESCENTS

Despite the abundance of studies on social understanding that were published in the last decades, some populations are still heavily unresearched. Particularly studies on ToM have largely focused on the developmental processes in children and on alterations in clinical populations, such as ASD and schizophrenia (e.g. Bora et al., 2009; Cadinu & Kiesner, 2000; Deschrijver et al., 2016; Wellman et al., 2001). While healthy adults increasingly came into the focus of research in recent years, the teenage population remains largely neglected. Pioneer studies found that empathy and ToM continue to undergo major behavioral and neural developmental processes in teenage years. However, findings are still too scarce to paint a conclusive picture (Allemand et al., 2015; Blakemore et al., 2007; Eisenberg et al., 2009).

Notably, the underrepresentation of adolescent sample groups in social cognition research is accompanied by a lack of appropriate measures to validly capture social understanding in this age group. Particularly with regard to the assessment of ToM, the existing measures were largely designed for children and/or clinical sample groups or for healthy adults. Hence, these tasks are either too easy or too difficult for teenagers, resulting in ceiling or floor effects when applied in this age group. Furthermore, the majority of ToM tasks relies on a binary response format that further obstructs the possibility of capturing variance in mental state representation in healthy and adolescent/adult individuals. A final shortcoming of classical ToM measures is the limited ecological validity that is introduced by the employment of paper-pencil tests to study human social interaction. The final manuscript of this dissertation, Manuscript 7, introduces a new paradigm that was designed to overcome these deficiencies by applying close-to-real-life video stimuli followed by multiple-choice questions, RT assessment and valence ratings to investigate ToM and empathy in teenagers. This is important groundwork for future research on social understanding in healthy and clinical adolescent sample groups. In addition, supplemental material to Manuscript 7, added exclusively to this dissertation, contains initial findings on the relationship between gaze behavior and social understanding in this age group.

MANUSCRIPT 6

Breil, C. & Böckler, A. (2020). The Lens Shapes the View: on Task Dependency in ToM Research. *Current Behavioral Neuroscience Reports, 7 (2), 41-50.* DOI: 10.1007/s40473-020-00205-6

This manuscript is a narrative review that summarizes the current status in ToM research with a special focus on the link between task choice and outcome specifics.

The ability to predict and explain the behavior of others has been in the focus of research for more than four decades and, over the years, a broad range of paradigms to investigate this multifaceted capability have been developed. Since none of the existing paradigms can fully capture ToM in its complexity, findings from different studies are heterogenous and their outcome largely depends on the task and stimulus materials that were employed (Schurz et al., 2014).

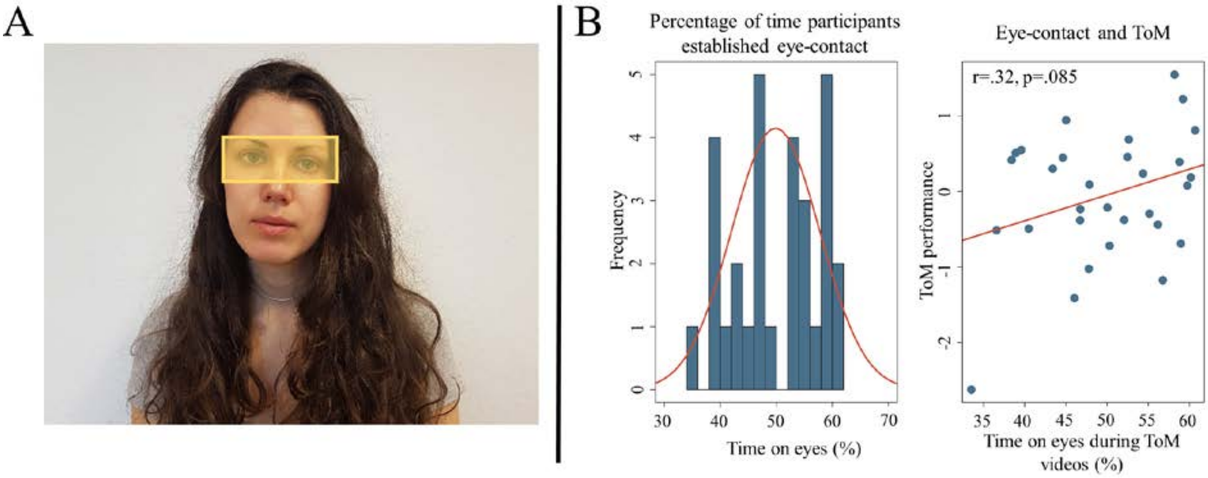
The development of ToM starts as early as in the first months of human life, and it is generally agreed on that this is a sequential process rather than a single accomplishment. While some studies that incorporated implicit measures, such as spontaneous gaze behavior, suggest that infants below the age of 2 years can represent (simple) mental states of other people (Baillargeon et al., 2010; Kovacs et al., 2010), performance in these tasks largely depends on formal and substantial characteristics of the paradigm (see Barone et al., 2019 for a review and meta-analysis). From 4 years on, a step-wise acquisition of ToM is accompanied by a progressive understanding of subjectivity, and linguistic task aspects increasingly modulate task performance (Atkinson et al., 2017; Burnel et al., 2018; Wellman et al., 2001; Wellman & Liu, 2004). Accordingly, inconsistencies between studies that worked with children of this age group can often be explained by differences in linguistic demands or in task complexity (Atkinson et al., 2017; Burnel et al., 2018; Kamawar & Olson, 2011; Miller, 2009; Rakoczy et al., 2015; Sullivan et al., 1994). The multi-faceted nature of ToM unfolds throughout adolescence and the endeavor of “measuring” this capability becomes increasingly difficult (Dumontheil et al., 2010; Osterhaus et al., 2016). Higher-order ToM appears to entail a large bouquet of distinct socio-cognitive competences, such as reasoning about rational behavior and about ambiguity (Osterhaus et al., 2016; Schurz et al., 2014), and hence decisions about the task and its specific aspects largely influence the outcome in ToM research. As a result, the endeavor of capturing ToM development throughout adolescence and adulthood requires an informed selection of a multiple-task battery.

Likewise, the effects of training programs targeting ToM in children, adults and older people critically depend on the specific content. Success rates and generalizability of such programs can be enhanced by tackling the whole spectrum of socio-cognitive competences with diversified and more true-to-life procedures (Begeer et al., 2011; Cavallini et al., 2015; Golan & Baron-Cohen, 2006; Kandalaft et al., 2013; Ozonoff & Miller, 1995).

One major hurdle is the challenge of capturing inter-individual variability in healthy adults. In neuroimaging research, relatively easy tasks are sufficient to stimulate ToM-related processes in order to capture the underlying neuronal processes. In contrast, investigating differences between healthy adult participants requires paradigms that encompass a higher level of difficulty in order to circumvent the typical ceiling-effects that are found with classical ToM paradigms. A particularly fruitful approach to overcome this issue is the incorporation of continuous measures, such as RTs, and more sophisticated task materials, such as video stimuli in combination with open or multiple-choice questions. Some recent examples of ToM paradigms that combine all or some of these innovations are the Edinburgh Social Cognition Test (Baksh et al., 2018), the Strange Stories Film Task (Murray et al., 2017) or the EmpaToM (Kanske et al., 2015b). All of these tasks benefit from considerably higher ecological validity due to their dynamic and more naturalistic video material (Alkire et al., 2018; Rice et al., 2016; Schilbach et al., 2013). Since its initial application in the study of 2015, the EmpaToM has proven to be sensitive for social cognition changes across the lifespan (Reiter et al., 2017) and for training-induced neural plasticity (Trautwein et al., 2020). Furthermore, an adapted version with reduced complexity exists that allows for simultaneous assessment of empathy and ToM in adolescents (Breil, Kanske, et al., 2021b, see Manuscript 7) and that could be applicable to clinical samples with relatively mild social cognition impairments, such as depressive or social anxiety disorders (Berecz et al., 2016;

Washburn et al., 2016). As such, the EmpaToM and its adaptations provide a promising basis for longitudinal studies of social cognition in healthy and clinical sample groups. Furthermore, pilot findings indicate idiosyncratic relationships of gaze behavior with empathy and ToM in healthy adults (*Figure 11*). Specifically, high empathic responding was related to shorter durations of eye contact with the narrator in negative videos while increased durations of eye contact were marginally correlated to better performance at subsequent ToM questions. These findings emphasize the far-reaching consequences of gaze behavior that pass beyond a modulation of low-level human attention onto higher-order social cognition.

Figure 11. *Pilot findings of the EmpaToM with eye tracking in Manuscript 6*



Note. Panel A: Example for the region of interest (eye region) for one of the narrators in the EmpaToM. Panels B: Pilot findings of gaze behavior in relation to Theory of Mind (ToM). The histogram displays how much time participants spent looking at the eye region during the videos (in percent). The scatter plot shows the correlation between the relative duration participants looked at the eye region during ToM videos and performance in ToM questions (composite score integrating speed and accuracy).

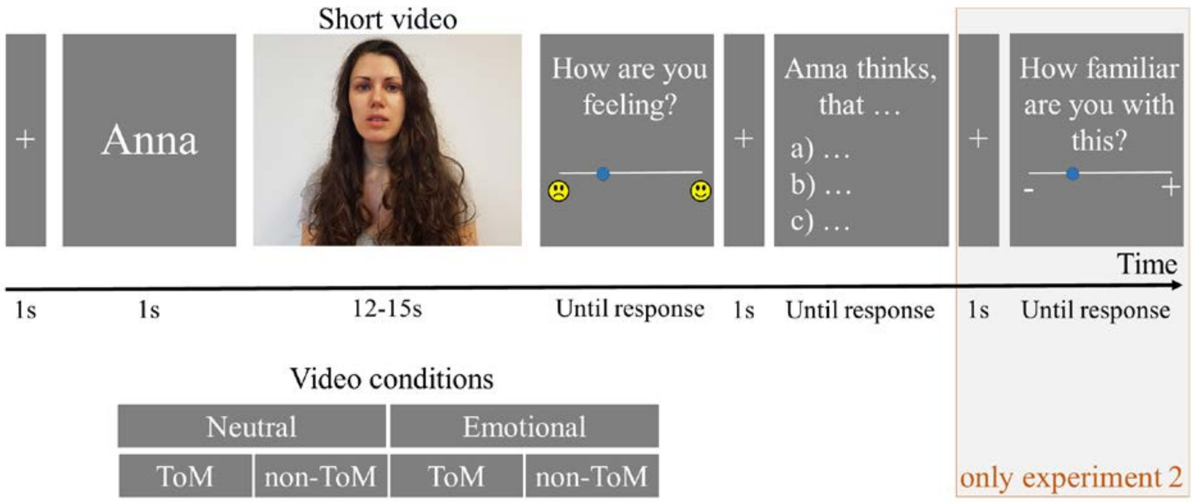
To conclude, research on ToM has come a great way since its first reference more than 40 years ago (Premack & Woodruff, 1978). Over the years, a vast amount of findings has accumulated and, more recently, systematic reviews and meta-analyses made an effort to integrate them (e.g. Barone et al., 2019; Heyes & Frith, 2014; Schurz et al., 2014, 2017, 2020). The compelling conclusion of these studies is that ToM is an intricate and multi-faceted construct that cannot be clearly distinguished from other higher-level cognitive processes such as language abilities and executive functions (Aboulafia-Brakha et al., 2011; Atkinson et al., 2017; C. M. Heyes & Frith, 2014; Saxe et al., 2006; Wade et al., 2018). Based on this notion, we endorse further research on the specific link between task characteristics and their outcome in ToM research that can foster the development of multiple-task batteries. These collections could in turn provide the foundation for further research on socio-cognitive capacities and a more profound understanding of the nature of ToM, its relation to other mental capacities and its lifelong developmental process.

MANUSCRIPT 7

Breil, C., Kanske, P., Pittig, R. & Böckler, A. (2021). A revised instrument for the assessment of empathy and Theory of Mind in adolescents: Introducing the EmpaToM-Y. *Behavior Research Methods.* Doi: 10.3758/s13428-021-01589-3

This manuscript reports the validation study of a novel instrument to investigate social understanding in neurotypical and clinical teenage samples. As highlighted in the previous sections of this chapter, research on empathy and ToM in adolescents is scarce and further obstructed by several shortcomings of the existing methods. Having identified this gap, we introduced a new instrument for simultaneous assessment of both capacities that is especially tailored to capture inter-individual variability in the ongoing developmental phase between child- and adulthood. Our new task, the EmpaToM-Y, is based on the EmpaToM (Kanske et al., 2015b). This paradigm employs naturalistic video stimuli followed by a rating of empathic affect sharing and multiple-choice questions with RT measurement for ToM assessment (*Figure 12*).

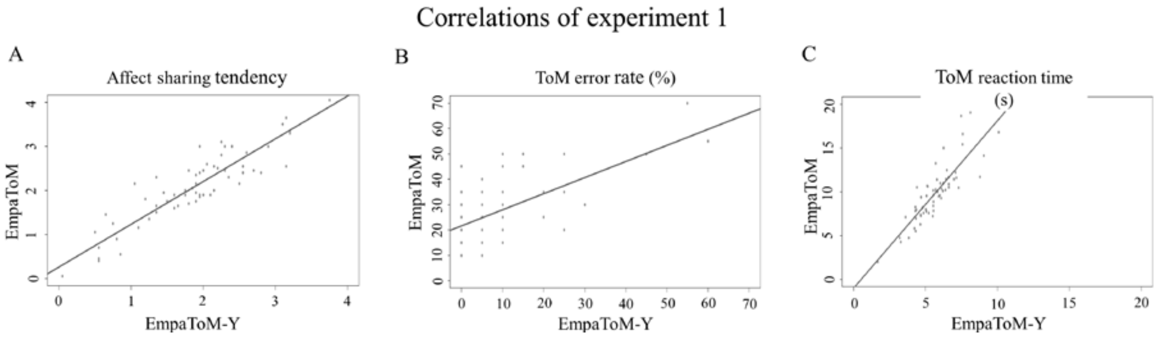
Figure 12. *Trial sequence of the EmpaToM-Y of Manuscript 7*



Note. After a fixation cross and the name of the person in the video are displayed for 1s each, a short video (12-15s) is played. The video is followed by a rating scale measuring empathic affect and a multiple-choice question for Theory of Mind assessment or factual reasoning, both displayed until a response is made. In experiment 2, this is followed by a second rating question to assess familiarity with the situation in the video.

We created youth-oriented story matter as a basis for new video stimuli and behaviorally tested our adapted version in conjunction with the original EmpaToM in a large adult sample as a first validation study (experiment 1). The results were promising, with high correlations for ToM error rates and RTs as well as for empathy ratings between the EmpaToM and the EmpaToM-Y (*Figure 13*).

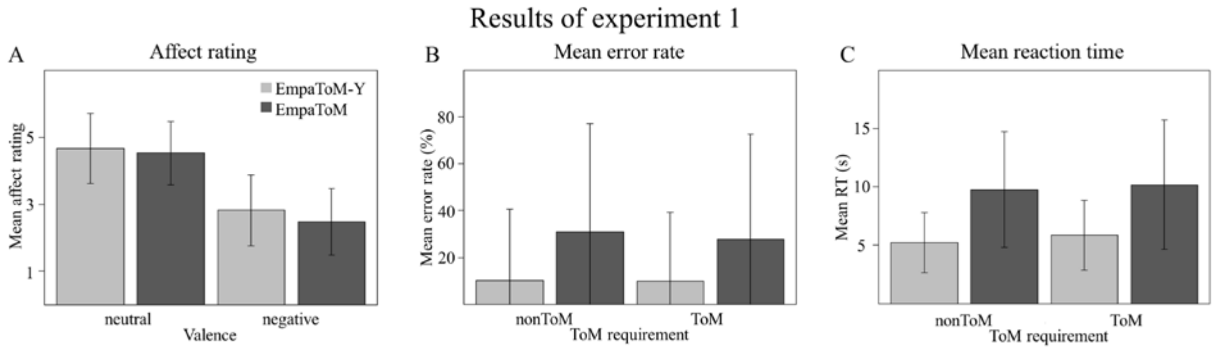
Figure 13. *Correlations of affect sharing tendencies as well as errors and response times in Theory of Mind questions between the EmpaToM and the EmpaToM-Y in experiment 1 of Manuscript 7*



Note. ToM = Theory of Mind. Panel A: Correlation of affect sharing tendency (difference between ratings after neutral and negative videos) between the EmpaToM and the EmpaToM-Y. Higher values indicate a higher individual tendency for empathic affect sharing. Panel B: Correlation of individual percentages of error rates for ToM questions between the two tasks. Panel C: Correlation of mean response times for questions with ToM requirements between both measures.

Furthermore, the direct comparison between both measures showed that the EmpaToM-Y was considerably easier than the original EmpaToM whilst still capturing inter-individual variability even in healthy adults (*Figure 14*). Based on these findings, we tested the feasibility of the new task in the intended age group by employing the EmpaToM-Y in a sample of 40 teenagers (aged 14-18 years) and augmented it with pupillometry and electrodermal activity (EDA) measurement as indicators of physiological arousal (experiment 2). Most notably, we found that, in contrast to findings in the adult sample in experiment 1, ToM questions were more difficult to adolescents than control questions, indicating that this capacity was not yet fully emerged in the young participants (*Figure 15*).

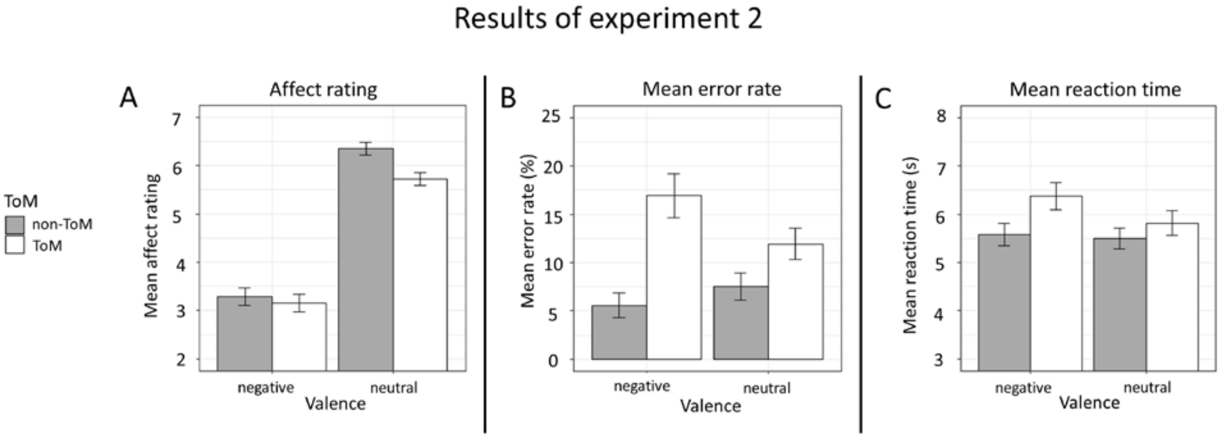
Figure 14. *Absolute affect ratings, error rates and response times per condition in the EmpaToM and the EmpaToM-Y in experiment 1 of Manuscript 7*



Note. ToM = Theory of Mind. RT = response time. Error bars represent standard errors. Panel A: Mean affect ratings on a 7-point scale. Panel B: Mean error rates at questions in %. Panel C: Mean RTs to questions in seconds.

Besides providing evidence for an ongoing development of ToM (but not empathy) in the age range between 14 and 18 years, this study shows general feasibility of the EmpaToM-Y in adolescent sample groups and the opportunity to capture interindividual variability in healthy teenagers. In combination with the well-established EmpaToM, our new measure opens the door for longitudinal studies assessing developmental patterns of social understanding from adolescence through adulthood. Effects in pupillometry and EDA were small in effect size in our study, indicating that larger samples are needed to reliably assess whether adolescent empathic responding is reflected in pupil dilation and physiological arousal.

Figure 15. *Affect rating and performance results by condition of the EmpaToM-Y in the adolescent sample of experiment 2 of Manuscript 7*



Note. ToM = Theory of Mind. Panel A: Mean affect ratings on a 9-point scale. Panel B: Mean error rates at questions in %. Panel C: Mean response times to questions in seconds.

ADDITIONAL DATA NOT REPORTED IN MANUSCRIPT 7: GAZE BEHAVIOR

In addition to the measures reported in Manuscript 7 and its appendix, we tracked participants eyes in experiment 2 to infer whether inter-individual differences in social understanding were related to specific patterns of visual attention. As outlined in the introduction of chapter 3, previous findings suggest that higher levels of empathy are related to an increased establishment of eye contact (Cowan et al., 2014; Dadds et al., 2014; Gehrer et al., 2020; Le et al., 2020; Moutinho et al., 2021). In contrast, while gaze behavior has been noted to play an important role in *ToM-related* processes such as joint attention and visual perspective taking, research on the direct relationship between eye contact and ToM is still scarce (Brennan et al., 2008; Charman et al., 2000; Clark & Krych, 2004; Symeonidou et al., 2016; Vaughan Van Hecke et al., 2007). We intended to bridge this gap by recording gaze behavior while participants watched a video of another person recounting an autobiographical episode and linking gaze patterns to subsequent performance in ToM questions and empathy ratings.

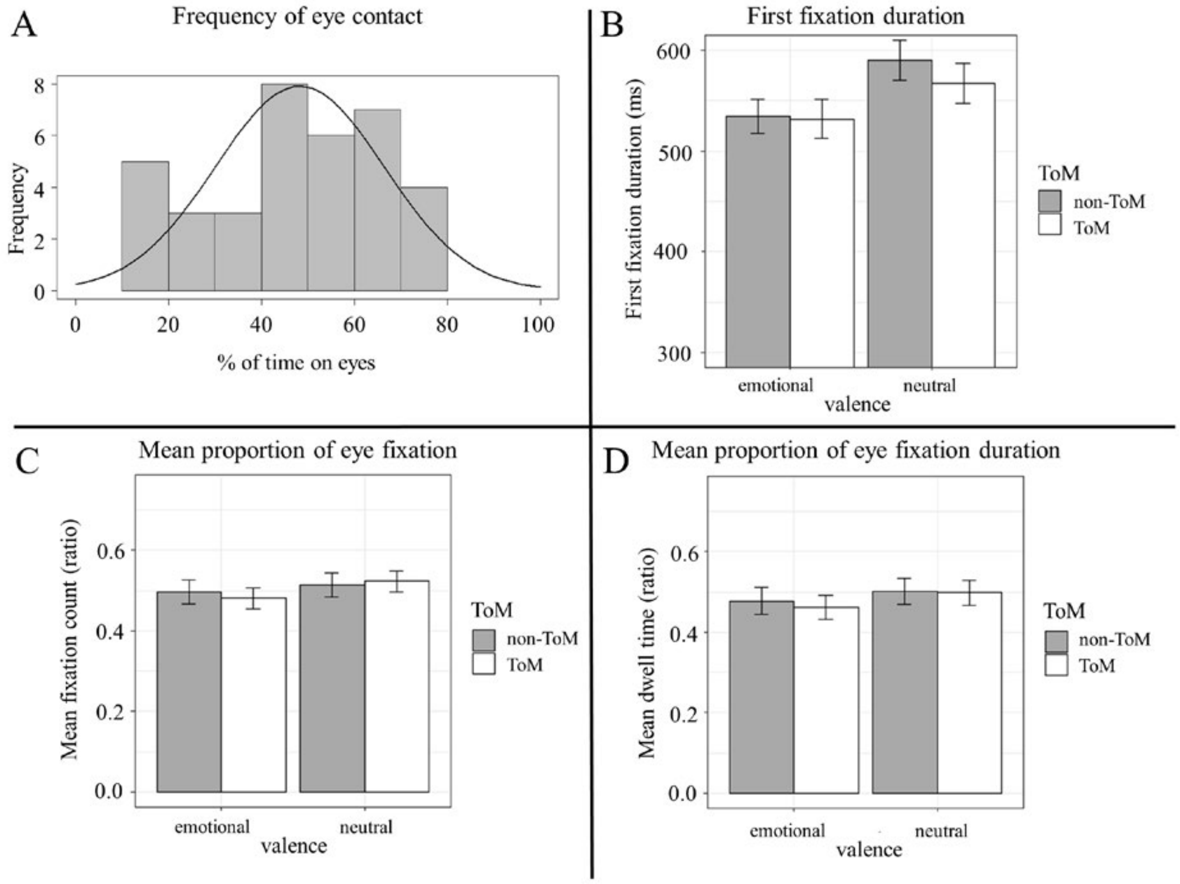
Procedure. The gaze behavior during the videos was tracked with an EyeLink 1000 Plus eye tracker (SR Research, Ontario, Canada) and analyzed with the corresponding software (Data Viewer version 3.2). Calibration was performed prior to training and to each test block. Additionally, a drift check was performed before every trial and calibration was initiated when the accuracy of the calibration parameters was too low. We created a dynamic interest area around the eyes of the narrator in each video. Because the eyes are the most preferred area of fixation in social scenes (Birmingham et al., 2008a, 2008b), we specifically considered the duration of the first fixation in this area. Of interest were furthermore the number of fixations falling on as well as the cumulative time spent on the eye region during the videos. Fixations within this region were set in relation to fixations that occurred on the total video screen (*eye ratio*) in order to control for measurement failures.

Analyses. We calculated separate 2 (video valence: neutral, negative) × 2 (ToM requirement: ToM, nonToM) repeated measures ANOVAs for the following dependent variables: (i) first eye fixation duration (duration of the first fixation in the eye region during the videos), (ii) eye fixation ratio (number of fixations on the eye region during the videos, in relation to all fixations on the screen) and (iii) eye dwell time ratio (cumulative time spent on the eye region during the videos, in relation to time spent on the screen). Post-hoc t-tests with Bonferroni-correction were performed to resolve ANOVA interaction effects.

To test whether gaze behavior is related to social understanding, we calculated the following correlations: (i) eye dwell time ratio and empathy rating, (ii) eye fixation ratio and empathy rating, (iii) eye dwell time ratio and accuracy at ToM questions, (iv) eye fixation ratio and accuracy at ToM questions.

Results. Findings on gaze behavior are visualized in *Figure 16*.

Figure 16. *Gaze behavior of adolescents during videos of the EmpaToM-Y in experiment 2 of Manuscript 7*



Note. Panel A: Distribution of the percentage of time during a video that the participants spent on the eyes of the narrator. Panel B: Mean duration of the first fixation of the eyes of the narrator. Panel C: Mean amount of fixation of the eye region in relation to fixations on the rest of the screen. Panel D: Mean cumulative time of eye fixation in relation to fixation time of the rest of the screen.

On average, the participants in experiment 2 spent 46.93% (SD = 18.03) of the time during a video on the eyes of the narrator. However, the variance between participants was large with a range between 0.69 to 84.44% of total video time (*Figure 16*, panel A).

Participants in experiment 2 avoided the gaze of the narrator more during videos with negative valence. This pattern was apparent both in terms of the relative number of fixations of the eye region compared to fixations of the total screen ($F(1,35) = 8.059$, $p = .007$, $\eta^2 = .008$; *Figure 16*, panel C) as well as in terms of the cumulative time that the participants spent on the eyes of the narrator compared to the total screen ($F(1,35) = 6.505$, $p = .015$, $\eta^2 = .006$; *Figure 16*, panel D). In line with this pattern, the first fixation on the eyes of the narrator was significantly shorter during emotional videos ($F(1,35) = 7.932$, $p = .008$, $\eta^2 = .039$; *Figure 16*, panel B). No other effects on gaze behavior were significant (all $p > .05$).

None of the tested correlations was significant after correction for multiple testing (all $p > .05$).

Discussion. We found consistent effects of valence on gaze behavior in this experiment. Adolescents in our study avoided direct eye contact with the person in the video more when this person recounted an emotionally arousing story. These effects were small but reliably found on all measures, i.e. number of fixations and the cumulative time spent on the eyes as well as the duration of the first eye fixation. Taken together, these results fit the idea of an avoidance of direct eye contact during an emotionally charged conversation in order to down-regulate one's own emotions (Kendon, 1967a). However, in contrast to previous findings, we did not find a significant relationship between gaze behavior and empathic affect sharing. Furthermore, our data do not indicate a relationship between gaze behavior and ToM in healthy adolescents (for different pilot findings in adults, see Breil & Böckler (2020); Manuscript 6). However, inter-individual variability was large on both measures and our sample size could have been insufficient to detect small effects. We endorse further research to test and replicate our findings on the relationship between social cognition and gaze behavior in larger groups.

The large variability in gaze behavior reflects profound differences in the time that individual participants established eye contact during the videos. Based on these results, a promising line of future research are investigations on the origins of inter-individual differences in gaze behavior and their relation to other social preferences and skills. Indeed, we found systematic differences between gaze behavior during videos with neutral and emotionally negative content. Albeit small, these effects consistently speak for an avoidance of direct eye contact in emotionally charged social situations: participants fixated the eyes of the narrator shorter and less often during videos with negative valence. In a similar vein, the first fixation on the eyes was briefer when the narrator was recounting an emotional episode from their life. Because the eyes are usually reported as one of the first regions visited during face perception (Hills et al., 2013), it can be assumed that the first fixation towards this area happened rather early during the trial, i.e. before the story unfolded its content. Hence, a shorter duration of the first fixation to the eyes presumably represents a reaction to the initial facial expression of the displayed person rather than to the story that was narrated. However, differences in gaze behavior were unrelated to subjective ratings of empathy and, similarly, our results suggest that differences in viewing patterns are unrelated to ToM performance in healthy adolescents.

INTERIM SUMMARY

In this chapter, I summarized two articles that focus on the assessment of human social understanding and provide evidence for a relationship between this capacity and gaze behavior.

In Manuscript 6, we adopted a broad view on the assessment of mentalizing processes. Putting a special emphasis on the characteristics of the specific paradigm and stimulus material, we highlight the striking relationship between task choice and outcome and make a call for more sophisticated and informed decisions on paradigm variations and control conditions in ToM research. Furthermore, we report pilot findings that speak to a relationship between gaze behavior and social understanding in healthy adults.

In the final study of this dissertation, we introduced a new instrument for the investigation of empathy and ToM in adolescent sample groups. Socio-cognitive research in this age group is scarce and accompanied by a lack of appropriate measures. Under these circumstances, the introduction of our new paradigm for the assessment of empathy and ToM in adolescents in Manuscript 7 is an important step towards a better understanding of developmental processes of social cognition, their precursors and their outcomes. Initial application in a group of teenagers showed that ToM capacities are still developing in the time period between 14 and 18 years. Eye tracking data that I added exclusively to this dissertation give initial evidence of a relationship between gaze behavior and emotional context in this age group while revealing a striking variability in gaze behavior between individuals.

LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

In conclusion, the articles of chapter 3 emphasize the importance of methodological aspects in the assessment of social cognition and give important impulses for future research in this field. We made a point for adopting a holistic view in social cognitive research that (1) employs naturalistic stimuli to increase ecological validity, (2) combines multiple methodological approaches to detect potential relations between behavioral, attentional and perceptual as well as physiological and gaze processes and (3) recognizes social cognition as a multifaceted construct that is subject to lifelong development and, hence, requires multi-methodological and longitudinal study designs across different age groups.

GENERAL DISCUSSION

In a social world, direct gaze is the foundation for communication and a functional group life. The studies that I reviewed in this dissertation highlight the manifold influences of gaze cues on human low-level and higher-order cognition.

In chapter 1, I summarized three studies that assess the boundary conditions of attention capture by direct gaze and how this effect is integrated with facial context information. All studies in this chapter employ modified versions of a target detection task adapted from Böckler et al. (2015; 2014) that orthogonally manipulates gaze and motion cues. We analyzed data of manual performance (RT, error rates) in all experiments as well as gaze behavior in experiment 2 of Manuscript 3.

In the first study (Manuscript 1), we conducted six experiments to systematically assess attention capture by ostensive cues with different levels of naturalistic and holistic social information. In particular, we employed photographic faces, arrows, isolated photographic or schematic eyes as well as schematic faces. We found the typical effects on attention only for photographic faces, suggesting that a socially meaningful facial context is a necessary prerequisite for the direct gaze effect. No effects of ostensive directional cues were found for any of the other stimuli employed in this study. In contrast, the motion effect remained stable across experiments (except for schematic faces with head turn), suggesting that it is less affected by context information. Taken together, these results indicate parallel but distinct underlying processing channels for attention capture by gaze and motion cues. This notion is in line with findings from previous studies (Böckler et al., 2015; Böckler, van der Wel, et al., 2014) and is further substantiated by the following two studies of this chapter.

The second study (Manuscript 2) consists of four experiments that parametrically manipulate the VA of stimulus perception. We found that the gaze effect collapsed while the motion effect was particularly strong in the peripheral visual field, where rod density is high but spatial resolution is low. We detected the opposite pattern for stimulus presentation in the central peripheral field: there was a large effect of gaze cues but no effect of motion. This pattern of results is in line with previous findings from functional magnetic resonance imaging (fMRI) studies that speak to a functional specialization of our visual system (Bressler et al., 2013; Roberts et al., 2007): while central vision appears to be involved in endogenous attention for fine detail perception, peripheral vision has been linked to action and object allocation in order to redirect foveal attention (Burnat, 2015; Corbetta & Shulman, 2002; Kowler, 2011). Furthermore, the results of Manuscript 2 show a generalizability of the gaze effect across different eccentricities, exemplifying the importance of spatial resolution abilities for attention capture by direct gaze and allowing us to pinpoint its boundaries: direct gaze is salient enough to capture attention only until a VA of 5.5°.

The last study of this chapter (Manuscript 3) further illustrates the impact of social context on the direct gaze effect by showing an emotion-specific integration process of gaze direction and emotion expression. In two experiments, we assessed the integration of gaze direction and facial emotion expression by concurrently presenting both cues in an adapted version of the target detection task by Böckler et al. (2015; 2014). We found that both cues flexibly interact in an emotion-specific fashion: expressions of anger and fear directly affected RTs at costs of the direct gaze effect. In contrast, expressions of happiness and disgust modulated the gaze effect in accordance with the approach/avoidance congruency hypothesis by Adams & Kleck (2003, 2005). In particular, manual responses to happy faces were facilitated when they depicted direct gaze while responses to disgusted faces were facilitated by averted gaze. Similarly, happy-direct and disgusted-averted faces were fixated earlier and longer than happy-averted and disgusted-direct stimuli. Eye tracking data furthermore suggest an early integration of both cues, with a process that starts around 275 ms for happy faces and around 375 ms for disgusted faces, and lasts around 225 ms for both emotions. A likely explanation for these effects is that happy and disgusted faces do not indicate potential threat in the environment and therefore pull less attention than expressions of anger and fear. Instead, they are immediately integrated with gaze information. In contrast, anger and fear are more aversive and emotional. This could be the reason why they are prioritized over gaze information at early processing stages (Ortony & Turner, 1990; Schupp et al., 2004; Vuilleumier, 2002a).

Taken together, the findings of this chapter give insight into the mechanisms and time course of integration of gaze direction with social and non-social cues. Investigating the basic attentional mechanisms that are driven by gaze cues will allow a better grasp on social attention, face processing and person perception in order to ultimately gain a comprehensive understanding of interpersonal communication and decision making.

There are two overall conclusions from the studies of this chapter: for one, direct gaze and motion cues rely on separate processing channels that work independent and in parallel. This notion is supported by the finding that the motion effect, in contrast to the direct gaze effect, remains relatively untouched by (social) context information and feature-related stimulus aspects (Manuscripts 1 and 3). The implementation of different stimuli across the three studies speaks to the reliability and generalizability of this finding. Furthermore, the findings of Manuscript 2 suggest that attentional effects of both cues are shaped by the properties of our visual system in their very own ways. Second, gaze direction is not perceived in isolation but is embedded in the context of other facial cues and social information. This finding suggests an interaction of bottom-up and top-down mechanisms that are part of a sophisticated system that is specialized in the processing of gaze direction.

Chapter 2 of this dissertation consists of two studies that investigate the impact of gaze and emotional context on the perception of our interaction partner as well as on the behavior towards them. Both studies employ a paradigm in which pre-recorded dyadic conversations followed by rating questions and economic games (Manuscript 5) are presented. Importantly, participants were instructed to adopt the perspective of one of the conversation partners while watching the videos in order to subsequently rate and interact with the second person.

The first study of this chapter (Manuscript 4) tested how the direction of gaze (direct, averted or mixed) and the emotional context of the story (neutral or negative) during a conversation shape the perception of the other person in terms of their social skills (empathy, perspective taking) and trustworthiness as well as the perceived closeness of relationship between the conversation partners. We found clear evidence that the gaze behavior as well as the emotional context of the story that is told influence how we perceive the other person and the relationship between the conversation partners. In particular, we found that persons who at least sometimes engaged in direct gaze during a conversation were perceived as more empathic and trustworthy, and to have a better understanding of the speakers' mental states. Furthermore, the relationship between the conversation partners was judged as more intimate by participants when the two people sometimes or constantly held eye contact. Importantly, these effects of gaze were modulated by the valence of the story that was shared in the conversation: During neutral narrations, gaze avoidance was perceived as an indication of low social skills, trustworthiness and emotional closeness. This was not the case during emotionally negative episodes, presumably because gaze avoidance was taken as an emotion regulation strategy and a means to give the narrator "space" (Kendon, 1967a). In sum, this study reveals complex and context-dependent effects of gaze in social interactions.

The second study of this chapter (Manuscript 5) builds on these findings by further assessing whether the effects translate onto prosocial behavior in economic games and, if they do, how person perception and prosocial decision-making are interrelated. To this end, we employed the same paradigm as in the previous study and quantified perceptions of empathy and perspective taking in a rating question format. Furthermore, we assessed prosocial behavior (trust or generosity) by means of economic games. In the first experiment, participants played the Trust Game (Berg et al., 1995a) with the person that was seen in the video while, in the second experiment, they played the Dictator Game (Camerer et al., 2004). In both experiments, participants were endowed with 10 chips in each round and could freely decide how many of the chips they would like to transfer to the person in the video. With regard to social skill ratings, we largely replicated findings from the first study of this chapter (Manuscript 4), with main effects of gaze and emotional context, as well as an interaction of both cues, albeit the latter effect was less reliably observed. Furthermore, we found similar effects on prosocial behavior: persons that displayed mixed gaze during the conversations received higher investments and donations in the economic games than people who avoided the gaze of the conversation partner, as did people who were listening to a negative compared to a neutral story. For generosity (i.e. non-strategic sharing in the Dictator Game in experiment 2), the emotional context of the story significantly modulated gaze effects in that gaze avoidance led to decreased sharing behavior only during neutral episodes. Importantly, mediation analyses in both experiments suggest that the main effects of gaze and emotional context on prosocial behavior were fully mediated by perceptions of empathy and perspective taking. In other words, individuals who displayed mixed gaze and/or listened to emotionally negative narrations were ascribed higher levels of social understanding which in turn led to higher investments and donations by participants.

The fact that we found increased levels of giving behavior in both experiments speaks to the fact that participant's decisions in the Trust Game (experiment 1) were not or not only driven by strategic considerations. Instead, participants were generally more generous and prosocial towards people who they considered to have higher social skills. These findings exemplify the striking influence of gaze cues on cognition and behavior and demonstrate once more that gaze is perceived and interpreted in context.

In the final chapter of this dissertation, chapter 3, I reviewed two articles that demonstrate the relationship between gaze behavior and social understanding and that advance methodological aspects of social cognition research.

This chapter starts with a narrative review (Manuscript 6) that summarizes and reviews recent findings in ToM research. A special focus lies on the impact of task choice on outcome in social cognitive research as well as on the (often neglected) importance of a holistic investigation of social cognition. Over 40 years of research on ToM has shown that this capacity emerges early in life. There is evidence that children can master easier versions of ToM tasks before they are able to understand or produce language, even though precisely this inability complicates reliable research in this age group and scientists are divided whether and which conclusions can be drawn from the existing studies. When language emerges, it is one of the factors that has profound impact on ToM performance in children. However, when controlling for this confound, results rather consistently indicate a step-wise acquisition of ToM with an increasing understanding of subjectivity. Over the teenage years, ToM unfolds in its complexity and a holistic assessment or training of all sub-components becomes increasingly difficult. A carefully selected, true-to-life multiple-task battery or training program is inevitable to master this challenge. In the past years, new paradigms have been developed that made headway with an ecologically valid ToM assessment and allow to capture inter-individual variability in healthy adults. An example of such a paradigm is the EmpaToM (Kanske et al., 2015b) that allows for simultaneous measurement of empathy and ToM in adults. Pilot data of an application of the EmpaToM with eye tracking substantiate the notion of dissociation between empathy and ToM by showing idiosyncratic relations of both capacities with gaze behavior. In particular, high empathic responding was related to increased gaze avoidance of the narrator while more eye contact was marginally related to better performance at subsequent ToM questions. These results exemplify the wide-ranging impact of gaze cues on higher-order cognitive processes. We close our review with a call for science that is specifically targeted at the relationship between task characteristics and outcome in ToM research in order to promote the development of more naturalistic multiple-task batteries which can foster a better understanding of ToM and its related processes.

The final manuscript of this dissertation, Manuscript 7, is the validation study of a new task that was specifically designed to overcome some of the previously discussed deficiencies and to fill a largely overlooked gap in ToM research. Based on the EmpaToM for adults (Kanske et al., 2015b), we designed a task to measure social understanding with naturalistic stimulus material in a group that received only little attention in recent years: adolescents. The task consists of 40 video items that are followed by multiple-choice questions and rating scales to assess empathy and ToM. We successfully validated the new EmpaToM-Y on the existing task in a group of adults, finding high correlations between the respective measures and considerably reduced error rates and RTs for the EmpaToM-Y. Hence, our task validly captures social understanding with reduced task complexity. In a second experiment, we applied the new task in a group of teenagers, proving general feasibility of the measures and providing evidence that ToM, but not empathy, is still developing between 14 and 18 years. Eye tracking data that were collected during this study and published exclusively with this dissertation once more indicate a relationship between gaze behavior and empathy: adolescents avoided eye contact with narrators who told an emotionally negative story. These results fit the general idea of gaze avoidance as a form of emotion regulation (Kendon, 1967a) and are in line with pilot findings on adults (see Manuscript 6).

Taken together, chapter 3 deals with the assessment of human social understanding and provides initial evidence for its link to individual gaze behavior both in adult and adolescent samples. Following the call for a more holistic and naturalistic investigation of social understanding in Manuscript 6, Manuscript 7 provides a task that was intended to advance assessment of empathy and ToM as well as their development in adolescent samples. Both studies contain eye tracking data that suggest a relationship between social understanding and gaze behavior in healthy samples. The articles of this chapter are groundwork towards an adequate life-long and all-embracing understanding of social cognition.

OVERARCHING FINDINGS

There are a few notable recurrences in the findings of the studies that I presented in this dissertation. Integrating them allows us to draw more general conclusions from my research.

One important and reliable observation is that context information is efficiently integrated with gaze behavior on several cognitive levels. As such, it affects not only our interpretation of social behavior but it also works on basic perceptual and attentional mechanisms. For example, we found that fundamental mechanisms such as attention capture by direct gaze collapse when naturalistic face information are missing. In Manuscript 1, we tested the effect of gaze direction on attention with six different ostensive stimuli and found evidence for the direct gaze effect only when the gaze information was embedded in a holistic and naturalistic facial context. Stimuli that lacked this context information, such as isolated eyes or schematic faces, failed to elicit the gaze effect. Similarly, we observed how quickly gaze direction and facial expression are integrated in Manuscript 3: when both cues were concurrently presented, we found a prioritization of emotion over gaze cues at early processing stages for threat-related expressions but a modulation of the gaze effect for biologically less relevant emotions. Eye tracking data suggest that this integration process occurs at early processing stages in an emotion-specific fashion. Taken together, the findings from chapter 1 suggest that gaze direction and face information are quickly and efficiently integrated to influence basic attentional mechanisms. Hence, gaze is not perceived in isolation but in context and gaze effects can be shaped by other cues.

Findings from chapters 2 and 3 support and extend this notion by showing that context effects in social situations also shape more complex and higher-order cognitive processes. In chapter 2, we observed that the emotional context of a conversation shapes how we interpret another person’s gaze behavior. In particular, we observed in Manuscripts 4 and 5 that (partial) gaze avoidance during neutral, everyday conversations makes a bad impression but is well accepted (or even rewarded) when highly emotional story matter is shared. This interplay of gaze behavior and emotional context has far-reaching effects on our perception of the other person's social skills which in turn shape our behavior towards them. More precisely, people who switched between direct and averted gaze while listening to a negative story were rated as more empathic and understanding of the narrator's mental states. As a result of this perception, they were trusted more with tokens in a Trust Game and received higher donations in a Dictator Game. In line with the emotion-regulation hypothesis (Kendon, 1967a, 1969), avoiding eye contact during emotionally arousing social situations could be a strategy to control one's own emotional state, or to give the other person more space.

This idea fits eye tracking data of chapter 3 suggesting that we also adapt our own gaze behavior to the emotional context of a conversation. Data of Manuscript 6 and results in the supplemental material of Manuscript 7 show that adult and adolescent participants avoided the gaze of the narrator more when they told a negative story. Hence, not only is our perception of another person's gaze shaped by the situational context (chapters 1 and 2). We also change our own gaze behavior depending on the emotional and mental state of our conversation partner (chapter 3).

Our findings have important implications for interpersonal communication in several areas of our lives. For example, workers in the health sector, such as medical doctors or psychotherapists, should deliberately adapt their gaze behavior to the situational context in order to create an atmosphere of safety and trust. In economy, both physical consultants and artificial agents should be trained to pay attention to the gaze behavior of the costumer as well as to establish or avoid eye contact depending on the situational and emotional context. Politicians could establish eye contact with the citizens to strengthen the impact of their speech and to build trust with the population. Of course, these techniques could also pay off in private communications, and individuals who struggle in social situations, such as patients with schizophrenia, ASD or social anxiety, should receive tailored gaze training to improve their social skills. At the same time, neurotypicals could receive inclusive education to prevent misinterpretations of altered gaze behavior. This guidance could ultimately help the affected patient groups build healthy relationships and increase their general well-being. Finally, future research should use our findings to explore the role of context on a broader scale, particularly in situations where eye contact is not possible and other sensory channels, such as hearing and touch, are even more important.

LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

While the studies in this dissertation are an important step towards a more ecologically valid assessment of gaze effects, they also demonstrate that research will need to strike on new paths and design novel paradigms and methods in order to fully capture the far-reaching effects of gaze behavior. A large part of research on social interaction has been conducted with stimuli that allow for high experimental control but provide no opportunity for social interaction. As a result, the social brain has been studied from an observer's perspective that is solely based on the representation of a third person's mind (Fuchs & De Jaegher, 2009; Lehmann et al., 2019a). Clearly, findings from research that is limited to an observer's perspective cannot generalize to full-blown social interactions as we experience them in our everyday lives. Particularly in the studies of chapter 2 of this dissertation, we have made effort to establish a more naturalistic setting and engage participants in a dynamic interaction. Nevertheless, the paradigms we employed miss central aspects of a natural, mutual interaction that typically involves moment-to-moment (re-)actions, coordination and resonance between two individuals. This criticism is not new and my dissertation lines up with findings from other studies that highlight the importance of studying social interaction in context (e.g. see Fuchs & De Jaegher, 2009; Lehmann et al., 2019). For example, perceiving live speech has been shown to result in enhanced neural activation in areas that are related to social cognitive processing and the mere belief that speech was live triggered activation in ToM-related areas such as the temporo-parietal junction or the precuneus (Rice et al., 2016). Similarly, peer-interaction, as opposed to listening and reacting to pre-recorded messages, activated regions of the mentalizing network as well as reward-related brain areas (Alkire et al., 2018). Underneath these findings lies the assumption that social cognition is fundamentally different during face-to-face social interaction compared to the mere observation of others (Krach et al., 2013; Laidlaw et al., 2011a; Schilbach et al., 2013).

Fortunately, the technology to "really go social" while ensuring a strictly controlled experimental setting already exists and recent research has advanced paradigms that incorporate more interactive elements or even live social interaction. For instance, virtual reality set ups allow for highly naturalistic *and* controlled stimulus presentation (Forbes et al., 2016; Kandalaft et al., 2013, 2013; Nijman et al., 2019). However, this technology comes with the drawback that participants are fully aware that they are interacting with an avatar and not with a real person. Perhaps a more promising approach is connecting two participants with live-video feed while the brain activation of one of them is recorded in an fMRI scanner (Redcay et al., 2010) or while both participant's eyes are tracked (Hessels et al., 2017, 2018). In a similar fashion, mobile eye trackers can be used to measure gaze behavior during live interactions (Freeth et al., 2013; Hanley et al., 2014; Magrelli et al., 2013) and motion trackers can help to analyze motor synchronization (Fitzpatrick et al., 2017; Romero et al., 2018).

Besides the problem of increasing ecological validity of one dependent variable, future research should aim to connect findings of different measures in order to form the "big picture". As an example, a combined eye tracking and fMRI study found that self-initiated joint attention in children and adolescents activated areas that are related to social cognition and social reward in adults (Oberwelland et al., 2016). Findings from another study combining both measures suggest that social interaction and social observation in adults are based on distinct neural circuits (Tylén et al., 2012). A particularly promising new line of research is focused on psychopharmacological mechanisms, such as the role of steroid hormones or neuropeptides, and their relation to behavior and brain activity. Steroid and neuropeptide mechanisms had and have virtually unlimited spatial and temporal flexibility to unfold their effects both during the evolution of mankind and in the developmental course of every human being. Recent research has highlighted the broad-scale impact of neuropeptides, such as OT and vasopressin (VAT), and of steroid hormones, in particular testosterone (T) and estradiol (E) on complex social behaviors (for a review, see Bos et al., 2012). An important advantage of this line of research is that pharmacological drug administration in comparison to placebo administration allows for experimental manipulation of effects within or between individuals. For example, administration of OT increased performance of ASD patients at a well-established ToM task (the reading the mind in the eyes task; RMET), and it led to higher sensitivity for positive emotion expressions in a morphing task (Domes et al., 2007; Marsh et al., 2010). Furthermore, a positive effect of OT on empathy and ToM has been reported by several studies (Bartz et al., 2010; Domes et al., 2007; Guastella et al., 2010) while T seems to have the opposite effect by reducing social cognition (Olsson et al., 2016; van Honk & J.L.G. Schutter, 2007). In a similar vein, OT enhances ratings of trustworthiness and trust behavior as well as positive social interactions whereas T has a negative effect on trust and increases aggression (Andari et al., 2010; Bos et al., 2010; Kosfeld et al., 2005; Mikolajczak et al., 2010). Seminal studies combining psychopharmacological drug administration and fMRI found that hormones also have an impact at a neuronal level. As an example, a single administration of T in young women has been found to reduce connectivity in regions of the “social brain network” during the RMET (Bos, 2016). In sum, hormones seem to play a causal role in human social cognition and social interaction. Considering their effects in research on social cognition could be particularly informative in studies that work with adolescent samples because this age group is marked by hormonal changes during puberty. Ideally, longitudinal studies should assess the developmental course of sex differences during and beyond adolescence (Goddings et al., 2019).

CONCLUDING REMARKS

The findings of this dissertation highlight the multifaceted influences of gaze cues on human attention and interpersonal communication. Social interactions are highly complex scenarios and gaze cues play a decisive role on several levels of human cognition. The eyes are a magnet for our attention and they help us form an impression about other people, communicate with them and understand their feelings and intentions. Importantly, as highlighted throughout this dissertation, gaze cues do not exert their influences in isolation. Instead, they are shaped by and integrated with context information and additional social cues, suggesting an intricate interplay of multiple-level mechanisms during social encounters. These findings are not only important work towards a thorough understanding of human social attention and cognition, they also highlight the importance of a holistic and more inclusive assessment of gaze cues and social understanding. Human interactions in the real world are complex and multidimensional scenarios. Accordingly, our brain is built to integrate all social and non-social aspects of a situation in order to guide adaptive behavior. Research that aims at generating findings that generalize outside of the laboratory needs to account for this complexity by adopting naturalistic paradigms and by explicitly considering context variables on multiple sensory channels.

In the present set of studies, we took different approaches to investigate the impact of gaze behavior on social attention and cognition. If you have read my whole dissertation, you will now know how powerful and far-reaching the effects of gaze are and how our brain is wired to attend and respond to gaze cues in the environment. This reflexive disposition is so strong that, sometimes, it can be quite relaxing to escape the gaze of others by hiding alone in your room to read a good book. However, if this book happens to be “I Am Legend” by Richard Matheson, you might be happy that other people are just a few meters away. You only need to open your window and gaze outside.

REFERENCES

Aboulafia-Brakha, T., Christe, B., Martory, M.-D., & Annoni, J.-M. (2011). Theory of mind tasks and executive functions: A systematic review of group studies in neurology: Theory of mind and executive functions. *Journal of Neuropsychology*, *5*(1), 39–55. https://doi.org/10.1348/174866410X533660

Abrams, R. A., & Christ, S. E. (2005). The onset of receding motion captures attention: Comment on Franconeri and Simons (2003). *Perception & Psychophysics*, *67*(2), 219–223. https://doi.org/10.3758/BF03206486

Adam, E. K., Chyu, L., Hoyt, L. T., Doane, L. D., Boisjoly, J., Duncan, G. J., Chase-Lansdale, P. L., & McDade, T. W. (2011). Adverse adolescent relationship histories and young adult health: Cumulative effects of loneliness, low parental support, relationship instability, intimate partner violence, and loss. *Journal of Adolescent Health*, *49*(3), 278–286.

Adams, R. B., & Kleck, R. E. (2003). Perceived Gaze Direction and the Processing of Facial Displays of Emotion. *Psychological Science*, *14*(6), 4.

Adams, R. B., & Kleck, R. E. (2005). Effects of Direct and Averted Gaze on the Perception of Facially Communicated Emotion. *Emotion*, *5*(1), 3–11.

Alkire, D., Levitas, D., Warnell, K. R., & Redcay, E. (2018). Social interaction recruits mentalizing and reward systems in middle childhood. *Human Brain Mapping*, *39*, 3928–3942. https://doi.org/10.1002/hbm.24221

Allemand, M., Steiger, A. E., & Fend, H. A. (2015). Empathy Development in Adolescence Predicts Social Competencies in Adulthood: Adolescent Empathy and Adult Outcomes. *Journal of Personality*, *83*(2), 229–241. https://doi.org/10.1111/jopy.12098

Almenberg, J., Dreber, A., Apicella, C. L., & Rand, D. (2011). Third party reward and punishment: Group size, efficiency and public goods. In *Psychology of Punishment*. Nova Publishing.

Altschuler, M., Sideridis, G., Kala, S., Warshawsky, M., Gilbert, R., Carroll, D., Burger-Caplan, R., & Faja, S. (2018). Measuring Individual Differences in Cognitive, Affective, and Spontaneous Theory of Mind Among School-Aged Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(11), 3945–3957. https://doi.org/10.1007/s10803-018-3663-1

Andari, E., Duhamel, J.-R., Zalla, T., Herbrecht, E., Leboyer, M., & Sirigu, A. (2010). Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proceedings of the National Academy of Sciences*, *107*(9), 4389–4394. https://doi.org/10.1073/pnas.0910249107

Anstis, S. (1974). A chart demonstrating variations in acuity with retinal position. *Vision research*, *14*(7), 589–592.

Anstis, S. (1998). Picturing peripheral acuity. *Perception*, *27*(7), 817–825.

Arango Tobón, O. E., Olivera-La Rosa, A., Restrepo Tamayo, V., & Puerta Lopera, I. C. (2017). Empathic skills and theory of mind in female adolescents with conduct disorder. *Revista Brasileira de Psiquiatria*, *40*(1), 78–82. https://doi.org/10.1590/1516-4446-2016-2092

Argyle, M., & Dean, J. (1965). Eye-Contact, Distance and Affiliation. *Sociometry*, *28*(3), 289. https://doi.org/10.2307/2786027

Argyle, M., & Ingham, R. (1972). Gaze, Mutual Gaze, and Proximity. *Semiotica*, *6*(1). https://doi.org/10.1515/semi.1972.6.1.32

Arizpe, J., Walsh, V., Yovel, G., & Baker, C. I. (2017). The categories, frequencies, and stability of idiosyncratic eye-movement patterns to faces. *Vision Research*, *141*, 191–203. https://doi.org/10.1016/j.visres.2016.10.013

Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology*, *164*, 225–238. https://doi.org/10.1016/j.jecp.2017.04.007

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Baksh, R. A., Abrahams, S., Auyeung, B., & MacPherson, S. E. (2018). The Edinburgh Social Cognition Test (ESCoT): Examining the effects of age on a new measure of theory of mind and social norm understanding. *PLOS ONE*, *13*(4), e0195818. https://doi.org/10.1371/journal.pone.0195818

Baltazar, M., Hazem, N., Vilarem, E., Beaucousin, V., Picq, J.-L., & Conty, L. (2014). Eye contact elicits bodily self-awareness in human adults. *Cognition*, *133*(1), 120–127. https://doi.org/10.1016/j.cognition.2014.06.009

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer Relations and the Understanding of Faux Pas: Longitudinal Evidence for Bidirectional Associations: Peer Relations Faux Pas. *Child Development*, *82*(6), 1887–1905. https://doi.org/10.1111/j.1467-8624.2011.01669.x

Barasch, A., Small, D., Levine, E. E., & Berman, J. (2014). Selfish Or Selfless? On the Signal Value of Emotion in Altruistic Behavior. *Journal of Personality and Social Psychology*, *107*(3), 393–413.

Bard, K. A., Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., Costall, A., & Matsuzawa, T. (2005). Group Differences in the Mutual Gaze of Chimpanzees (Pan Troglodytes). *Developmental Psychology*, *51*(4), 616–624.

Baron-Cohen, S. (1997). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *The maladapted mind: Classic readings in evolutionary psychopathology*, 207–239.

Baron-Cohen, S. (2000). Theory of mind and autism: A review. *International review of research in mental retardation*, *23*, 169–184.

Baron-Cohen, S., & Cross, P. (1992a). Reading the eyes: Evidence for the role of perception in the development of a theory of mind. *Mind & Language*, *7*(1-2), 172–186.

Baron-Cohen, S., & Cross, P. (1992b). Reading the Eyes: Evidence for the Role of Perception in the Development of a Theory of Mind. *Mind & Language*, *7*(1–2), 172–186. https://doi.org/10.1111/j.1468-0017.1992.tb00203.x

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind” ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning Disorders*, *5*(1), 47–78.

Barone, P., Corradi, G., & Gomila, A. (2019). Infants’ performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior and Development*, *57*, 101350. https://doi.org/10.1016/j.infbeh.2019.101350

Barthelme, S. (2019). *eyelinker: Import ASC Files from EyeLink Eye Trackers.* (R package version 0.2.0) [Computer software]. https://CRAN.R-project.org/package=eyelinker

Bartz, J. A., Zaki, J., Bolger, N., Hollander, E., Ludwig, N. N., Kolevzon, A., & Ochsner, K. N. (2010). Oxytocin Selectively Improves Empathic Accuracy. *Psychological Science*, *21*(10), 1426–1428. https://doi.org/10.1177/0956797610383439

Bavelas, J. B., & Gerwing, J. (2011). The Listener as Addressee in Face-to-Face Dialogue. *International Journal of Listening*, *25*(3), 178–198. https://doi.org/10.1080/10904018.2010.508675

Bayle, D. J., Schoendorff, B., Hénaff, M.-A., & Krolak-Salmon, P. (2011). Emotional Facial Expression Detection in the Peripheral Visual Field. *PLoS ONE*, *6*(6), e21584. https://doi.org/10.1371/journal.pone.0021584

Begeer, S., Gevers, C., Clifford, P., Verhoeve, M., Kat, K., Hoddenbach, E., & Boer, F. (2011). Theory of Mind Training in Children with Autism: A Randomized Controlled Trial. *Journal of Autism and Developmental Disorders*, *41*(8), 997–1006. https://doi.org/10.1007/s10803-010-1121-9

Behrens, F., Snijdewint, J. A., Moulder, R. G., Prochazkova, E., Sjak-Shie, E. E., Boker, S. M., & Kret, M. E. (2020). Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-76539-8

Bekkering, E., & Shim, J. P. (2006). Trust in videoconferencing. *Communications of the ACM*, *49*(7), 103–107. https://doi.org/10.1145/1139922.1139925

Berecz, H., Tényi, T., & Herold, R. (2016). Theory of Mind in Depressive Disorders: A Review of the Literature. *Psychopathology*, *49*(3), 125–134. https://doi.org/10.1159/000446707

Berg, J., Dickhaut, J., & McCabe, K. (1995a). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122–142.

Berg, J., Dickhaut, J., & McCabe, K. (1995b). Trust, reciprocity, and social history. In *Games and Economic Behavior* (Bd. 10, Nummer 1, S. 122–142). https://doi.org/10.1006/game.1995.1027

Berman, J. Z., & Silver, I. (2022). Prosocial behavior and reputation: When does doing good lead to looking good? *Current Opinion in Psychology*, *43*, 102–107. https://doi.org/10.1016/j.copsyc.2021.06.021

Bernhardt, B. C., Klimecki, O. M., Leiberg, S., & Singer, T. (2014). Structural Covariance Networks of the Dorsal Anterior Insula Predict Females' Individual Differences in Empathic Responding. *Cerebral Cortex*, *24*(8), 2189–2198. https://doi.org/10.1093/cercor/bht072

Bernhardt, B. C., Valk, S. L., Silani, G., Bird, G., Frith, U., & Singer, T. (2014). Selective Disruption of Sociocognitive Structural Brain Networks in Autism and Alexithymia. *Cerebral Cortex*, *24*(12), 3258–3267. https://doi.org/10.1093/cercor/bht182

Bianco, F., & Lecce, S. (2016). Translating child development research into practice: Can teachers foster children's theory of mind in primary school? *British Journal of Educational Psychology*, *86*(4), 592–605. https://doi.org/10.1111/bjep.12125

Bianco, F., Lecce, S., & Banerjee, R. (2016). Conversations about mental states and theory of mind development in middle childhood: A training study. *Journal of Experimental Child Psychology*, *149*, 41–61.

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

Bio, B. J., Webb, T. W., & Graziano, M. S. A. (2018). Projecting one's own spatial bias onto others during a theory-of-mind task. *Proceedings of the National Academy of Sciences*, *115*(7), E1684–E1689. https://doi.org/10.1073/pnas.1718493115

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

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). Social Attention and Real-World Scenes: The Roles of Action, Competition and Social Content. *Quarterly Journal of*

Experimental Psychology, *61*(7), 986–998. https://doi.org/10.1080/17470210701410375

Blair, R. J. R. (2018). Traits of empathy and anger: Implications for psychopathy and other disorders associated with aggression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1744), 20170155. https://doi.org/10.1098/rstb.2017.0155

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267–277. https://doi.org/10.1038/nrn2353

Blakemore, S.-J., den Ouden, H., Choudhury, S., & Frith, C. (2007). Adolescent development of the neural circuitry for thinking about intentions. *Social Cognitive and Affective Neuroscience*, *2*(2), 130–139. https://doi.org/10.1093/scan/nsm009

Blakemore, S.-J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., Segebarth, C., & Decety, J. (2001). How the brain perceives causality: An event-related fMRI study: *Neuroreport*, *12*(17), 3741–3746. https://doi.org/10.1097/00001756-200112040-00027

Böckler, A., Herrmann, L., Trautwein, F.-M., Holmes, T., & Singer, T. (2017). Know Thy Selves: Learning to Understand Oneself Increases the Ability to Understand Others. *Journal of Cognitive Enhancement*, *1*(2), 197–209. https://doi.org/10.1007/s41465-017-0023-6

Böckler, A., Hömke, P., & Sebanz, N. (2014). Invisible Man: Exclusion From Shared Attention Affects Gaze Behavior and Self-Reports. *Social Psychological and Personality Science*, *5*(2), 140–148. https://doi.org/10.1177/1948550613488951

Böckler, A., Tusche, A., & Singer, T. (2016). The Structure of Human Prosociality: Differentiating Altruistically Motivated, Norm Motivated, Strategically Motivated, and Self-Reported Prosocial Behavior. *Social Psychological and Personality Science*, *7*(6), 530–541. https://doi.org/10.1177/1948550616639650

Böckler, A., van der Wel, R. P. R. D., & Welsh, T. N. (2014). Catching Eyes: Effects of Social and Nonsocial Cues on Attention Capture. *Psychological Science*, *25*(3), 720–727. https://doi.org/10.1177/0956797613516147

Böckler, A., van der Wel, R. P. R. D., & Welsh, T. N. (2015). Eyes only? Perceiving eye contact is neither sufficient nor necessary for attentional capture by face direction. *Acta Psychologica*, *160*, 134–140. https://doi.org/10.1016/j.actpsy.2015.07.009

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia Research*, *109*(1–3), 1–9. https://doi.org/10.1016/j.schres.2008.12.020

Bos, P. A. (2016). *Testosterone reduces functional connectivity during the 'Reading the Mind in the Eyes' Test*. 8.

Bos, P. A., Panksepp, J., Bluthé, R.-M., & Honk, J. van. (2012). Acute effects of steroid hormones and neuropeptides on human social–emotional behavior: A review of single administration studies. *Frontiers in Neuroendocrinology*, *33*(1), 17–35. https://doi.org/10.1016/j.yfrne.2011.01.002

Bos, P. A., Terburg, D., & van Honk, J. (2010). Testosterone decreases trust in socially naive humans. *Proceedings of the National Academy of Sciences*, *107*(22), 9991–9995. https://doi.org/10.1073/pnas.0911700107

Bosacki, S., & Wilde Astington, J. (2001). Theory of Mind in Preadolescence: Relations Between Social Understanding and Social Competence. *Social Development*, *8*(2), 237–255. https://doi.org/10.1111/1467-9507.00093

Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of Vision*, *16*(2), 10. https://doi.org/10.1167/16.2.10

Bourgou, S., Halayem, S., Amado, I., Triki, R., Bourdel, M. C., Franck, N., Krebs, M. O., Tabbane, K., & Bouden, A. (2016). Theory of mind in adolescents with early-onset schizophrenia: Correlations with clinical assessment and executive functions. *Acta Neuropsychiatrica*, *28*(4), 232–238. https://doi.org/10.1017/neu.2016.3

Boyd, R., & Richerson, P. J. (2009). *Culture and the evolution of human cooperation*. 8.

Boyer, T. W., & Wang, M. (2018). Direct gaze, eye movements, and covert and overt social attention processes. *Attention, Perception, & Psychophysics*, *80*(7), 1654–1659. https://doi.org/10.3758/s13414-018-1590-z

Bradford, E. E. F., Gomez, J.-C., & Jentzsch, I. (2019). Exploring the role of self/other perspective-shifting in theory of mind with behavioural and EEG measures. *Social Neuroscience*, *14*(5), 530–544. https://doi.org/10.1080/17470919.2018.1514324

Breil, C., & Böckler, A. (2020). The Lens Shapes the View: On Task Dependency in ToM Research. *Current Behavioral Neuroscience Reports*, *7*(2), 41–50. https://doi.org/10.1007/s40473-020-00205-6

Breil, C., & Böckler, A. (2021). Look away to listen: The interplay of emotional context and eye contact in video conversations. *Visual Cognition*, *29*(5), 277–287. https://doi.org/10.1080/13506285.2021.1908470

Breil, C., Huestegge, L., & Böckler, A. (2021). From eye to arrow: Attention capture by direct gaze requires more than just the eyes. *Attention, Perception & Psychophysics*.

Breil, C., Kanske, P., Pittig, R., & Böckler, A. (2021). A revised instrument for the assessment of empathy and Theory of Mind in adolescents: Introducing the EmpaToM-Y. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01589-3

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

Bressler, D. W., Fortenbaugh, F. C., Robertson, L. C., & Silver, M. A. (2013). Visual spatial attention enhances the amplitude of positive and negative fMRI responses to visual stimulation in an eccentricity-dependent manner. *Vision Research*, *85*, 104–112. https://doi.org/10.1016/j.visres.2013.03.009

Brown, M. M., Thibodeau, R. B., Pierucci, J. M., & Gilpin, A. T. (2017). Supporting the development of empathy: The role of theory of mind and fantasy orientation. *Social Development*, *26*(4), 951–964. https://doi.org/10.1111/sode.12232

Brunet, E., Sarfati, Y., Hardy-Baylé, M.-C., & Decety, J. (2000). A PET Investigation of the Attribution of Intentions with a Nonverbal Task. *NeuroImage*, *11*(2), 157–166. https://doi.org/10.1006/nimg.1999.0525

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1).

Burnat, K. (2015). Are Visual Peripheries Forever Young? *Neural Plasticity*, *2015*, 1–13. https://doi.org/10.1155/2015/307929

Burnel, M., Perrone-Bertolotti, M., Reboul, A., Baciu, M., & Durrleman, S. (2018). Reducing the language content in ToM tests: A developmental scale. *Developmental Psychology*, *54*(2), 293–307. https://doi.org/10.1037/dev0000429

Burns, J. A., & Kintz, B. L. (1976). Eye contact while lying during an interview. *Bulletin of the Psychonomic Society*, *7*(1), 87–89. https://doi.org/10.3758/BF03337131

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

Bzdok, D., & Dunbar, R. I. (2020). The neurobiology of social distance. *Trends in cognitive sciences*, *24*(9), 717–733.

Cacioppo, J. T., & Cacioppo, S. (2014). Social Relationships and Health: The Toxic Effects of Perceived Social Isolation. *Social and Personality Psychology Compass*, 15.

Cadinu, M. R., & Kiesner, J. (2000). Children’s development of a theory of mind. *European Journal of Psychology of Education*, *15*(2), 93–111. https://doi.org/10.1007/BF03173169

Calvo, M. G., Fernández-Martín, A., & Nummenmaa, L. (2014). Facial expression recognition in peripheral versus central vision: Role of the eyes and the mouth. *Psychological Research*, *78*(2), 180–195. https://doi.org/10.1007/s00426-013-0492-x

Calvo, M. G., & Nummenmaa, L. (2008). Detection of Emotional Faces: Salient Physical Features Guide Effective Visual Search. *Journal of Experimental Psychology: General*, *137*(3), 471–494.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Camerer, C. F., Ho, T.-H., & Chong, J. K. (2004). Behavioural game theory: Thinking, learning and teaching. In *Advances in understanding strategic behaviour* (S. 120–180). Springer.

Cañigueral, R., & Hamilton, A. F. de C. (2019). The Role of Eye Gaze During Natural Social Interactions in Typical and Autistic People. *Frontiers in Psychology*, *10*, 560. https://doi.org/10.3389/fpsyg.2019.00560

Canty, A. L., Neumann, D. L., & Shum, D. H. K. (2017). Using virtual reality to assess theory of mind subprocesses and error types in early and chronic schizophrenia. *Schizophrenia Research: Cognition*, *10*, 15–19. https://doi.org/10.1016/j.scog.2017.09.001

Carrasco, M., McLean, T. L., Katz, S. M., & Frieder, K. S. (1998). Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation and spatial frequency. *Vision Research*, *38*(3), 347–374. https://doi.org/10.1016/S0042-6989(97)00152-1

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313–2335. https://doi.org/10.1002/hbm.20671

Caspi, A., & Shiner, R. L. (2006). Personality Development. In N. Eisenberg (Hrsg.), *Handbook of Child Psychology, Vol. 3. Social, Emotional, and Personality Development* (6. Aufl., S. 300–365). Wiley.

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

Cavallini, E., Bianco, F., Bottiroli, S., Rosi, A., Vecchi, T., & Lecce, S. (2015). Training for generalization in Theory of Mind: A study with older adults. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01123

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15*(4), 481–498. https://doi.org/10.1016/S0885-2014(01)00037-5

Chen, M. (2002). Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconference. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *4*(1), 49–56.

Chen, Y.-C., & Yeh, S.-L. (2012). Look into my eyes and I will see you: Unconscious processing of human gaze. *Consciousness and Cognition*, *21*(4), 1703–1710. https://doi.org/10.1016/j.concog.2012.10.001

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual Differences in Frequency and Saliency of Speech-Accompanying gestures: The role of cognitive abilities and empathy. *Journal of Exprimental Psychology: General*, *143*(2), 694–709.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4

Conty, L., George, N., & Hietanen, J. K. (2016). Watching Eyes effects: When others meet the self. *Consciousness and Cognition*, *45*, 184–197. https://doi.org/10.1016/j.concog.2016.08.016

Conty, L., Russo, M., Loehr, V., Hugueville, L., Barbu, S., Huguet, P., Tijus, C., & George, N. (2010). The mere perception of eye contact increases arousal during a word-spelling task. *Social Neuroscience*, *5*(2), 171–186. https://doi.org/10.1080/17470910903227507

Conty, L., Tijus, C., Hugueville, L., Coelho, E., & George, N. (2006). Searching for asymmetries in the detection of gaze contact versus averted gaze under different head views: A behavioural study. *Spatial Vision*, *19*(6), 529–545. https://doi.org/10.1163/156856806779194026

Conway, J. R., & Bird, G. (2018). Conceptualizing degrees of theory of mind. *Proceedings of the National Academy of Sciences*, *115*(7), 1408–1410. https://doi.org/10.1073/pnas.1722396115

Cook, M. (1977). Gaze and Mutual Gaze in Social Encounters: How long—And when—We look others „in the eye“ is one of the main signals in nonverbal communication. *American Scientist*, *65*, 328–333.

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

Cowan, D. G., Vanman, E. J., & Nielsen, M. (2014). Motivated empathy: The mechanics of the empathic gaze. *Cognition and Emotion*, *28*(8), 1522–1530. https://doi.org/10.1080/02699931.2014.890563

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(48), 17321–17325. https://doi.org/10.1073/pnas.1424572112

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Cui, M., Zhu, M., Lu, X., & Zhu, L. (2019). Implicit Perceptions of Closeness From the Direct Eye Gaze. *Frontiers in Psychology*, *9*, 2673. https://doi.org/10.3389/fpsyg.2018.02673

Dadds, M. R., Allen, J. L., McGregor, K., Woolgar, M., Viding, E., & Scott, S. (2014). Callous-unemotional traits in children and mechanisms of impaired eye contact during expressions of love: A treatment target? *Journal of Child Psychology and Psychiatry*, *55*(7), 771–780.

Daft, R. L., & Lengel, R. H. (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, *32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

Daum, M. M., & Gredebäck, G. (2011). Spatial cueing by referential human gestures, arrows and mechanical devices. *International Journal of Mind, Brain and Cognition*, *2*(1–2), 113–126.

Davis, H. (1980). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–132.

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441. https://doi.org/10.1016/j.tics.2006.08.008

Decety, J., & Michalska, K. J. (2010). Neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood: Changes in circuits underlying empathy and sympathy. *Developmental Science*, *13*(6), 886–899. https://doi.org/10.1111/j.1467-7687.2009.00940.x

Deschrijver, E., Bardi, L., Wiersema, J. R., & Brass, M. (2016). Behavioral measures of implicit theory of mind in adults with high functioning autism. *Cognitive Neuroscience*, *7*(1–4), 192–202. https://doi.org/10.1080/17588928.2015.1085375

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of Mind, Gender, and Social Experiences in Middle Childhood. *Child Development*, *84*(3), 989–1003. https://doi.org/10.1111/cdev.12017

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious Facial Reactions to Emotional Facial Expressions. *Psychological Science*, *11*(1), 86–89. https://doi.org/10.1111/1467-9280.00221

Dockett, S., & Perry, B. (1996). Young children’s construction of knowledge. *Australasian Journal of Early Childhood*, *21*(4), 6–11.

Doi, H., Ueda, K., & Shinohara, K. (2009). Neural correlates of the stare-in-the-crowd effect. *Neuropsychologia*, *47*(4), 1053–1060. https://doi.org/10.1016/j.neuropsychologia.2008.11.004

Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin Improves “Mind-Reading” in Humans. *Biological Psychiatry*, *61*(6), 731–733. https://doi.org/10.1016/j.biopsych.2006.07.015

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

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, *111*(15), E1454–E1462. https://doi.org/10.1073/pnas.1322355111

Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: A Movie for the Assessment of Social Cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623–636. https://doi.org/10.1007/s10803-006-0107-0

Eisenberg, N., Morris, A. S., McDaniel, B., & Spinrad, T. L. (2009). Moral cognitions and prosocial responding in adolescence. In R. M. Lerner & L. Steinberg (Hrsg.), *Handbook of adolescent psychology* (S. 229–265). John Wiley & Sons Inc.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604. https://doi.org/10.1016/S0149-7634(00)00025-7

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Erlandsson, A., Wingren, M., & Andersson, P. A. (2020). Type and amount of help as predictors for impression of helpers. *PLoS ONE*, *15*(12 December), 1–23. https://doi.org/10.1371/journal.pone.0243808

Exline, R., Gray, D., & Schuette, D. (1965). Visual behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology*, *1*(3), 201–209. https://doi.org/10.1037/h0021865

Farroni, T., Massaccesi, S., Pividori, Donatella, & Johnsonm Mark. (2004). Gaze Following in Newborns. *Infancy*, *5*(1), 39–60.

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

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

Fawcett, C., Wesevich, V., & Gredebäck, G. (2016). Pupillary Contagion in Infancy: Evidence for Spontaneous Transfer of Arousal. *Psychological Science, 27*(7), 997–1003. https://doi.org/10.1177/0956797616643924

FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. *NeuroImage*, *105*, 347–356. https://doi.org/10.1016/j.neuroimage.2014.10.043

Ferguson, D. J. (2018). _*eyetrackingR*_ (R package version 0.1.8) [Computer software]. http://www.eyetracking-R.com

Fitzpatrick, P., Romero, V., Amaral, J. L., Duncan, A., Barnard, H., Richardson, M. J., & Schmidt, R. C. (2017). Social Motor Synchronization: Insights for Understanding Social Behavior in Autism. *Journal of Autism and Developmental Disorders*, *47*(7), 2092–2107. https://doi.org/10.1007/s10803-017-3124-2

Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind? *Journal of Experimental Child Psychology*, *162*, 209–224. https://doi.org/10.1016/j.jecp.2017.05.005

Forbes, P. A. G., Pan, X., & de C. Hamilton, A. F. (2016). Reduced Mimicry to Virtual Reality Avatars in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *46*(12), 3788–3797. https://doi.org/10.1007/s10803-016-2930-2

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

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

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional Effects of Counterpredictive Gaze and Arrow Cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319–329. https://doi.org/10.1037/0096-1523.30.2.319

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

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

Frith, C. D., & Corcoran, R. (1996). Exploring ‘theory of mind’ in people with schizophrenia. *Psychological Medicine*, *26*(3), 521–530. https://doi.org/10.1017/S0033291700035601

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, *8*(4), 465–486. https://doi.org/10.1007/s11097-009-9136-4

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the Intentional Stance in a Competitive Game. *NeuroImage*, *16*(3), 814–821. https://doi.org/10.1006/nimg.2002.1117

Gehrer, N. A., Duchowski, A. T., Jusyte, A., & Schönenberg, M. (2020). Eye contact during live social interaction in incarcerated psychopathic offenders. *Personality Disorders: Theory, Research, and Treatment*, *11*(6), 431–439.

Gillespie-Lynch, K., Sepeta, L., Wang, Y., Marshall, S., Gomez, L., Sigman, M., & Hutman, T. (2012). Early Childhood Predictors of the Social Competence of Adults with Autism. *Journal of Autism and Developmental Disorders*, *42*(2), 161–174. https://doi.org/10.1007/s10803-011-1222-0

Goddings, A.-L., Beltz, A., Peper, J. S., Crone, E. A., & Braams, B. R. (2019). Understanding the Role of Puberty in Structural and Functional Development of the Adolescent Brain. *Journal of Adolescence*, *29*(1), 32–53.

Goffman, E. (2008). *Behavior in public places*. Simon and Schuster.

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology*, *18*(02). https://doi.org/10.1017/S0954579406060305

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants. *Pediatrics*, *56*(4), 544–549.

Greene, D. J., Colich, N., Iacoboni, M., Zaidel, E., Bookheimer, S. Y., & Dapretto, M. (2011). Atypical neural networks for social orienting in autism spectrum disorders. *NeuroImage*, *56*(1), 354–362. https://doi.org/10.1016/j.neuroimage.2011.02.031

Gruber, N., Müri, R. M., Mosimann, U. P., Bieri, R., Aeschimann, A., Zito, G. A., Urwyler, P., Nyffeler, T., & Nef, T. (2014). Effects of age and eccentricity on visual target detection. *Frontiers in Aging Neuroscience*, *5*, 101. https://doi.org/10.3389/fnagi.2013.00101

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

Hamilton, A. F. de C. (2016). Gazing at me: The importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150080. https://doi.org/10.1098/rstb.2015.0080

Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of theoretical biology*, *7*(1), 17–52.

Hanley, M., Riby, D. M., McCormack, T., Carty, C., Coyle, L., Crozier, N., Robinson, J., & McPhillips, M. (2014). Attention during social interaction in children with autism: Comparison to specific language impairment, typical development, and links to social cognition. *Research in Autism Spectrum Disorders*, *8*(7), 908–924. https://doi.org/10.1016/j.rasd.2014.03.020

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Harrison, C., Binetti, N., Coutrot, A., Johnston, A., & Mareschal, I. (2018). Personality Traits Do Not Predict How We Look at Faces. *Perception*, *47*(9), 976–984. https://doi.org/10.1177/0301006618788754

Hartikainen, K. M., Ogawa, K. H., & Knight, R. T. (2000). Transient interference of right hemispheric function due to automatic emotional processing. *Neuropsychologia*, *38*(12), 1576–1580. https://doi.org/10.1016/S0028-3932(00)00072-5

Hasni, A. A., Adamson, L. B., Williamson, R. A., & Robins, D. L. (2017). Adding sound to theory of mind: Comparing children's development of mental-state understanding in the auditory and visual realms. *Journal of Experimental Child Psychology, 164*, 239–249. https://doi.org/10.1016/j.jecp.2017.07.009

Haxby, J. V., & Gobbini, M. I. (2011). *Distributed Neural Systems for Face Perception*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.013.0006

Hayes, A. (2017). *Introduction to Mediation, Moderation and Conditional Process Analysis: A Regression based approach*. Guilford Publications.

Hays, R., & Revicki, D. A. (2005). Reliability and validity (including responsiveness). In P. Fayers & R. Hays (Hrsg.), *Assessing quality of life in clinical trials: Methods and practice* (2. Aufl.). Oxford University Press.

Hayward, D. A., & Ristic, J. (2015). Exposing the cuing task: The case of gaze and arrow cues. *Attention, Perception, & Psychophysics*, *77*(4), 1088–1104. https://doi.org/10.3758/s13414-015-0877-6

Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*, *35*(3), 454–462. https://doi.org/10.1111/bjdp.12186

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging*, *28*(3), 826–839. https://doi.org/10.1037/a0030677

Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *Journal of Vision*, *10*(10), 21–21. https://doi.org/10.1167/10.10.21

Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*, *27*(5), 856–881. https://doi.org/10.3758/s13423-020-01715-w

Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *71*(3), 226–242. https://doi.org/10.1037/cep0000113

Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2018). Eye contact takes two – autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*, *9*(2), jep.062917. https://doi.org/10.5127/jep.062917

Heyes, C. (2003). Four routes of cognitive evolution. *Psychological Review*, *110*(4), 713–727. https://doi.org/10.1037/0033-295X.110.4.713

Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, *9*(2), 131–143. https://doi.org/10.1177/1745691613518076

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, *21*(1), 101–114. https://doi.org/10.1017/S0140525X98000703

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190), 1243091–1243091. https://doi.org/10.1126/science.1243091

Hietanen, J. K. (2018). Affective Eye Contact: An Integrative Review. *Frontiers in Psychology*, *9*, 1587. https://doi.org/10.3389/fpsyg.2018.01587

Hietanen, J. K., Leppänen, J. M., Nummenmaa, L., & Astikainen, P. (2008). Visuospatial attention shifts by gaze and arrow cues: An ERP study. *Brain research*, *1215*, 123–136.

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

Hietanen, J. O., & Hietanen, J. K. (2017). Genuine eye contact elicits self-referential processing. *Consciousness and Cognition*, *51*, 100–115.

Hietanen, J., & Yrttimaa, K. (2005). Where a person with a squint is actually looking: Gaze-cued orienting in crossed eyes. *Visual Cognition*, *12*(1), 117–126. https://doi.org/10.1080/13506280444000184

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). First fixations in face processing: The more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica*, *142*(2), 211–219.

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *55*, 498–509. https://doi.org/10.1016/j.neubiorev.2015.06.002

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and Listening with the Eyes: Gaze Signaling during Dyadic Interactions. *PLOS ONE*, *10*(8), e0136905.

Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., Curtiss, J., & Harris, P. L. (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition*, *150*, 200–212. https://doi.org/10.1016/j.cognition.2016.01.006

Huestegge, L., & Böckler, A. (2016). Out of the corner of the driver's eye: Peripheral processing of hazards in static traffic scenes. *Journal of Vision*, *16*(2), 11. https://doi.org/10.1167/16.2.11

Hughes, C., & Leekam, S. (2004). What are the Links Between Theory of Mind and Social Relations? Review, Reflections and New Directions for Studies of Typical and Atypical Development. *Social Development*, *13*(4), 590–619. https://doi.org/10.1111/j.1467-9507.2004.00285.x

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric Evaluation of the Theory of Mind Inventory (ToMI): A Study of Typically Developing Children and Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *42*(3), 327–341. https://doi.org/10.1007/s10803-011-1244-7

Inquisit 5 [Computer software]. (2016).

Israelashvili, J., Sauter, D., & Fischer, A. (2019). How Well Can We Assess Our Ability to Understand Others' Feelings? Beliefs About Taking Others' Perspectives and Actual

Understanding of Others' Emotions. *Frontiers in Psychology*, *10*, 2475. https://doi.org/10.3389/fpsyg.2019.02475

Itier, R. J., & Neath-Tavares, K. N. (2017). Effects of task demands on the early neural processing of fearful and happy facial expressions. *Brain Research*, *1663*, 38–50. https://doi.org/10.1016/j.brainres.2017.03.013

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*(1), 2.

Jebara, N., Pins, D., Despretz, P., & Boucart, M. (2009). Face or building superiority in peripheral vision reversed by task requirements. *Advances in Cognitive Psychology*, *5*(1), 42–53. https://doi.org/10.2478/v10053-008-0065-5

Jenkins, A. C., Karashchuk, P., Zhu, L., & Hsu, M. (2018). Predicting human behavior toward members of different social groups. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(39), 9696–9701. https://doi.org/10.1073/pnas.1719452115

Jiang, J., Borowiak, K., Tudge, L., Otto, C., & von Kriegstein, K. (2016). Neural mechanisms of eye contact when listening to another person talking. *Social Cognitive and Affective Neuroscience*, nsw127. https://doi.org/10.1093/scan/nsw127

Johansson Nolaker, E., Murray, K., Happé, F., & Charlton, R. A. (2018). Cognitive and affective associations with an ecologically valid test of theory of mind across the lifespan. *Neuropsychology*, *32*(6), 754–763. https://doi.org/10.1037/neu0000464

Johnson, S. A., Filliter, J. H., & Murphy, R. R. (2009). Discrepancies between self- and parent-perceptions of Autistic traits and empathy in high functioning children and adolescents on the Autism spectrum. *Journal of Autism and Developmental Disorders*, *39*(12), 1706–1714. https://doi.org/10.1007/s10803-009-0809-1

Kaisler, R. E., & Leder, H. (2016a). Trusting the Looks of Others: Gaze Effects of Faces in Social Settings. *Perception*, *45*(8), 875–892. https://doi.org/10.1177/0301006616643678

Kaisler, R. E., & Leder, H. (2016b). Trusting the Looks of Others: Gaze Effects of Faces in Social Settings. *Perception*, *45*(8), 875–892. https://doi.org/10.1177/0301006616643678

Kaland, N., Callesen, K., Møller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of Children and Adolescents with Asperger Syndrome or High-functioning Autism on Advanced Theory of Mind Tasks. *Journal of Autism and Developmental Disorders*, *38*(6), 1112–1123. https://doi.org/10.1007/s10803-007-0496-8

Kamawar, D., & Olson, D. R. (2011). Thinking about representations: The case of opaque contexts. *Journal of Experimental Child Psychology*, *108*(4), 734–746. https://doi.org/10.1016/j.jecp.2010.10.005

Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual Reality Social Cognition Training for Young Adults with High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *43*(1), 34–44. https://doi.org/10.1007/s10803-012-1544-6

Kanske, P., Böckler, A., Trautwein, F. M., & Singer, T. (2015a). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and Theory of Mind. *NeuroImage*, *122*, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Kanske, P., Böckler, A., Trautwein, F.-M., Parianen Lesemann, F. H., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective Neuroscience*, *11*(9), 1383–1392. https://doi.org/10.1093/scan/nsw052

Kanske, P., Böckler, A., Trautwein, F.-M., & Singer, T. (2015b). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind. *NeuroImage*, *122*, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Kendon, A. (1967a). Some functions of gaze direction in social interaction. *Acta Psychologica*, *26*, 22–63.

Kendon, A. (1967b). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22–63. https://doi.org/10.1016/0001-6918(67)90005-4

Kendon, A. (1969). The consistency of gaze patterns in social interaction. *British Journal of Psychology, 60*(4), 481–494.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking Perspective in Conversation: The Role of Mutual Knowledge in Comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Khalid, S., Deska, J. C., & Hugenberg, K. (2016). The Eyes Are the Windows to the Mind: Direct Eye Gaze Triggers the Ascription of Others' Minds. *Personality and Social Psychology Bulletin*, *42*(12), 1666–1677. https://doi.org/10.1177/0146167216669124

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*(2), 191–204. https://doi.org/10.1111/j.2044-8295.1998.tb02680.x

Kindermann, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*(2), 157–166.

Kircher, T., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., & Krach, S. (2009). Online mentalising investigated with functional MRI. *Neuroscience Letters*, *454*(3), 176–181. https://doi.org/10.1016/j.neulet.2009.03.026

Kirchner, J. C., Hatri, A., Heekeren, H. R., & Dziobek, I. (2011). Autistic Symptomatology, Face Processing Abilities, and Eye Fixation Patterns. *J Autism Dev Disord*, 10.

Kitterle, F. L. (1986). Psychophysics of lateral tachistoscopic presentation. *Brain and Cognition*, *5*(2), 131–162. https://doi.org/10.1016/0278-2626(86)90052-7

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

Klimecki, O. M., Leiberg, S., Lamm, C., & Singer, T. (2013). Functional Neural Plasticity and Associated Changes in Positive Affect After Compassion Training. *Cerebral Cortex*, *23*(7), 1552–1561. https://doi.org/10.1093/cercor/bhs142

Kononenko, D., & Lempitsky, V. (2015). Learning to look up: Realtime monocular gaze correction using machine learning. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4667–4675. https://doi.org/10.1109/CVPR.2015.7299098

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*(7042), 673–676. https://doi.org/10.1038/nature03701

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

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

Krach, S., Müller-Pinzler, L., Westermann, S., & Paulus, F. M. (2013). Advancing the neuroscience of social emotions with social immersion. *Behavioral and Brain Sciences*, *36*(4), 427–428. https://doi.org/10.1017/S0140525X12001951

Kret, M. E., Fischer, A. H., & De Dreu, C. K. W. (2015). Pupil mimicry correlates with trust in in-group partners with dilating pupils. *Psychological Science*, *26*(9), 1401–1410. https://doi.org/10.1177/0956797615588306

Kulke, L., Johannsen, J., & Rakoczy, H. (2019). Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PLOS ONE*, *14*(3), e0213772.

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. *Psychological Science*, *29*(6), 888–900. https://doi.org/10.1177/0956797617747090

Kuzmanovic, B., Georgescu, A. L., Eickhoff, S. B., Shah, N. J., Bente, G., Fink, G. R., & Vogeley, K. (2009). Duration matters: Dissociating neural correlates of detection and evaluation of social gaze. *NeuroImage*, *46*(4), 1154–1163. https://doi.org/10.1016/j.neuroimage.2009.03.037

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

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

Laillier, R., Viard, A., Caillaud, M., Duclos, H., Bejanin, A., de La Sayette, V., Eustache, F., Desgranges, B., & Laisney, M. (2019). Neurocognitive determinants of theory of mind across the adult lifespan. *Brain and Cognition*, *136*, 103588. https://doi.org/10.1016/j.bandc.2019.103588

Langdon, R., Connors, M., & Connaughton, E. (2017). Indirect task instructions better reveal theory-of-mind impairment, independent of executive dysfunction, in schizophrenia. *Psychiatry Research*, *256*, 342–344. https://doi.org/10.1016/j.psychres.2017.06.064

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Le, J., Kou, J., Zhao, W., Fu, M., Zhang, Y., Becker, B., & Kendrick, K. M. (2020). Oxytocin Facilitation of Emotional Empathy Is Associated With Increased Eye Gaze Toward the Faces of Individuals in Emotional Contexts. *Frontiers in Neuroscience*, *14*, 803. https://doi.org/10.3389/fnins.2020.00803

Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014). Training Preschoolers on First-Order False Belief Understanding: Transfer on Advanced ToM Skills and Metamemory. *Child Development*, *85*(6), 2404–2418. https://doi.org/10.1111/cdev.12267

Lecce, S., Bianco, F., Devine, R. T., Hughes, C., & Banerjee, R. (2014). Promoting theory of mind during middle childhood: A training program. *Journal of Experimental Child Psychology*, *126*, 52–67. https://doi.org/10.1016/j.jecp.2014.03.002

Lecce, S., Bottiroli, S., Bianco, F., Rosi, A., & Cavallini, E. (2015). Training older adults on Theory of Mind (ToM): Transfer on metamemory. *Archives of Gerontology and Geriatrics*, *60*(1), 217–226.

Lee, D. H., & Anderson, A. K. (2017). Reading What the Mind Thinks From How the Eye Sees. *Psychological Science*, *28*(4), 494–503. https://doi.org/10.1177/0956797616687364

Lehmann, K., Maliske, L., Böckler, A., & Kanske, P. (2019a). Social Impairments in Mental Disorders: Recent Developments in Studying the Mechanisms of Interactive Behavior. *Clinical Psychology in Europe*, *1*(2), e33143. https://doi.org/10.32872/cpe.v1i2.33143

Lehmann, K., Maliske, L., Böckler, A., & Kanske, P. (2019b). Social Impairments in Mental Disorders: Recent Developments in Studying the Mechanisms of Interactive Behavior. *Clinical Psychology in Europe*, *1*(2), e33143. https://doi.org/10.32872/cpe.v1i2.33143

Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS ONE*, *6*(3). https://doi.org/10.1371/journal.pone.0017798

Leppanen, J., Sedgewick, F., Treasure, J., & Tchanturia, K. (2018). Differences in the Theory of Mind profiles of patients with anorexia nervosa and individuals on the autism spectrum: A meta-analytic review. *Neuroscience & Biobehavioral Reviews, 90,* 146–163. https://doi.org/10.1016/j.neubiorev.2018.04.009

Li, D., Li, X., Yu, F., Chen, X., Zhang, L., Li, D., Wei, Q., Zhang, Q., Zhu, C., & Wang, K. (2017). Comparing the ability of cognitive and affective Theory of Mind in adolescent onset schizophrenia. *Neuropsychiatric Disease and Treatment*, *Volume 13,* 937–945. https://doi.org/10.2147/NDT.S128116

Lifesize. (2019). *2019 Impact of Video Conferencing Report*. Lifesize. https://www.lifesize.com/en/ldp/the-future-of-video-communication-and-meeting-productivity

List, J. A. (2007). On the interpretation of giving in Dictator Games. *Journal of Political Economy*, *115*(3), 482–493.

Lockwood, P. L., Hamonet, M., Zhang, S. H., Ratnavel, A., Salmony, F. U., Husain, M., & Apps, M. A. J. (2017). Prosocial apathy for helping others when effort is required. *Nature Human Behaviour*, *1*(7), 1–23. https://doi.org/10.1038/s41562-017-0131

Loh, K., & Redd, S. (2008). Understanding and preventing computer vision syndrome. *Malaysian family physician: the official journal of the Academy of Family Physicians of Malaysia*, *3*(3), 128.

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of Perceiving Eye-Gaze and Head Direction with Peripheral Vision: Implications for the Dynamics of Eye-Gaze Behavior. *Perception*, *37*(9), 1443–1457. https://doi.org/10.1068/p5896

Lundqvist, D., & Ohman, A. (2005). Emotion regulates attention: The relation between facial configurations, facial emotion, and visual attention. *Visual Cognition*, *12*(1), 51–84. https://doi.org/10.1080/13506280444000085

Magrelli, S., Jermann, P., Noris, B., Ansermet, F., Hentsch, F., Nadel, J., & Billard, A. (2013). Social orienting of children with autism to facial expressions and speech: A study with a wearable eye-tracker in naturalistic settings. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00840

Mäkelä, P., Näsänen, R., Rovamo, J., & Melmoth, D. (2001). Identification of facial images in peripheral vision. *Vision Research*, *41*(5), 599–610. https://doi.org/10.1016/S0042-6989(00)00259-5

Mann, S., Ewens, S., Shaw, D., Vrij, A., Leal, S., & Hillman, J. (2013). Lying Eyes: Why Liars Seek Deliberate Eye Contact. *Psychiatry, Psychology and Law*, *20*(3), 452–461. https://doi.org/10.1080/13218719.2013.791218

Marsh, A. A., Yu, H. H., Pine, D. S., & Blair, R. J. R. (2010). Oxytocin improves specific recognition of positive facial expressions. *Psychopharmacology*, *209*(3), 225–232. https://doi.org/10.1007/s00213-010-1780-4

Martin, G. B., & Clark, R. D. (1982). Distress crying in neonates: Species and peer specificity. *Developmental psychology*, *18*(1), 3.

Martínez-Velázquez, E. S., Ahuatzin González, A. L., Chamorro, Y., & Sequeira, H. (2020). The Influence of Empathy Trait and Gender on Empathic Responses. A Study With Dynamic Emotional Stimulus and Eye Movement Recordings. *Frontiers in Psychology*, *11*, 23. https://doi.org/10.3389/fpsyg.2020.00023

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396.

Mason, M. F., Tatkow, E. P., & Macrae, C. N. (2005). The look of love: Gaze shifts and person perception. *Psychological science*, *16*(3), 236–239.

Matheson, R. (2007). *I am Legend*. Macmillan.

McCrackin, S. D., & Itier, R. J. (2019). Perceived Gaze Direction Differentially Affects Discrimination of Facial Emotion, Attention, and Gender – An ERP Study. *Frontiers in Neuroscience*, *13*, 517. https://doi.org/10.3389/fnins.2019.00517

McKee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision Research*, *24*(1), 25–32.

McKone, E. (2004). Isolating the special component of face recognition: Peripheral identification and a Mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 181.

Meins, E. (2012). Social relationships and children's understanding of mind: Attachment, internal states, and mind-mindedness. In M. Siegal & L. Surian (Hrsg.), *Access to Language and Cognitive Development* (S. 22–43). Oxford University Press.

Mikolajczak, M., Pinon, N., Lane, A., de Timary, P., & Luminet, O. (2010). Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance. *Biological Psychology*, *85*(1), 182–184. https://doi.org/10.1016/j.biopsycho.2010.05.010

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773. https://doi.org/10.1037/a0016854

Mitchell, P., Sheppard, E., & Cassidy, S. (2021). Autism and the double empathy problem: Implications for development and mental health. *British Journal of Developmental Psychology*, *39*(1), 1–18. https://doi.org/10.1111/bjdp.12350

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mojzisch, A., Schilbach, L., Helmert, J. R., Velichkovsky, B. M., & Vogeley, K. (2006). The effects of self-involvement on attention, arousal, and facial expression during social interaction with virtual others: A psychophysiological study. *Social Neuroscience*, *1*(3–4), 184–195.

Moutinho, R., Castro, S. L., & Silva, S. (2021). Look at me: The relation between empathy and fixation on the emotional eye-region in low vs. high social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 70,* 101610. https://doi.org/10.1016/j.jbtep.2020.101610

Murnighan, J. K., & Wang, L. (2016). The social world as an experimental game. *Organizational Behavior and Human Decision Processes*, *136*, 80–94. https://doi.org/10.1016/j.obhdp.2016.02.003

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., Hammond, N., Murphy, D., & Happé, F. (2017). A new test of advanced theory of mind: The "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders: A new test of advanced theory of mind. *Autism Research*, *10*(6), 1120–1132. https://doi.org/10.1002/aur.1744

N'Diaye, K., Sander, D., & Vuilleumier, P. (2009). Self-relevance processing in the human amygdala: Gaze direction, facial expression, and emotion intensity. *Emotion*, *9*(6), 798–806. https://doi.org/10.1037/a0017845

Neath, K. N., & Itier, R. J. (2015). Fixation to features and neural processing of facial expressions in a gender discrimination task. *Brain and Cognition*, *99*, 97–111. https://doi.org/10.1016/j.bandc.2015.05.007

Neath-Tavares, K. N., & Itier, R. J. (2016). Neural processing of fearful and happy facial expressions during emotion-relevant and emotion-irrelevant tasks: A fixation-to-feature approach. *Biological Psychology*, *119*, 122–140. https://doi.org/10.1016/j.biopsycho.2016.07.013

Newman-Norlund, R., Bosga, J., Meulenbroek, R., & Bekkering, H. (2008). Anatomical substrates of cooperative joint-action in a continuous motor task: Virtual lifting and balancing. *NeuroImage*, *41*(1), 169–177. https://doi.org/10.1016/j.neuroimage.2008.02.026

Nijman, S. A., Veling, W., Greaves-Lord, K., Vermeer, R. R., Vos, M., Zandee, C. E. R., Zandstra, D. C., Geraets, C. N. W., & Pijnenborg, G. H. M. (2019). Dynamic Interactive Social Cognition Training in Virtual Reality (DiSCoVR) for social cognition and social functioning in people with a psychotic disorder: Study protocol for a multicenter randomized controlled trial. *BMC Psychiatry*, *19*(1), 272. https://doi.org/10.1186/s12888-019-2250-0

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

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *NeuroImage*, *130*, 248–260. https://doi.org/10.1016/j.neuroimage.2016.02.026

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

Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*, 60–74. https://doi.org/10.1016/j.cognition.2018.01.001

Oliver, L. D., Vieira, J. B., Neufeld, R. W. J., Dziobek, I., & Mitchell, D. G. V. (2018). Greater involvement of action simulation mechanisms in emotional vs cognitive empathy. *Social Cognitive and Affective Neuroscience*, *13*(4), 367–380. https://doi.org/10.1093/scan/nsy013

Olsson, A., Kopsida, E., Sorjonen, K., & Savic, I. (2016). Testosterone and estrogen impact social evaluations and vicarious emotions: A double-blind placebo-controlled study. *Emotion*, *16*(4), 515–523. https://doi.org/10.1037/a0039765

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Ornaghi, V., Brockmeier, J., & Grazzani, I. (2014). Enhancing social cognition by training children in emotion understanding: A primary school study. *Journal of Experimental Child Psychology*, *119*, 26–39. https://doi.org/10.1016/j.jecp.2013.10.005

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315.

Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of Advanced Theory-of-Mind Tasks. *Child Development*, *87*(6), 1971–1991. https://doi.org/10.1111/cdev.12566

Ozonoff, S., & Miller, J. N. (1995). Teaching theory of mind: A new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders*, *25*(4), 415–433. https://doi.org/10.1007/BF02179376

Öztürk, Y., Özyurt, G., Turan, S., Mutlu, C., Tufan, A. E., & Akay, A. P. (2020). *Relationships Between Theory of Mind (ToM) and Attachment Properties in Adolescent with Social Axiety Disorder*. *57*(1), 65–70.

Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *274*, 81–93.

Palanica, A., & Itier, R. J. (2011). Searching for a perceived gaze direction using eye tracking. *Journal of Vision*, *11*(2), 19–19. https://doi.org/10.1167/11.2.19

Palanica, A., & Itier, R. J. (2012). Attention Capture by Direct Gaze is Robust to Context and Task Demands. *Journal of Nonverbal Behavior*, *36*(2), 123–134. https://doi.org/10.1007/s10919-011-0128-z

Palanica, A., & Itier, R. J. (2014). Effects of peripheral eccentricity and head orientation on gaze discrimination. *Visual Cognition*, *22*(9–10), 1216–1232. https://doi.org/10.1080/13506285.2014.990545

Palanica, A., & Itier, R. J. (2015). Eye gaze and head orientation modulate the inhibition of return for faces. *Attention, Perception, & Psychophysics*, *77*(8), 2589–2600. https://doi.org/10.3758/s13414-015-0961-y

Palanica, A., & Itier, R. J. (2017). Asymmetry in Gaze Direction Discrimination Between the Upper and Lower Visual Fields. *Perception*, *46*(8), 941–955. https://doi.org/10.1177/0301006616686989

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

Paulus, D. C. (2006). *Der Saarbrückener Persönlichkeitsfragebogen (IRI) zur Messung von Empathie*. http://bildungswissenschaften.uni-saarland.de/personal/paulus/empathy/SPF_Artikel.pdf

Perner, J., & Wimmer, H. (1985). “John Thinks That Mary Thinks That. . .” Attribution of Second-Order Beliefs by 5- to 1 O-Year-Old Children. *Journal of Experimental Child Psychology*, *39*(3), 437–471.

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

Peterson, C. C., Peterson, J. L., & Webb, J. (2000). Factors influencing the development of a theory of mind in blind children. *British Journal of Developmental Psychology*, *18*(3), 431–447. https://doi.org/10.1348/026151000165788

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2516–2528. https://doi.org/10.1016/j.neubiorev.2013.07.017

Piedmont, R. L. (2014). Inter-item Correlations. In A. C. Michalos (Hrsg.), *Encyclopedia of Quality of Life and Well-Being Research* (S. 3303–3304). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1493

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the Detection of Signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174.

Preckel, K., Kanske, P., Singer, T., Paulus, F. M., & Krach, S. (2016). Clinical trial of modulatory effects of oxytocin treatment on higher-order social cognition in autism spectrum disorder: A randomized, placebo-controlled, double-blind and crossover trial. *BMC Psychiatry*, *16*(1), 329. https://doi.org/10.1186/s12888-016-1036-x

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

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria. https://www.R-project.org/.

Rakoczy, H. (2017). In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese*, *194*(3), 689–707. https://doi.org/10.1007/s11229-015-0860-8

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit Theory of Mind Is Even More Unified Than Previously Assumed: Belief Ascription and Understanding

Aspectuality Emerge Together in Development. *Child Development*, *86*(2), 486–502. https://doi.org/10.1111/cdev.12311

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage*, *50*(4), 1639–1647. https://doi.org/10.1016/j.neuroimage.2010.01.052

Reiter, A. M. F., Kanske, P., Eppinger, B., & Li, S.-C. (2017). The Aging of the Social Mind—Differential Effects on Components of Social Understanding. *Scientific Reports*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-10669-4

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, *51*(3), 279–290.

Ricciardelli, P., Lugli, L., Pellicano, A., Iani, C., & Nicoletti, R. (2016). Interactive effects between gaze direction and facial expression on attentional resources deployment: The task instruction and context matter. *Scientific Reports*, *6*(1), 21706. https://doi.org/10.1038/srep21706

Rice, K., Moraczewski, D., & Redcay, E. (2016). Perceived live interaction modulates the developing social brain. *Social Cognitive and Affective Neuroscience*, *11*(9), 1354–1362. https://doi.org/10.1093/scan/nsw060

Riechelmann, E., Gamer, M., Böckler, A., & Huestegge, L. (2020). How ubiquitous is the direct gaze advantage? Evidence for an averted gaze advantage in a gaze discrimination task. *Attention, Perception, & Psychophysics*.

Riechelmann, E., Raettig, T., Böckler, A., & Huestegge, L. (2019). Gaze Interaction: Anticipation-Based Control of the Gaze of Others. *Psychological Research*, *85*, 302–321. https://doi.org/10.1007/s00426-019-01257-4

Rigoulot, S., D'Hondt, F., Honoré, J., & Sequeira, H. (2012). Implicit emotional processing in peripheral vision: Behavioral and neural evidence. *Neuropsychologia*, *50*(12), 2887–2896. https://doi.org/10.1016/j.neuropsychologia.2012.08.015

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, *9*(3), 507–513. https://doi.org/10.3758/BF03196306

Roberts, M., Delicato, L. S., Herrero, J., Gieselmann, M. A., & Thiele, A. (2007). Attention alters spatial integration in macaque V1 in an eccentricity-dependent manner. *Nature Neuroscience*, *10*(11), 1483–1491.

Romero, V., Fitzpatrick, P., Roulier, S., Duncan, A., Richardson, M. J., & Schmidt, R. C. (2018). Evidence of embodied social competence during conversation in high functioning children with autism spectrum disorder. *PLOS ONE*, *13*(3), e0193906. https://doi.org/10.1371/journal.pone.0193906

Rosi, A., Cavallini, E., Bottiroli, S., Bianco, F., & Lecce, S. (2016). Promoting theory of mind in older adults: Does age play a role? *Aging & Mental Health*, *20*(1), 22–28. https://doi.org/10.1080/13607863.2015.1049118

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

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

Ruffman, T., Puri, A., Galloway, O., Su, J., & Taumoepeau, M. (2018). Variety in parental use of "want" relates to subsequent growth in children's theory of mind. *Developmental Psychology*, *54*(4), 677–688. https://doi.org/10.1037/dev0000459

Sadler, M. (2020). *COVID-19 Software Industry Statistics*. TrustRadius. https://www.trustradius.com/vendor-blog/covid-19-software-industry-data-and-statistics

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., & Scherer, K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion. *European Journal of Cognitive Psychology*, *19*(3), 470–480. https://doi.org/10.1080/09541440600757426

Santiesteban, I., Gibbard, C., Drucks, H., Clayton, N., Banissy, M. J., & Bird, G. (2021). Individuals with Autism Share Others' Emotions: Evidence from the Continuous Affective Rating and Empathic Responses (CARER) Task. *Journal of Autism and Developmental Disorders*, *51*(2), 391–404. https://doi.org/10.1007/s10803-020-04535-y

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

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience*, *1*(3–4), 284–298. https://doi.org/10.1080/17470910601000446

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, *19*(2), 65–72. https://doi.org/10.1016/j.tics.2014.11.007

Schenkel, L. S., Marlow-O’Connor, M., Moss, M., Sweeney, J. A., & Pavuluri, M. N. (2008). Theory of mind and social inference in children and adolescents with bipolar disorder. *Psychological Medicine*, *38*(6), 791–800. https://doi.org/10.1017/S0033291707002541

Scherer, S. E., & Schiff, M. R. (1973). Perceived intimacy, physical distance and eye contact. *Perceptual and motor skills*, *36*(3), 835–841.

Schilbach, L. (2015a). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, *3*, 130–135. https://doi.org/10.1016/j.cobeha.2015.03.006

Schilbach, L. (2015b). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, *3*, 130–135. https://doi.org/10.1016/j.cobeha.2015.03.006

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex*, *130*, 362–386. https://doi.org/10.1016/j.cortex.2020.06.010

Schindler, S., & Kissler, J. (2016). Selective visual attention to emotional words: Early parallel frontal and visual activations followed by interactive effects in visual cortex. *Human Brain Mapping*, *37*(10), 3575–3587.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The Facilitated Processing of Threatening Faces: An ERP Analysis. *Emotion*, *4*(2), 189–200. https://doi.org/10.1037/1528-3542.4.2.189

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2020). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*. https://doi.org/10.1037/bul0000303

Schurz, M., Tholen, M. G., Perner, J., Mars, R. B., & Sallet, J. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: A review using probabilistic atlases from different imaging modalities: Brain Anatomy Underlying Temporo-Parietal Junction. *Human Brain Mapping*, *38*(9), 4788–4805. https://doi.org/10.1002/hbm.23675

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., De Brito, S. A., McCrory, E. J. P., & Viding, E. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*(1), 53–63. https://doi.org/10.1093/scan/nsr023

Sekuler, R., Watamaniuk, S. N., & Blake, R. (2002). Motion perception. *Stevens' Handbook of Experimental Psychology*.

Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition*, *12*(1), 127–144. https://doi.org/10.1080/13506280444000157

Senju, A., Hasegawa, T., & Tojo, Y. (2005). Does perceived direct gaze boost detection in adults and children with and without autism? The stare-in-the-crowd effect revisited. *Visual Cognition*, *12*(8), 1474–1496. https://doi.org/10.1080/13506280444000797

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

Senju, A., & Johnson, M. H. (2009b). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134. https://doi.org/10.1016/j.tics.2008.11.009

Senju, A., Kikuchi, Y., Hasegawa, T., Tojo, Y., & Osanai, H. (2008). Is anyone looking at me? Direct gaze detection in children with and without autism. *Brain and Cognition, 67*(2), 127–139. https://doi.org/10.1016/j.bandc.2007.12.001

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*(13), 3054–3067. https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of Mind and Emotion Regulation Difficulties in Adolescents With Borderline Traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(6), 563-573.e1. https://doi.org/10.1016/j.jaac.2011.01.017

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Available at SSRN 2160588*.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, *30*(6), 855–863. https://doi.org/10.1016/j.neubiorev.2006.06.011

Singer, T., & Lamm, C. (2009). The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Spunt, R. P., & Adolphs, R. (2014). Validating the Why/How contrast for functional MRI studies of Theory of Mind. *NeuroImage*, *99*, 301–311. https://doi.org/10.1016/j.neuroimage.2014.05.023

Sripada, C. S., Angstadt, M., Banks, S., Nathan, P. J., Liberzon, I., & Phan, K. L. (2009). Functional neuroimaging of mentalizing during the trust game in social anxiety disorder: *NeuroReport*, *20*(11), 984–989. https://doi.org/10.1097/WNR.0b013e32832d0a67

Stallen, M., & Sanfey, A. G. (2013). The Cooperative Brain. *The Neuroscientist*, *19*(3), 292–303. https://doi.org/10.1177/1073858412469728

Stein, T., Senju, A., Peelen, M. V., & Sterzer, P. (2011). Eye contact facilitates awareness of faces during interocular suppression. *Cognition*, *119*(2), 307–311. https://doi.org/10.1016/j.cognition.2011.01.008

Stietz, J., Jauk, E., Krach, S., & Kanske, P. (2019). Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Frontiers in Psychiatry*, *10*, 126. https://doi.org/10.3389/fpsyt.2019.00126

Stoet, G. (2017a). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Stoet, G. (2017b). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 82.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers Can Attribute Second-Order Beliefs. *Developmental Psychology, 30*(3), 395–402.

Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology, 149,* 81–97. https://doi.org/10.1016/j.jecp.2015.11.007

Symeonidou, I., Dumontheil, I., Ferguson, H. J., & Breheny, R. (2020). Adolescents are delayed at inferring complex social intentions in others, but not basic (false) beliefs: An eye-movement investigation. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021820920213

Taubert, J., van Golde, C., & Verstraten, F. A. J. (2017). Who is the Usual Suspect? Evidence of a Selection Bias Toward Faces That Make Direct Eye Contact in a Lineup Task. *I-Perception*, *8*(1), 204166951769041. https://doi.org/10.1177/2041669517690411

Taumoepeau, M., & Ruffman, T. (2008). Stepping Stones to Others' Minds: Maternal Talk Relates to Child Mental State Language and Emotion Understanding at 15, 24, and 33 Months. *Child Development*, *79*(2), 284–302. https://doi.org/10.1111/j.1467-8624.2007.01126.x

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

Tholen, M. G., Trautwein, F., Böckler, A., Singer, T., & Kanske, P. (2020). Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. *Human Brain Mapping*, *41*, 2611–2628. https://doi.org/10.1002/hbm.24966

Tipples, J. (2005). Orienting to Eye Gaze and Face Processing. *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 843–856. https://doi.org/10.1037/0096-1523.31.5.843

To, M. P. S., Regan, B. C., Wood, D., & Mollon, J. D. (2011). Vision out of the corner of the eye. *Vision Research*, 12.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005a). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005b). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., & Vaish, A. (2013). Origins of Human Cooperation and Morality. *Annual Review of Psychology*, *64*(1), 231–255. https://doi.org/10.1146/annurev-psych-113011-143812

Tompkins, V. (2015). Improving low-income preschoolers' theory of mind: A training study. *Cognitive Development*, *36*, 1–19. https://doi.org/10.1016/j.cogdev.2015.07.001

Trautwein, F.-M., Kanske, P., Böckler, A., & Singer, T. (2020). Differential benefits of mental training types for attention, compassion, and theory of mind. *Cognition*, *194*, 104039. https://doi.org/10.1016/j.cognition.2019.104039

Tylén, K., Allen, M., Hunter, B. K., & Roepstorff, A. (2012). Interaction vs. observation: Distinctive modes of social cognition in human brain and behavior? A combined fMRI and eye-tracking study. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00331

Valk, S. L., Bernhardt, B. C., Trautwein, F.-M., Böckler, A., Kanske, P., Guizard, N., Collins, D. L., & Singer, T. (2017). Structural plasticity of the social brain: Differential change after socio-affective and cognitive mental training. *Science Advances*, *3*(10), e1700489.

van der Wel, R. P., Welsh, T., & Böckler, A. (2018). Talking heads or talking eyes? Effects of head orientation and sudden onset gaze cues on attention capture. *Attention, Perception, & Psychophysics*, *80*(1), 1–6. https://doi.org/10.3758/s13414-017-1462-y

van Hazebroek, B. C. M., Olthof, T., & Goossens, F. A. (2017). Predicting aggression in adolescence: The interrelation between (a lack of) empathy and social goals: Predicting Aggression in Adolescence. *Aggressive Behavior*, *43*(2), 204–214. https://doi.org/10.1002/ab.21675

van Honk, J., & J.L.G. Schutter, D. (2007). Testosterone Reduces Conscious Detection of Signals Serving Social Correction: Implications for Antisocial Behavior. *Psychological Science*, *18*(8), 663–667. https://doi.org/10.1111/j.1467-9280.2007.01955.x

van Zonneveld, L., Platje, E., de Sonneville, L., van Goozen, S., & Swaab, H. (2017). Affective empathy, cognitive empathy and social attention in children at high risk of criminal behaviour. *Journal of Child Psychology and Psychiatry*, *58*(8), 913–921. https://doi.org/10.1111/jcpp.12724

Vaughan Van Hecke, A., Mundy, P. C., Acra, C. Françoise., Block, J. J., Delgado, C. E. F., Parlade, M. V., Meyer, J. A., Neal, A. R., & Pomares, Y. B. (2007). Infant Joint Attention, Temperament, and Social Competence in Preschool Children. *Child Development*, *78*(1), 53–69. https://doi.org/10.1111/j.1467-8624.2007.00985.x

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332

Von Gaudecker, H.-M., Holler, R., Janys, L., Siflinger, B., & Zimpelmann, C. (2020). *Labour supply in the early stages of the CoViD-19 Pandemic: Empirical Evidence on hours, home office, and expectations*.

Von Grünau, M., & Anston, C. (1995). The detection of gaze direction: A stare-in-the-crowd effect. *Perception*, *24*(11), 1297–1313.

Vrana, S. R. (1993). The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG. *Psychophysiology*, *30*(3), 279–286. https://doi.org/10.1111/j.1469-8986.1993.tb03354.x

Vuilleumier, P. (2002a). Facial expression and selective attention. *Current Opinion in Psychiatry*, *15*(3), 291–300.

Vuilleumier, P. (2002b). Facial expression and selective attention: *Current Opinion in Psychiatry*, *15*(3), 291–300. https://doi.org/10.1097/00001504-200205000-00011

Wade, M., Prime, H., Jenkins, J. M., Yeates, K. O., Williams, T., & Lee, K. (2018). On the relation between theory of mind and executive functioning: A developmental cognitive neuroscience perspective. *Psychonomic Bulletin & Review*, *25*(6), 2119–2140. https://doi.org/10.3758/s13423-018-1459-0

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive and Affective Neuroscience*, *1*(2), 107–121. https://doi.org/10.1093/scan/nsl018

Washburn, D., Wilson, G., Roes, M., Rnic, K., & Harkness, K. L. (2016). Theory of mind in social anxiety disorder, depression, and comorbid conditions. *Journal of Anxiety Disorders*, *37*, 71–77. https://doi.org/10.1016/j.janxdis.2015.11.004

Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasue, H., Nakamura, M., Miyashita, Y., & Masuda, N. (2014). Two distinct neural mechanisms underlying indirect reciprocity. *Proceedings of the National Academy of Sciences*, *111*(11), 3990–3995. https://doi.org/10.1073/pnas.1318570111

Wedekind, C., & Milinski, M. (2000). Cooperation through Image Scoring in Humans. *Science*, *288*(5467), 850–852.

Weimer, A. A., Parault Dowds, S. J., Fabricius, W. V., Schwanenflugel, P. J., & Suh, G. W. (2017). Development of constructivist theory of mind from middle childhood to early adulthood and its relation to social cognition and behavior. *Journal of Experimental Child Psychology*, *154*, 28–45. https://doi.org/10.1016/j.jecp.2016.10.002

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, *61*, 161–265.

Wesselmann, E. D., Cardoso, F. D., Slater, S., & Williams, K. D. (2012). To Be Looked at as Though Air: Civil Attention Matters. *Psychological Science*, *23*(2), 166–168. https://doi.org/10.1177/0956797611427921

Wheeler, R. W., Baron, J. C., Michell, S., & Ginsburg, H. J. (1979). Eye contact and the perception of intelligence. *Bulletin of the Psychonomic Society*, *13*(2), 101–102. https://doi.org/10.3758/BF03335025

Williams, J. H. G., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *NeuroImage*, *25*, 133–140.

Williams, K. D., Shore, W. J., & Grahe, J. E. (1998). The silent treatment: Perceptions of its behaviors and associated feelings. *Group Processes & Intergroup Relations*, *1*(2), 117–141.

Williams, M., Moss, S., Bradshaw, J., & Mattingley, J. (2005). Look at me, I'm smiling: Visual search for threatening and nonthreatening facial expressions. *Visual Cognition*, *12*(1), 29–50. https://doi.org/10.1080/13506280444000193

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Winter, K., Spengler, S., Bermpohl, F., Singer, T., & Kanske, P. (2017). Social cognition in aggressive offenders: Impaired empathy, but intact theory of mind. *Scientific Reports, 7*(1), 1–10. https://doi.org/10.1038/s41598-017-00745-0

Wolf, L., Freund, Z., & Avidan, S. (2010). An eye for an eye: A single camera gaze-replacement method. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 817–824. https://doi.org/10.1109/CVPR.2010.5540133

Wood, E., Baltrušaitis, T., Morency, L.-P., Robinson, P., & Bulling, A. (2018). GazeDirector: Fully Articulated Eye Gaze Redirection in Video. *Computer Graphics Forum, 37*(2), 217–225. https://doi.org/10.1111/cgf.13355

Wu, Z., Chen, X., Gros-Louis, J., & Su, Y. (2018). 'She is looking at me! Shall I share?' How Chinese and American preschoolers respond to eye gaze during sharing. *Social Development, 27*(2), 447–460. https://doi.org/10.1111/sode.12278

Xiao, Y., Geng, F., Riggins, T., Chen, G., & Redcay, E. (2019). Neural correlates of developing theory of mind competence in early childhood. *NeuroImage, 184*, 707–716. https://doi.org/10.1016/j.neuroimage.2018.09.079

Yokoyama, T., Sakai, H., Noguchi, Y., & Kita, S. (2014). Perception of Direct Gaze Does Not Require Focus of Attention. *Scientific Reports, 4*, 3858. https://doi.org/10.1038/srep03858

Yokoyama, T., & Takeda, Y. (2019). Gaze Cuing Effects in Peripheral Vision. *Frontiers in Psychology, 10*, 708. https://doi.org/10.3389/fpsyg.2019.00708

Yu, H.-H., Verma, R., Yang, Y., Tibballs, H. A., Lui, L. L., Reser, D. H., & Rosa, M. G. P. (2010). Spatial and temporal frequency tuning in striate cortex: Functional uniformity and specializations related to receptive field eccentricity. *European Journal of Neuroscience, 31*(6), 1043–1062. https://doi.org/10.1111/j.1460-9568.2010.07118.x

Ze, O., Thoma, P., & Suchan, B. (2014). Cognitive and affective empathy in younger and older individuals. *Aging & Mental Health*, *18*(7), 929–935. https://doi.org/10.1080/13607863.2014.899973

Zhou, C., Jiang, T., & Zhu, L. (2018). Direct Gaze Blurs Self-Other Boundaries. *The Journal of General Psychology*, *145*(3), 280–295. https://doi.org/10.1080/00221309.2018.1469465

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and Nonverbal Communication of Deception. In *Advances in Experimental Social Psychology* (Bd. 14, S. 1–59). Elsevier. https://doi.org/10.1016/S0065-2601(08)60369-X

APPENDIX

CURRICULUM VITAE

LIST OF PUBLICATIONS

References marked with an asterisk indicate studies included in the dissertation

Peer-reviewed journal articles

*Breil, C., Micheli, L.R. & Böckler, A. (unpublished). Golden gazes: Gaze direction and emotional context promote prosocial behavior by modulating perceptions of others' empathy and perspective-taking.

*Breil, C., Raettig, T., Pittig, R., van der Wel, R., Welsh, T. & Böckler, A. (submitted). Don't look at me like that: integration of gaze direction and facial expression. *Journal of Experimental Psychology: Human Perception and Performance.*

*Breil, C., Huestegge, L., Pittig, R. & Böckler, A. (under review). How eccentricity modulates attention capture by direct gaze and facial motion. *Journal of Vision.*

*Breil, C., Huestegge, L. & Böckler, A. (2021). From eye to arrow: Attention capture by direct gaze requires more than just the eyes. *Attention, Perception & Psychophysics.* DOI: 10.3758/s13414-021-02382-2

*Breil, C., Kanske, P., Pittig, R. & Böckler, A. (2021). A revised instrument for the assessment of empathy and Theory of Mind in adolescents: Introducing the EmpaToM-Y. *Behavior Research Methods.* DOI: 10.3758/s13428-021-01589-3

*Breil, C. & Böckler, A. (2021) Look away to listen: The interplay of emotional context and eye contact in video conversations. *Visual Cognition.* DOI: 10.1080/13506285.2021.1908470

*Breil, C. & Böckler, A. (2020). The lens shapes the view: On task dependency in ToM research. *Current Behavioral Neuroscience Reports, 7*(2), 41-50. DOI: 10.1007/s40473-020-00205-6

Didino, D., Breil, C., & Knops, A. (2019). The influence of semantic processing and response latency on the SNARC effect. *Acta Psychologica*, 196, 75-86.

Abstracts and conference presentations

Breil, C. & Böckler, A. (2021, September). Listen through your eyes: The interplay of gaze behavior and emotional context in video conversations. 49th Meeting of the European Brain and Behaviour Society in Lausanne, Switzerland.

Breil, C. & Böckler, A. (2021, June). Prosocial behavior is influenced by perceptions of other people's social cognition. 1st International Conference on Social Neuroscience in Ecologically Valid Conditions in Moscow, Russia.

Breil, C. & Böckler, A. (2020, July). Heads and hearts: Investigation of Theory of Mind and Empathy in Adolescents. 9th A-DOK in Regensburg, Germany.

Breil, C. & Böckler, A. (2020, March). Validating the EmpaToM-Y: A new instrument to assess social understanding in adolescents. In Dobel, C., Giesen, C., Grigutsch, L., A., Kaufmann, J., M., Kovács, G., Meissner, F., Rothermund, K., & Schweinberger, S., R. (Eds). Abstracts of the 62nd Conference of Experimental Psychologists Jena (p. 39), Lengerich: Pabst Science Publishers (Conference canceled).

Breil, C. & Böckler, A. (2020, January). Assessment of social understanding in adolescents: Introducing the EmpaToM-Y. 1st GK Doctoral Symposium on Cognitive Science in Tuebingen, Germany.

Breil, C. & Böckler, A. (2019, April). From eye to arrow: Influences of non-social and social cues on attention capture. 61st Conference of Experimental Psychologists (TeaP) in London, UK.

STATEMENT OF INDIVIDUAL AUTHOR CONTRIBUTIONS AND OF LEGAL SECOND PUBLICATION RIGHTS

Publication 1 (complete reference): Breil, C., Huestegge, L. & Böckler, A. (2021). From eye to arrow: Attention capture by direct gaze requires more than just the eyes. *Attention, Perception & Psychophysics.* DOI: 10.3758/s13414-021-02382-2

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AB	CB	LH		
Data Collection	CB				
Data Analysis and Interpretation	CB	AB	LH		
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	CB	AB	LH		

Explanations (if applicable): This manuscript is currently in press.

Publication 2 (complete reference): Breil, C., Huestegge, L., Pittig, R. & Böckler, A. (under review). How eccentricity modulates attention capture by direct gaze and facial motion. *Journal of Vision*.

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AB	LH	RP	CB	
Data Collection	RP				
Data Analysis and Interpretation	CB	AB	RP	LH	
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	CB	AB	LH	RP	

Explanations (if applicable): This manuscript is currently under review.

Publication 3 (complete reference): Breil, C., Raettig, T., Pittig, R., van der Wel, R., Welsh, T. & Böckler, A. (submitted). Don't look at me like that: integration of gaze direction and facial expression. *Journal of Experimental Psychology: Human Perception and Performance.*

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AB	RK			
Data Collection	RK				

Data Analysis and Interpretation	TR	CB			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	CB	AB	TR	TW, RvdW	RK

Explanations (if applicable): This manuscript has been submitted to the Journal of Experimental Psychology: Human Perception and Performance.

Publication 4 (complete reference): Breil, C. & Böckler, A. (2021). Look away to listen: The interplay of emotional context and eye contact in video conversations. *Visual Cognition.* DOI: 10.1080/13506285.2021.1908470					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AB AB	CB CB			
Data Collection	CB	AB			
Data Analysis and Interpretation	CB	AB			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	CB	AB			

Explanations (if applicable): This manuscript has been published.

Publication 5 (complete reference): Breil, C., Micheli, L.R. & Böckler, A. (unpublished). Golden gazes: Gaze direction and emotional context promote prosocial behavior by modulating perceptions of others' empathy and perspective-taking					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	LRM AB	CB CB	AB LRM		
Data Collection	CB				
Data Analysis and Interpretation	CB	LRM	AB		
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	LRM	CB	AB		

Explanations (if applicable): This manuscript is currently in preparation.

Publication 6 (complete reference): Breil, C., Kanske, P., Pittig, R. & Böckler, A. (2021). A revised instrument for the assessment of empathy and Theory of Mind in adolescents: Introducing the EmpaToM-Y. *Behavior Research Methods*.DOI: 10.3758/s13428-021-01589-3					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AB CB	CB AB	PK PK	RP RP	
Data Collection	CB				
Data Analysis and Interpretation	CB	AB	PK	RP	
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	CB	AB	PK	RP	

Explanations (if applicable): This manuscript has been published.

Publication 7(complete reference): Breil, C. & Böckler, A. (2020). The Lens Shapes the View: on Task Dependency in ToM Research. *Current Behavior Neuroscience Reports, 7(2), 41-50.* DOI: 10.1007/s40473-020-00205-6					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development					
Data Collection					
Data Analysis and Interpretation					
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	CB	AB			

Explanations (if applicable): This manuscript has been published. Since the article is a narrative review, there was no data collection or analysis involved.

The doctoral researcher confirms that she/he has obtained permission from both the publishers and the co-authors for legal second publication.
The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.
I also confirm my primary supervisor's acceptance.

Christina Breil 12.11.2021, Würzburg

Doctoral Researcher's Name Date Place Signature

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

Publication 1 (complete reference): Breil, C., Huestegge, L. & Böckler, A. (2021). From eye to arrow: Attention capture by direct gaze requires more than just the eyes. *Attention, Perception & Psychophysics.* DOI: 10.3758/s13414-021-02382-2

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CB	AB	LH		
2	CB	AB	LH		
3	CB	AB	LH		

Explanations (if applicable):

Publication 2 (complete reference): Breil, C., Huestegge, L., Pittig, R. & Böckler, A. (under review). How eccentricity modulates attention capture by direct gaze and facial motion. *Journal of Vision.*

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CB	AB	RP	LH	
2	CB	AB	RP	LH	
3	CB	AB	RP	LH	
4	CB	AB	RP	LH	

Explanations (if applicable):

Publication 3 (complete reference): Breil, C., Raettig, T., Pittig, R., van der Wel, R., Welsh, T. & Böckler, A. (submitted). Don't look at me like that: integration of gaze direction and facial expression. *Journal of Experimental Psychology: Human Perception and Performance.*

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CB	AB	TR		
2	CB	AB	TR		
3	CB	AB	TR		
4	TR	AB	CB		
5	TR	AB	CB		
6	TR	AB	CB		

Explanations (if applicable):

Publication 4 (complete reference): Breil, C. & Böckler, A. (2021). Look away to listen: The interplay of emotional context and eye contact in video conversations. *Visual Cognition.* DOI: 10.1080/13506285.2021.1908470

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CB	AB			
2	CB	AB			

Explanations (if applicable):

Publication 5 (complete reference): Breil, C., Micheli, L.R. & Böckler, A. (unpublished). Golden gazes: Gaze direction and emotional context promote prosocial behavior by modulating perceptions of others' empathy and perspective-taking					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	LRM	CB	AB		
2	CB	LRM	AB		
3	CB	LRM	AB		
4	CB	LRM	AB		
5	CB	LRM	AB		

Explanations (if applicable):

Publication 6 (complete reference): Breil, C., Kanske, P., Pittig, R. & Böckler, A. (2021). A revised instrument for the assessment of empathy and Theory of Mind in adolescents: Introducing the EmpaToM-Y. *Behavior Research Methods.* DOI: 10.3758/s13428-021-01589-3					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CB	AB	PK	RP	
2	CB	AB	PK	RP	
3	CB	AB	PK	RP	
4	CB	AB	PK	RP	

Explanations (if applicable):

Publication 7 (complete reference): Breil, C. & Böckler, A. (2020). The Lens Shapes the View: on Task Dependency in ToM Research. *Current Behavior Neuroscience Reports, 7(2), 41-50.* DOI: 10.1007/s40473-020-00205-6					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CB	AB			
2	CB	AB			

Explanations (if applicable):

I also confirm my primary supervisor's acceptance.

Christina Breil 12.11.2021, Würzburg

Doctoral Researcher's Name Date Place Signature

AFFIDAVIT

I hereby confirm that my thesis entitled *“Look at me and I will feel you: eye contact and social understanding”* is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date   Signature

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich an Eides statt, die Dissertation *„Schau mich an und ich sehe dich: Blickkontakt und Sozialverstehen“* eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

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

Ort, Datum   Unterschrift

MANUSCRIPT 1

FROM EYE TO ARROW: ATTENTION CAPTURE BY DIRECT GAZE REQUIRES MORE THAN JUST THE EYES

Christina Breil[1], Lynn Huestegge[1], Anne Böckler[2,3]

[1]Julius-Maximilians-University of Würzburg, Würzburg, Germany

[2]Leibniz University Hannover, Hannover, Germany

[3]Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

This article has been published in *Attention, Perception & Psychophysics* (2021) with DOI: 10.3758/s13414-021-02382-2

Abstract

Human attention is strongly attracted by direct gaze and sudden onset motion. The sudden direct gaze effect refers to the processing advantage for targets appearing on peripheral faces that suddenly establish eye contact. Here, we investigate the necessity of social information for attention capture by (sudden onset) ostensive cues. Six experiments involving 204 participants applied (1) naturalistic faces, (2) arrows, (3) schematic eyes, (4) naturalistic eyes, or schematic facial configurations (5) without or (6) with head turn to an attention capture paradigm. Trials started with two stimuli oriented towards the observer and two stimuli pointing into the periphery. Simultaneous to target presentation, one direct stimulus changed to averted and one averted stimulus changed to direct, yielding a 2×2 factorial design with direction and motion cues being absent or present. We replicated the (sudden) direct gaze effect for photographic faces, but found no corresponding effects in experiments 2-6. Hence, a holistic and socially meaningful facial context seems vital for attention capture by direct gaze.

Statement of significance

The present study highlights the significance of context information for social attention. Our findings demonstrate that the direct gaze effect, that is, the prioritization of direct gaze over averted gaze, critically relies on the presentation of a meaningful holistic and naturalistic facial context. This pattern of results is evidence in favor of early effects of surrounding social information on attention capture by direct gaze.

*Keywords***:** social cognition, attention capture, direct gaze, social cues, face perception, social interaction

Introduction

Faces are special to us. In our everyday lives, we encounter a vast amount of information that is relevant for our well-being, yet we exhibit a striking susceptibility for facial configurations from early on (Goren et al., 1975). One of the first and most frequently fixated regions within the human face are the eyes (Arizpe et al., 2017). Eye gaze conveys essential information about attentional, intentional and emotional states and is indispensable for social communication (Schilbach, 2015b; Tomasello & Carpenter, 2007). Accordingly, we are specialized in detecting the direction of another's attention, and human eyes with their white sclera seem particularly effective in conveying this information (Emery, 2000). Neurons in the superior temporal sulcus of monkeys and humans specifically respond to the direction of the eyes (Perret et al., 1985), and several facial features further emphasize the salience of this region (Emery, 2000).

A well-known effect in the social attention literature is "gaze following", the finding that we rapidly shift attention according to others' gaze direction, resulting in a processing advantage for this location (Driver et al., 1999; Friesen & Kingstone, 1998; Frischen et al., 2007). In addition, humans are extraordinarily sensitive to direct gaze. Researchers have proposed that direct gaze immediately activates sub-cortical structures and facilitates subsequent cognitive and perceptual processing (Senju & Johnson, 2009b). Even though - under some circumstances such as very brief or masked presentations - a processing advantage for averted gaze has also been reported (e.g. Riechelmann et al., 2020), direct gaze seems to be preferred over averted gaze from childhood on (Farroni et al., 2004) and constitutes a magnet for human attention also in adulthood (Mojzisch et al., 2006; Palanica & Itier, 2012). We detect a specific face among other faces faster when it directly looks at us, which has been labeled as the "stare-in-the-crowd-effect" (Doi et al., 2009; Palanica & Itier, 2011; Von Grünau & Anston, 1995). Moreover, discriminating direct from averted gaze is still accurate when a second task is performed concurrently, whereas discriminating averted left from averted right gaze suffers from dual task demands (Yokoyama et al., 2014).

Attentional capture by direct gaze is particularly pronounced when eye contact co-occurs with sudden onset motion of the face, two cues that seem to influence information processing additively and in parallel (Böckler, van der Wel, et al., 2014). In this task, participants identified targets that were presented on the forehead of one of four face images in a 2×2 within-subjects design with gaze direction (direct or averted) and apparent face motion (static or sudden) as within-subject factors. With this initial combination of gaze and motion cues, Böckler et al. (2014) found that targets are classified faster when they were presented on faces that suddenly established eye contact (sudden direct gaze effect).

Until now, the sudden direct gaze effect has been investigated exclusively with images of real faces. In contrast to gaze following research, where effects of a variety of social and non-social ostensive stimuli (such as arrows) have been systematically addressed and demonstrated (Friesen & Kingstone, 1998; Frischen et al., 2007; J. Hietanen & Yrttimaa, 2005; Ristic et al., 2002; Tipples, 2005), the role of particular stimulus features on attention capture by direct gaze remains unknown. Generalizing findings from one paradigm to the other is, however, precarious due to fundamental differences between them: While cues are centrally presented and specifically attended to in gaze cueing, stimuli appear in the periphery and serve as distractors in our task. At the current state of research, one cannot estimate the extent to which the observed effect relies on direct *gaze* at all as compared to the mere feeling of “being addressed”. Here, we present an experimental series that is specifically designed to close this gap. We probe whether and to what degree the sudden direct gaze effect relies on naturalistic and holistic social information. Specifically, we ran the attention capture paradigm by Böckler et al. (2014) with six different sets of stimuli: photographs of real faces (photographic and holistic face information) similar to the original study, arrows (no social but directional information), isolated schematic eyes (absence of both photographic and holistic face information), photographs of isolated eyes (photographic; no holistic face information), and schematic faces (holistic; no photographic face information). Following the notion that congruence of head and eye orientation shapes the detection of gaze direction (Conty et al., 2006), two versions of schematic faces were employed: One with frontal head view in all experimental conditions (*no head turn*; experiment 5) and one switching between frontal and deviated head view between conditions, hence creating the impression of a head turn movement similar to the one in experiment 1 (experiment 6). A feature that is common to all experiments (probably except for arrows) is the ostensive signal of being addressed: the stimulus is either targeted towards the observer or into the periphery. This setup hence manipulated the degree to which holistic and photographic social information was provided and allowed targeting the boundary conditions that enable attention capture by direct gaze. Arrow stimuli were implemented in order to directly compare social with directional information.

We hypothesized that the gaze effect would decrease together with the level of holistic and naturalistic social information. Specifically, we expected the strongest gaze effect with photographs of real faces (experiment 1) and attenuated or absent gaze/direction effects for arrows (experiment 2), isolated eyes (experiment 3 and 4) and schematic faces (experiments 5 and 6).

Materials and methods

Experimental setup and procedure

We employed the paradigm of Böckler et al. (2014) with six different stimulus sets. In the original experiment, participants saw two displays, each consisting of four images of the same face positioned around a central fixation cross. Participants were repeatedly instructed to keep their eyes fixated on this cross throughout the experiment. In the first display, two of the faces depicted direct gaze while the other two faces looked to the side. Each of the four faces had the number "8" positioned on their forehead. After 1500ms, the number-8 figures were replaced by three distractor letters ("E"/"U") and one target letter ("H"/"S") to which the participants were required to respond by pressing "H" or "S" with the index fingers of both hands on a keyboard. Simultaneous to target presentation, two of the faces changed their orientation: one direct face suddenly changed to averted (*sudden-averted*) and one averted face suddenly looked straight ahead (*sudden-direct*). The other two faces remained static (*static-direct; static-averted*). Across 384 trials, identity and position of the target and distraction letters as well as locations of gaze and motion cues appeared equally often in all possible combinations. A sample trial sequence with photographic face stimuli from our experiment 1 is displayed in Figure 1.

Participants

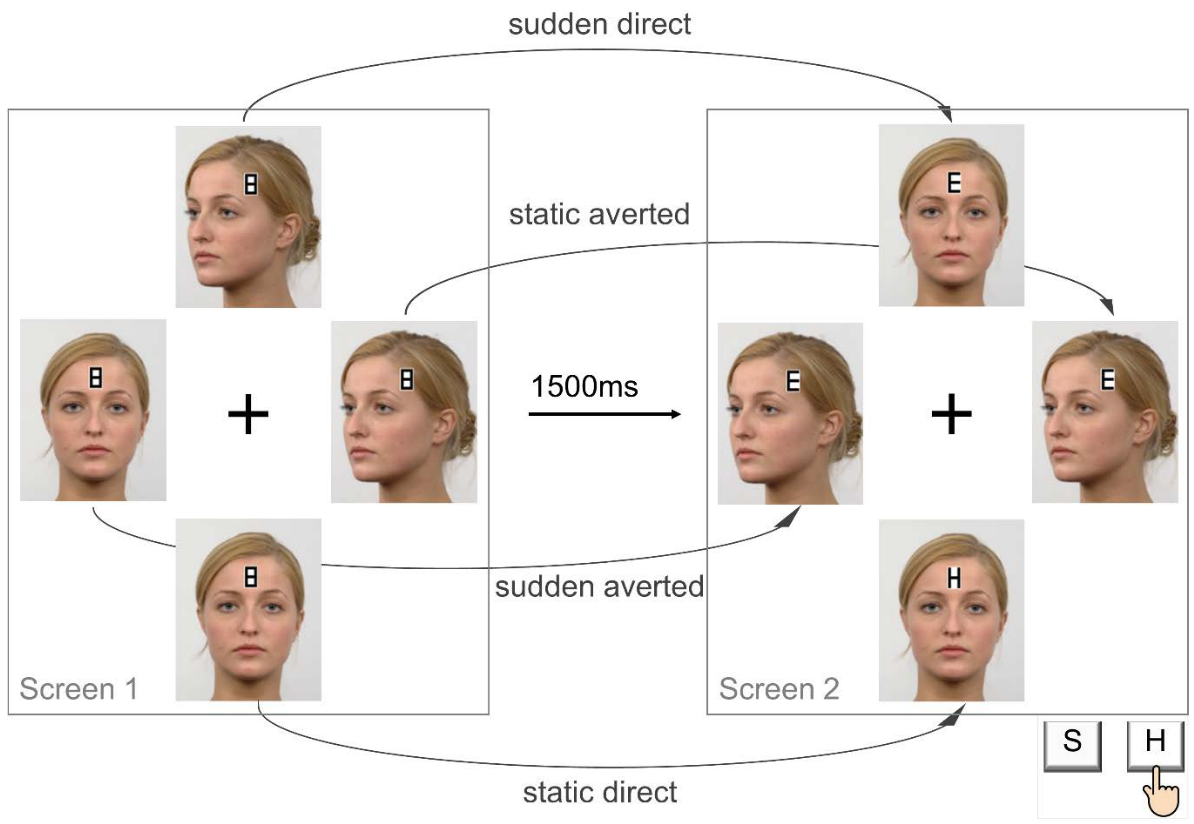
The number of participants for each experiment was determined using G*power3 (Faul et al., 2007b), assuming 80% power and an α of .05 with a small effect size, resulting in 34 participants for each experiment. In sum, we tested 204 participants with normal or corrected to normal vision (Table 1). All participants gave informed consent and were compensated with 7€ or course credit. The present study complies with the ethical standards of the 1964 Declaration of Helsinki regarding the treatment of human participants.

Stimuli

Stimuli of all experiments are displayed in Figure 2. We chose one female face from the Radboud Face Database (RaFD) (Langner et al., 2010) for experiment 1 that we showed either in direct of averted position. The images were 200×250 pixels (1.21×1.52° of visual angle). To investigate whether directional, symbolic signals are sufficient for attention capture, we used arrows in experiment 2. In experiments 3 and 4, we employed isolated eyes to address the necessity of a holistic facial context. The eyes in experiment 4 also stem from the RaFD. For experiment 3, we designed schematic eyes on the basis of images from the RaFD.

Figure 1

Sample trial sequence of experiment 1



Note. Number-8-figures overlaid the four stimuli in screen 1 and were replaced by one target and three distraction letters after 1500ms. Simultaneously, one direct stimulus changed to averted and one averted stimulus changed to direct while the other two stimuli remained unchanged, resulting in four experimental conditions. Participants were required to react as fast as possible to the target letter by pressing the corresponding response key. This set-up was kept for experiments 1-6, but stimuli varied (see figure 2).

Table 1

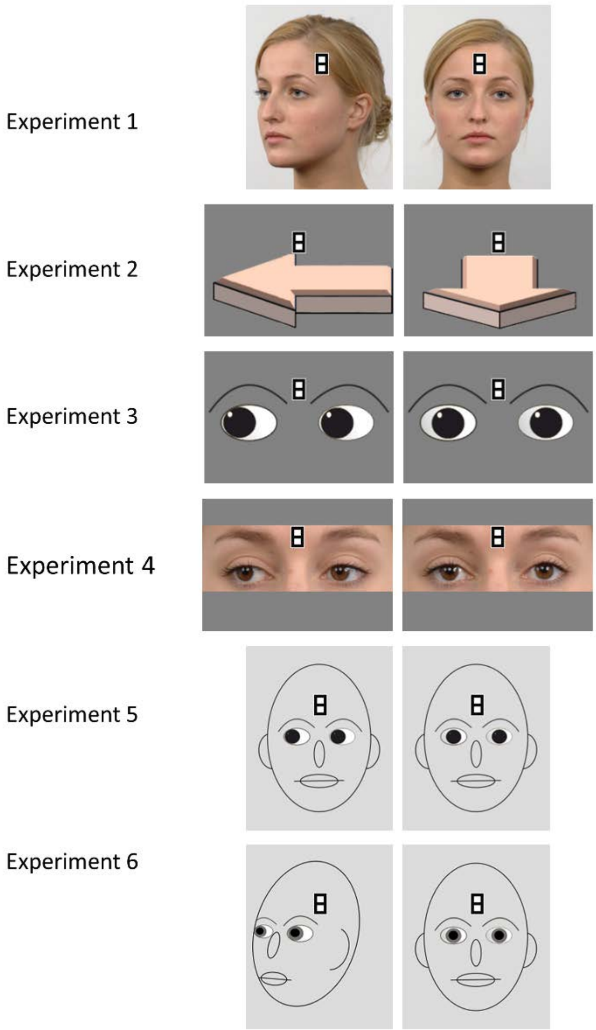
Data exclusions and gender, age and handedness of participants of experiments 1-6

Experiment	Total N I final sample	Excluded due to mean error rate +2SD	Females	Mean age (SD)	Right-handed
1	33	1	25	22.23 (±3.22)	25
2	32	2	24	23.91 (±3.60)	27
3	32	2	26	27.10 (±9.11)	31
4	33	1	22	22.45 (±2.61)	32
5	32	2	24	24.31 (±4.85)	30
6	32	2	26	23.91 (±4.01)	29

The eye regions of the same images were used for experiment 4. For both experiments, the images were 259×180 pixels (1.57×1.09° of visual angle). For experiment 5, we inserted the comic-style eyes from experiment 3 into a schematic facial configuration that was based on the images of the *direct* condition of experiment 1. However, analogous to experiments 3-4, we kept frontal head orientation for averted stimuli, hence constricting the illusory motion to the area of the eyes. In experiment 6, we took the comic-style faces one step closer to the photographic faces of experiment 1 by rotating the averted-stimulus by 45° to create the impression of a head-turning movement and by adding pupils to the eyes. Across all experiments, we devoted special attention to keeping all relevant aspects of the stimuli as similar as possible.

Figure 2

Stimuli of experiments 1-6



Note. Experiment 1: images of real faces (replication). Experiment 2: arrows (no social but directional information). Experiment 3: schematic eyes (no photographic and no holistic context). Experiment 4: images of real eyes (photographic social information; no holistic context). Experiment 5: schematic face without head turn (no photographic social information; holistic context). Experiment 6: schematic face with head turn (no photographic social information; holistic context).

Analyses

Reaction time (RT) was defined as the time window from target onset until the first key press. In each experiment, participants with error rates +2SD above the global mean were removed. In the remaining data sets, RTs ±2SD of the participant's mean in each condition and all RTs of trials that were associated with errors were excluded from further analysis. Table 2 provides an overview of error and exclusion rates as well as mean RTs for each condition of each experiment. RTs of all experiments are visualized in Figure 2. All data was submitted to two 2 × 2 × 6 mixed effects ANOVAs with the within-subject factors motion (static, sudden) and gaze/direction (direct, averted) and the between-subject factor experiment (1-6), entering mean RTs and error rates as dependent variables. Differences between experiments were further investigated with individual ANOVAs for each experiment. Finally, for each of experiments 2-6, an ANOVA on RTs with the between-subjects factor *experiment* was conducted to compare it to experiment 1. For better interpretation of null results, we drew on Bayesian statistics in addition to traditional null-hypothesis testing. In each case of a non-significant gaze/direction or motion effect, we performed Bayesian t-tests to calculate non-directional Bayes factors (BFs) with a prior distribution value of 1. Following Rouder et al. (2009), BFs were computed as f (data | H_0) / f (data | H_1) and interpreted as evidence for the null hypothesis when BF > 3 or as evidence for the alternative hypothesis when BF < 1/3.

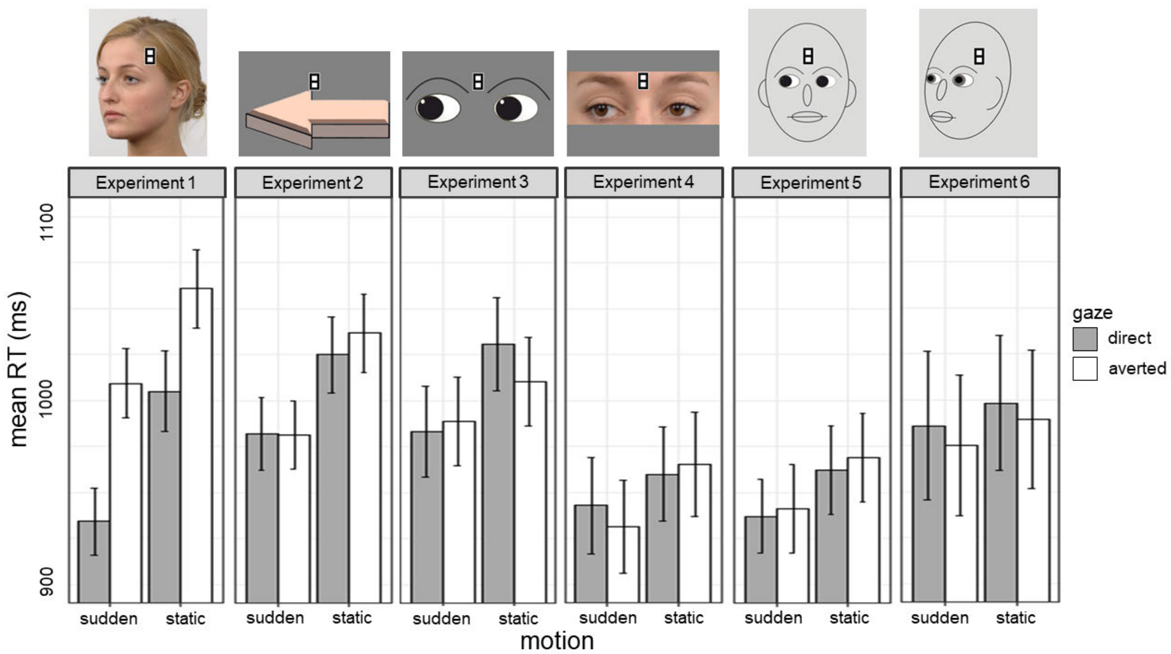
Considering the non-normal distribution of data, we took an additional, alternative approach to statistical analysis. First, participant- and trial-wise exclusions were based on predefined threshold values instead of on means and SDs. Hence, in each experiment, data sets of participants who performed below or at chance (error rate ≥ 50%) were excluded from the analysis. In the remaining data sets, trials with RTs below 150ms or above 2500ms were removed. In a second step, RTs were log transformed and entered into a 2 × 2 repeated measures ANOVA with the within-subject factors motion (static, sudden) and gaze/direction (direct, averted) for each experiment individually. In a similar way to our original analysis, Bonferroni corrected *t*-tests were applied to resolve interaction effects and non-directional Bayes factors were calculated for non-significant main effects of gaze or motion. The results of this analysis, which revealed a highly similar pattern as the results described in the following section, can be found in Appendix A.

Results and discussion

The mean RTs of each combination of gaze and motion for each of experiments 1-6 are displayed in Figure 2.

Figure 2

Mean RTs for all conditions for each of experiments 1-6

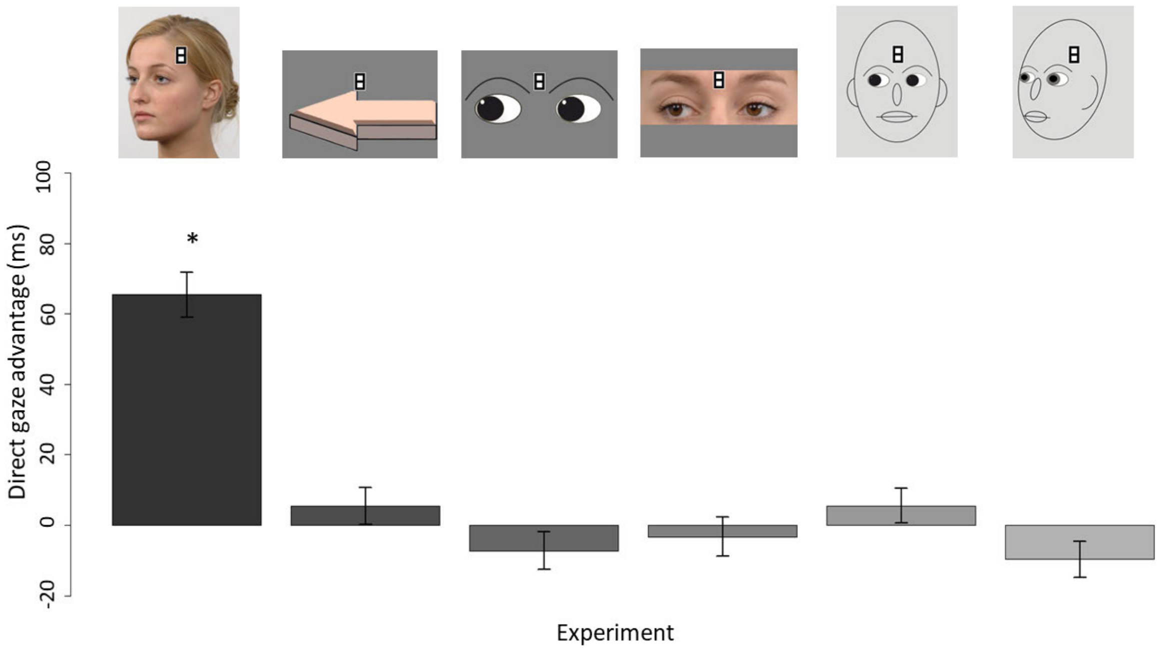


Note. Mean response times for targets appearing on stimuli directed towards participants are presented in grey; mean response times for targets appearing on averted stimuli are depicted in white. Error bars represent standard errors.

The size of the direct gaze effect for each experiment is visualized in Figure 3. The data sets that the following analyses are based on are available in the Open Science Framework (DOI: 10.17605/OSF.IO/2JZGS). None of the experiments was preregistered.

Figure 3

Direct gaze/direction advantage for each of experiments 1-6



Note. Direct gaze/direction advantage calculated as mean RT of correct responses for targets appearing on averted stimuli – mean RT of correct responses for targets appearing on direct stimuli. Error bars represent standard errors.

Table 2

Reaction times (RT), error rates and exclusion rates across experiments 1-6

Experiment	Condition										Exclusion rate
	total		sudden direct		static direct		sudden averted		static averted		
	Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	
1 (photographic faces)	1002 (±123)	3.67 (±3.03)	934 (±104)	2.72 (±1.84)	1005 (±126)	3.73 (±3.50)	1009 (±109)	4.14 (±2.89)	1061 (±123)	4.1 (±3.49)	7.18
2 (arrows)	1006 (±115)	2.9 (±2.33)	982 (±112)	2.64 (±2.75)	1025 (±117)	2.83 (±2.17)	981 (±106)	2.67 (±2.23)	1037 (±120)	3.45 (±2.14)	6.25
3 (schematic eyes)	1003 (±139)	3.1 (±2.84)	983 (±139)	3.16 (±2.94)	989 (±136)	3.48 (±3.05)	989 (±136)	2.54 (±2.47)	1011 (±136)	3.22 (±2.90)	6.3

4 (photographic eyes)	950 (±151)	3.18 (±2.65)	943 (±151)	3.09 (±2.82)	960 (±147)	3.50 (±2.87)	932 (±146)	2.97 (±2.32)	965 (±163)	3.16 (±2.66)	7.44
5 (schematic faces without head turn)	952 (±130)	3.72 (±2.83)	937 (±114)	3.48 (±3.10)	962 (±136)	3.74 (±2.98)	941 (±137)	3.55 (±2.49)	969 (±136)	4.10 (±2.8)	8.14
6 (schematic faces with head turn)	987 (±215)	3.52 (±2.78)	986 (±230)	3.84 (±2.78)	998 (±207)	2.77 (±2.33)	975 (±218)	3.48 (±2.73)	990 (±213)	4.00 (±3.19)	8.24

Note. Mean RTs of correct responses in ms; error and exclusion rates in %.

Omnibus analysis

To test for overall differences between experiments, we performed an omnibus analysis by entering the mean correct RTs of all six experiments with the between-subjects factor experiment (1, 2, 3, 4, 5 and 6) and the within-subject factors gaze/direction (direct, averted) and motion (static, sudden) to a mixed effects ANOVA. We found a significant gaze/direction effect across all experiments ($F(1, 188) = 15.87$, $p = .001$, $\eta^2 = .001$) as well as a motion effect ($F(1, 188) = 82.96$, $p < .001$, $\eta^2 = .014$). Hence, overall, participants responded faster to targets appearing on stimuli that were direct towards them compared to away from them, and to targets appearing on stimuli that changed direction simultaneous to target presentation. Critically, the interaction effect of gaze/direction × experiment ($F(5, 188) = 24.28$, $p < .001$, $\eta^2 = .008$) was also significant, indicating that the size of the gaze/direction effect differed substantially between the six experiments. We also found a small two-way interaction between motion × experiment ($F(5, 188) = 3.45$, $p = .005$, $\eta^2 = .003$) and a small three-way interaction ($F(5, 188) = 2.835$, $p = .017$, $\eta^2 = .001$), suggesting a modulation of the motion effect as well as a modulation of the interplay between gaze/direction and motion by experiment. Performing the same analysis with error rates as a dependent variable revealed a significant main effect of motion ($F(1, 188) = 5.41$, $p = .021$, $\eta^2 = .003$), indicating more errors for static stimuli, in line with RT results. No other effects were significant (all *p*s > .05). Exact *p*-values and effect sizes for all non-significant effects are reported in Table 3 in the Appendix B.

To disentangle interaction effects of this initial omnibus analysis, individual analyses on RTs were performed for each experiment and results are reported in the following.

Experiment 1: photographs of real faces

As in the original study (Böckler, van der Wel, et al., 2014), RTs were shorter when targets were presented on a face with direct gaze compared to averted gaze ($F(1, 32) = 84.11$,

$p < .001$, $\eta^2 = .076$). We also found a significant motion effect with shorter RTs in the *sudden* condition ($F(1, 32) = 17.80$, $p < .001$, $\eta^2 = .067$). In line with RT results, participants produced more errors in response to averted compared to direct faces, as evident in a significant main effect of gaze on error rates ($F(1, 32) = 6.87$, $p = .013$, $\eta^2 = .023$). No other effects were significant (all *p*s > .05).

This pattern of results replicates earlier findings of attention capture by direct gaze.

Experiment 2: arrows

In contrast to Experiment 1, the main effect of gaze/direction was not significant ($F(1,31) = 0.902$, $p = .350$; $BF_{01} = 3.5$). The main effect of motion was significant, with shorter RTs for moving stimuli ($F(1, 31) = 32.10$, $p < .001$, $\eta^2 = .046$). None of the other effects were significant for RTs and error rates (all *p*s > .05).

We conducted an ANOVA on RTs with the additional between-subjects factor experiment (1, 2) to systematically test for differences between the gaze/direction effects for pictures of faces and arrows. Besides significant overall effects for gaze/direction ($F(1, 61) = 60.31$, $p < .001$, $\eta^2 = .022$) and motion ($F(1, 61) = 52.71$, $p < .001$, $\eta^2 = .016$), we found a significant gaze/direction × experiment interaction ($F(1, 61) = 42.52$, $p = .001$, $\eta^2 = .016$), emphasizing the difference between magnitudes of the gaze/direction effect between naturalistic face stimuli and arrows. The three-way interaction effect of gaze/direction × motion × experiment was also significant ($F(1, 61) = 6.75$, p $.012$, $\eta^2 = .002$). Because the two-way interaction of gaze/direction and motion was not even close to significance in either of the individual experiments, we refrained from further analyzing this interaction. No other effects were significant (all *p*s > .05).

These results indicate that being pointed at by a directional symbolic stimulus such as an arrow does not capture attention.

Experiment 3: schematic eyes

The main effect of gaze/direction was not significant (F (1,31) = 1.832, p = .186; BF_{01} = 2.36). We found a significant main effect of motion (F(1, 31) = 24.79, p < .001, η^2 = .016), with faster RTs to targets appearing on moving stimuli. The gaze × motion interaction effect was significant (F(1, 31) = 6.49, p = .016, η^2 = .002), with faster reactions to targets appearing on averted stimuli compared to direct stimuli in the static condition, but not in the sudden condition (static: t(31) = 2.504, p = .035; sudden: p = .802). No other effects were significant in RTs or error rates (all ps > .05).

The ANOVA with the additional between-subjects factor experiment (1, 3) revealed significant main effects of gaze (F(1, 63) = 42.27, p < .001, η^2 = .013) and motion (F(1, 61) = 34.81, p < .001, η^2 = .035) as well as significant gaze × experiment interaction (F(1, 61) = 65.87, p = .001, η^2 = .020), emphasizing the difference in the direct gaze effect between real faces and schematic eyes. The three-way interaction of gaze × motion × experiment also reached significance (F(1, 61) = 9.44, p = .003, η^2 = .002), reflecting the presence of a gaze × motion interaction for schematic eyes, which was absent for faces. No other effects were significant (all ps > .05).

These results suggest that eye gaze without photographic facial context information is insufficient to trigger the direct gaze advantage.

Experiment 4: photographs of real eyes

We found the typical main effect of motion on RTs (F(1, 32) = 12.96, p = .001, η^2 = .007), but no main effect of gaze (F (1,32) = 0.302, p = .586; BF_{01} = 4.73). No other effects were significant (all ps > .05).

Adding the between-subjects factor experiment (1, 4), we found significant main effects of gaze (F(1, 64) = 45.03, p < .001, η^2 = .014) and motion (F(1, 64) = 28.81, p < .001, η^2 =

.026), as well as a significant interaction effect of gaze × experiment ($F(1, 64) = 54.93$, $p < .001$, $\eta^2 = .016$), supporting the difference in gaze effects between photographs of faces and eyes. In addition, there was a motion × experiment interaction ($F(1, 64) = 4.94$, $p = .030$, $\eta^2 = .016$) due to a smaller motion effect for photographs of eyes compared to faces, and a gaze × motion × experiment interaction ($F(1, 61) = 6.06$, $p = .017$, $\eta^2 = .001$). No other effects were significant (all *p*s > .05).

These findings further emphasize that the presentation of a holistic face is vital for the direct gaze effect. We found no statistically significant gaze effect for photographs of real eyes and the BF provided substantial evidence for the null hypothesis: This conclusion is in line with Böckler et al. (2015) reporting a collapse of the (sudden) direct gaze effect when the integration of eye and face was disrupted by presenting faces upside down.

Experiment 5: schematic faces without head-turn

Again, we found a significant main effect of motion on RTs ($F(1, 31) = 19.18$, $p < .001$, $\eta^2 = .010$) and no attention capture effect for direct gaze (F (1,31) = 1.22, $p = .278$; $BF_{01} = 3.42$). No other effects were significant (all *p*s > .05).

Adding the between-subjects factor experiment (1, 5), we found significant main effects of gaze ($F(1, 63) = 65.55$, $p < .001$, $\eta^2 = .021$) and motion ($F(1, 63) = 30.45$, $p < .001$, $\eta^2 = .031$). Again, we found a significant interaction of gaze × experiment ($F(1, 63) = 46.70$, $p < .001$, $\eta^2 = .015$), indicating that the gaze effect was substantially larger for photographic face stimuli. There was a small interaction of gaze × motion in this ANOVA ($F(1, 63) = 4.77$, $p = .033$, $\eta^2 = .005$) suggesting that the motion effect was also larger for photographs of real faces than for schematic faces. No other effects were significant (all *p*s > .05).

Revealing no direct gaze effect in schematic faces, these findings further emphasize the relevance of naturalistic social cues for the direct gaze advantage.

Experiment 6: schematic faces with head-turn

No significant effects were found for experiment 6 (all *p*s > .05; gaze: $F(1,31) = 3.623$; $p = .066$; $BF_{01} = 1.09$; motion: $F(1,31) = 1.667$; $p = .206$; $BF_{01} = 2.54$).

When adding the between-subjects factor experiment (1, 6), we found significant main effects of gaze ($F(1, 63) = 40.06$, $p < .001$, $\eta^2 = .007$) and motion ($F(1, 63) = 17.33$, $p = .003$, $\eta^2 = .012$) as well as a significant gaze × experiment interaction effect ($F(1, 63) = 72.712$, $p < .001$, $\eta^2 = .01$), suggesting a gaze effect for photographic face stimuli only. The experiment × motion interaction was also significant ($F(1, 63) = 7.13$, $p = .01$, $\eta^2 = .005$) because the motion effect was absent for schematic and turning faces. No other effects were significant (all *p*s > .05).

The absence of a direct gaze advantage in schematic faces with head-turns further supports the necessity of naturalistic social stimuli for the direct gaze effect. In addition, we found an absence of a motion effect for the schematic faces with head-turn. Given that the head-turn orientation in schematic faces was kept identical to that in the original photographic faces (experiment 1), the absence of a motion effect is not due to a mere reduction of the extent of motion per se. Nonetheless, our constructed schematic faces did not elicit apparent motion effects, even though the same faces without head-turn did (experiment 5).

General Discussion

The present study investigated the aptitude of various ostensive stimuli to capture attention. Six experiments systematically manipulated the degree of photographic and holistic social context information and assessed the (sudden) direct gaze effect. Specifically, we compared (1) photographic human faces, (2) arrows, (3) schematic and (4) photographic isolated eye stimuli, and schematic face stimuli (5) without or (6) with head turn.

Firstly, results of experiment 1 revealed a reliable direct gaze effect for human faces, replicating prior studies (Böckler, van der Wel, et al., 2014; Boyer & Wang, 2018) and substantiating the notion of an exceptional processing of direct gaze cues (Senju & Johnson, 2009b). In contrast, no attention capture was found for symbolic self-directed cues (arrows, experiment 2). This finding is somewhat surprising given that spatial cueing is reliably observed for arrows (Daum & Gredebäck, 2011), but fits the notion that reflexive attention to arrows and to biologically relevant gaze cues are based on distinct neural systems (Ristic et al., 2002).

Interestingly, we found no direct gaze effect for schematic or photographic eyes outside of a facial context. These results seem at odds with some previous findings of processing advantages for direct over averted gaze with eyes-only stimuli (Y.-C. Chen & Yeh, 2012; Conty et al., 2006; Greene et al., 2011; D. A. Hayward & Ristic, 2015; Senju et al., 2005). However, this discrepancy can be accounted for when taking a closer look at the tasks. In two of the abovementioned studies (Conty et al., 2006; Senju et al., 2005), participants were asked to detect direct gaze stimuli among averted gaze distractors as quickly as possible. Hence, while gaze was a distractor in our task, it was the target in these experiments. Without task-driven requirements to process gaze characteristics, isolated eye stimuli in our experiment may not have been encoded sufficiently to capture attention. Furthermore, results of Conty et al. (2006) indicate another critical mediator of gaze processing: Even for eyes-only stimuli, direct gaze was more salient than averted gaze when the visible part of the face (the region around the eyes) was oriented in the same direction as the pupils. This aspect was lacking in our averted eyes-only stimuli (see Figure 1). Note, however, that congruency between head and eye orientation was not sufficient to elicit a direct gaze effect with schematic face stimuli in our study (experiment 6), indicating that naturalness of the stimuli is a further prerequisite for the direct gaze effect as investigated with our paradigm. Finally, a critical difference between our paradigm and spatial cueing paradigms, such as in Hayward and Ristic (2015), is the location of cue presentation within the visual field: While cues are usually presented centrally and hence overtly fixated in spatial cueing, participants in our experiments were explicitly instructed to fixate on a centrally presented cross throughout the task so that cues would appear in the periphery and be covertly attended. This difference can have a substantial impact on (social) attention (Boyer & Wang, 2018; Riechelmann et al., 2020). Perception is most accurate and contrast-sensitive in the foveal region and decreases from the central-to-peripheral gradient of the visual field (Burnat, 2015; Kitterle, 1986). While face processing is particularly impressive even at high eccentricities (Hershler et al., 2010) it remains unclear whether this processing advantage extends onto isolated eyes and, hence, specific investigations of related effects are necessary.

Remarkably, the gaze cueing paradigm robustly produces orientation effects with numerous stimulus types, including photographic and schematic face stimuli, both upright and inverted (Tipples, 2005), faces with strabismus (J. Hietanen & Yrttimaa, 2005) and even arrows (Ristic et al., 2002). Within this paradigm, systematic investigation of facial feature information indicated that local processing of the eyes has a major impact on reflexive orienting (Frischen et al., 2007; Tipples, 2005). It might be worthwhile to conduct similar investigations with the attention capture paradigm to assess systematically the impact of local feature information on the sudden direct gaze effect. From what we know, it appears that the attention capture paradigm, in contrast to that of spatial cueing, requires a holistic facial context that allows for a meaningful interpretation of the embedded eyes. In previous experiments, images of realistic faces elicited a direct gaze effect even when only the eyes were moving (van der Wel et al., 2018) and that this effect collapsed when the eyes were closed or when the integration of eyes and face was disrupted by inverting face stimuli (Böckler et al., 2015). However, visibility of the eyes was not strictly necessary: The effect still occurred when eyes were covered with opaque sunglasses. Taken together, these findings indicate that, instead of being drawn by direct gaze *per se,* our preferential attention to the ostensive signal of "being addressed by someone" has an impact on attention capture (Csibra & Gergely, 2009).

In contrast to the pronounced direct gaze effect for the more realistic photographic face stimuli, there was no indication of such an effect for schematic facial configurations, irrespective of head orientation. One explanation is that the naturalness of the stimuli is a further crucial factor for direct gaze advantages (A. F. de C. Hamilton, 2016). Naturalistic stimuli convey a larger potential to interact socially, and the mere opportunity to do so can alter social attention (Laidlaw et al., 2011; but see also Riechelmann, Raettig, et al., in press). A promising line of future research may be to gradually and independently manipulate the degrees of naturalness of stimuli (e.g. by employing avatars) and of the interaction opportunity (e.g. by employing real or virtual reality setups; see Rubo et al. (2020)) in order to tackle the minimum requirements and relative contribution of realistic face information and interaction opportunity for attention capture by direct gaze.

Taken together, we replicated the direct gaze effect for photographic faces, but did not find an effect of gaze/arrow direction in any other stimulus configuration. While we are aware that the interpretation of null results is tricky, further results support our conclusions. First, between-experiment ANOVAs revealed significant differences between direct gaze effects of experiment 1 and all other experiments. Second, numerical differences between averted minus direct stimuli in experiments 2-6 are also negligible (maximum 3ms), which suggests that the absence of significant effects is not due to mere power issues. Finally, frequentist inferences of non-significant gaze effects were supported by Bayesian results with none of the BFs in experiments 2-6 providing evidence for the alternative hypothesis (no BF $< 1/3$). For three experiments, namely experiments 2, 4 and 5, the BFs provided clear evidence for the null hypothesis (BF > 3) while, for experiments with ambiguous evidence (experiments 3 and 6), RTs were slightly faster for *averted* stimuli rather than for direct stimuli. Of course, further research is necessary to strengthen the absence of direct gaze effects in reduced social stimuli and to further explore the boundary conditions and underlying factors of this absence.

Although we varied our stimulus set along several dimensions, performance for targets on moving stimuli was generally better than on stimuli that remained static. In other words, except for experiment 6, the motion effect remained present across the degree of social information, the degree to which this information was naturalistic, and the degree to which isolated features versus holistic faces were presented. This pattern confirms the finding that sudden onset motion captures attention (Abrams & Christ, 2005) and demonstrates generalizability to a variety of stimuli. In this light, the finding that the motion effect in schematic faces (experiment 5) vanishes when a head-turn movement is introduced (experiment 6) is particularly surprising. One possibility is that motion effects are not as stable as generally assumed. However, considering findings from experiments 1-5 and the vast body of literature on motion effects, including previous studies using the same paradigm (Böckler et al., 2015; Byer et al., 2018), this seems rather unlikely. An alternative explanation is that other attentional or volitional processes cancelled out motion effects in experiment 6. That is, basic or configural aspects of the head-turn stimuli might have prevented them from inducing an apparent motion effect. Critically, findings from Bayes analyses on motion effects in experiment 6 were ambiguous, indicating that further research is necessary to replicate and potentially clarify these (null-) effects.

Conclusions

To conclude, the present results indicate that even though attention capture by direct gaze is triggered by small stimulus facets such as the eyes, it critically depends on facial context information. Direction information that was conveyed by symbolic stimuli, by isolated eyes or even by schematic faces with head-turns identical to those in photographs did not catch attention in a similar manner. This pattern speaks to the idea that social context information has an early effect on attention capture and can modulate our subsequent perception, cognition and interaction (Laidlaw et al., 2011b). Hence, instead of blindly being drawn by gaze wherever we spot it, we may only catch the eye of someone we can potentially interact with.

Acknowledgements

This work was supported by the Emmy Noether Programme of the German Research Foundation (grant number BO 4962/1-1). We thank Jason Tipples, Sara McCrackin and one anonymous reviewer for their constructive suggestions that helped us to strengthen our manuscript.

MANUSCRIPT 2

HOW ECCENTRICITY MODULATES ATTENTION CAPTURE BY DIRECT GAZE AND FACIAL MOTION

Christina Breil[1], Lynn Huestegge[1], Pittig, R.[1], Anne Böckler[2,3]

[1]Julius-Maximilians-University of Würzburg, Würzburg, Germany

[2]Leibniz University Hannover, Hannover, Germany

[3]Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

This manuscript is under review at the *Journal of Vision.*

Abstract

We investigated how effects of direct gaze and facial motion onset depend on presentation location, specifically distance to fixation. Participants responded to targets that were presented on one of four faces that suddenly established direct or averted gaze. Between subjects, faces were presented at different distances to central fixation, spanning 3.3°, 4.3°, 5.5° or 6.5° of the visual field. Replicating previous studies, we found processing advantages for direct gaze and motion onset. Critically, while motion effects remained strong with increasing distance to the center, they were strongly attenuated near center. In contrast, direct gaze effects were powerful at closer eccentricities (3.3° to 5.5°) but expectedly collapsed in the periphery (6.5°). These findings demonstrate how exactly the two distinct processing pathways for peripheral gaze and facial motion cues affect processing, based on functional differences between central and peripheral retinal regions. Moreover, the results further highlight the importance of taking specific stimulus types and properties into account when studying perception and attention in the periphery.

Keywords: direct gaze, motion, social attention, visual attention, spatial attention, stimulus eccentricity

Introduction

Social interactions play a central role in our everyday lives. In order to interpret and react to the utterances and actions of the people around us, we often need to accurately perceive their direction of gaze. In fact, humans seem to master this skill effectively (Emery, 2000; Tomasello & Carpenter, 2007). For example, we reflexively shift our attention in response to averted gaze cues and follow the gaze of others to external locations in an almost automatic fashion, providing the basis for joint attention (Driver et al., 1999; Friesen & Kingstone, 1998; Frischen et al., 2007). On the other hand, another attentional mechanism in response to gaze cues is that we are highly sensitive to other people looking at us (Riechelmann et al., 2019). Eye contact has powerful effects on human attention and modulates subsequent cognitive processes (Kleinke, 1986; Senju & Johnson, 2009b).

However, in real life, direct gaze seldomly occurs in isolation. Instead, gaze is often paired with other cues that affect our attention and perception. For example, when another person raises the head to look at us, direct gaze co-occurs with sudden onset motion, another strong cue for attention capture (Abrams & Christ, 2005). Nevertheless, previous studies showed that direct gaze and sudden onset motion exert their influences independently from each other, suggesting separate underlying attentional channels (Böckler et al., 2015; Böckler, van der Wel, et al., 2014). In the underlying experiment by Böckler et al. (2014), participants were required to detect a target that was randomly presented on one of four faces that gazed either directly at the observer or into the periphery (*gaze*: direct or averted) and had either made a head movement or not (*motion*: sudden or static). Results showed that target classification was facilitated by both direct gaze and motion, but it was most efficient when the two factors coincided in time and space, that is, when the target was presented at the location of sudden motion onset and direct gaze. In contrast, direct gaze still exerted a facilitating effect on target classification, indicating independent sources of influence for gaze and motion cues on attention. This notion was substantiated by further studies with the same paradigm (Böckler et al., 2015; Breil, Huestegge, et al., 2021).

In their experiments, Böckler et al. (2015; 2014) and Breil et al. (2021) paid special attention to holding the location of the stimuli constant by instructing participants to fixate on a central cross, thereby ensuring that all stimuli would appear in the close periphery but within foveal boundaries of the visual field (4.3° visual angle [VA]). This aspect is particularly important because the properties of visual processing shift greatly from center to periphery: Perception is most accurate and contrast-sensitive in the foveal region while the periphery is known to be particularly sensitive to motion cues (Burnat, 2015; Kitterle, 1986; Yu et al., 2010). However, similar to many other attention phenomena, faces are an exception to the rule: They are processed fairly accurately and often better than other objects even when presented at greater eccentricities (Boucart et al., 2016; Hershler et al., 2010), and this advantage is especially pronounced for the encoding of emotional, particularly of happy expressions (Bayle et al., 2011; Calvo et al., 2014; Rigoulot et al., 2012). While other objects are always processed based on single features, both configural and feature-based processing is possible for faces (McKone, 2004).

It has been hypothesized that gaze cues were particularly important for our survival as social beings (Frischen et al., 2007). In line with this idea, it seems relevant that we are able to perceive and react not only to people that are directly in front of us but also to those in our immediate surroundings. Previous studies reported that judgements of gaze direction are reliable up to 4 or 5° (Loomis et al., 2008; Palanica & Itier, 2017) and that attention orienting in response to gaze cues is accurate when the gaze stimulus is presented at 5.0°, but no longer at 7.5° (Yokoyama & Takeda, 2019). Hence, both face recognition and face discrimination suffer from increasing presentation eccentricity. While this dependence of processing accuracy on eccentricity may in part rely on stimulus features such as contrast and size, pure size-scaling cannot equate foveal performance in peripheral presentation (Jebara et al., 2009;

Mäkelä et al., 2001; see Strasburger et al., 2011 for a review). Rather, the superiority in peripheral vision observed for faces (and sometimes also for other objects) depends on the semantic category as well as on the task and its specific spatial scale requirements (Huestegge & Böckler, 2016; Jebara et al., 2009).

Taken together, a general direct gaze advantage for centrally presented faces has been repeatedly demonstrated (for an overview see Senju & Johnson, 2009), but in some circumstances there also appear to be advantages for processing of averted gaze: Riechelmann et al. (2020) found that, for very brief presentations, averted gaze can be detected more easily than direct gaze. One potentially important factor here could be central vs. peripheral processing. In particular, only few is known about the nature of a direct gaze effect, i.e. the effect of direct gaze on attention and perception, at different locations relative to current fixation. Previous findings suggest that attention capture by direct gaze may only hold within foveal boundaries (Loomis et al., 2008; Palanica & Itier, 2014, 2015, 2017; Yokoyama & Takeda, 2019) and due to the decrease in resolution ability in the periphery one would certainly expect a breakdown of the direct gaze effect at some point. The present study is designed to specify these boundary conditions.

In contrast, the detection of motion seems to be preserved even at high eccentricities, most likely due to greater retinal rod density. This is thought to facilitate foveal attention shifts in order to initiate appropriate reactions to potential threats in the environment (Burnat, 2015; Kowler, 2011). Based on this notion, we hypothesize that also facial motion should exert a strong influence on attention allocation even when stimuli are presented more peripherally than in the seminal study by Böckler et al. (2014). However, it seems less clear to what extent one might expect effects of facial motion at smaller eccentricities from current fixation and the present study is designed to address this open issue.

Taken together, previous research suggests that any effects of object eccentricity on attentional processing are to a great extent determined by the type of object, the task at hand,

and the object features to be processed. However, a systematic assessment of eccentricity effects on the processing of facial motion and direct gaze within a single comprehensive study is still lacking. In order to address this research gap, we parametrically manipulated presentation eccentricities in the Böckler et al. (2014, 2015) paradigm and investigated how it shapes the effects of direct gaze and motion onset. As previous research has shown that peripheral processing abilities strongly depend on stimulus type and characteristics (e.g. Huestegge & Böckler, 2016), investigating the boundary conditions of the perception of basic social signals and their integration with context information is an important foundation for a more complete understanding of social attention and behavior. Addressing this matter, we applied the paradigm of Böckler et al. (2015; 2014) and systematically varied the distance between the facial stimuli and the central fixation cross. In particular, we conceptually replicated the experiment with the original distance of 4.3° VA (experiment 2). Furthermore, we either moved the stimuli closer to the central fixation cross (experiment 1: -25%, 3.3°VA) or further away from it (experiment 3: +25%, 5.5°VA; experiment 4: +50%, 6.5°VA; see Figure 2). The central issues were a) at what eccentricity the direct gaze effect might eventually break down, and b) to what extent the sudden facial motion onset effect might be modulated as a function of eccentricity.

Methods

According to recommendations in the context of the open science movement, the method section will focus on the description of sample sizing and selection, all manipulations, all data exclusions (drop-out/outlier procedures) as well as all measures (Simmons et al., 2012).

Participants

A total of 85 participants were randomly assigned to one of the four experiments, of which the data of 82 participants were entered into the final analysis. Data exclusion rates as well as details about age, gender and handedness for each experiment are presented in Table

1. Sensitivity analyses in G*Power 3.1 (Faul et al., 2007b) showed that with our sample size, an effect size of f=0.16 or $\eta^2 \geq 0.025$ can be detected with a statistical power of $1-\beta=.80$ in an ANOVA with an alpha-level of .05. This effect size is comparably smaller than those revealed for direct gaze and motion onset effects in previous studies (e.g. in Böckler et al., 2014). All participants gave written informed consent prior to testing and were compensated monetarily or with course credit. The present study is compliant with the ethical standards of the 1964 Declaration of Helsinki regarding the treatment of human participants and was approved by the local ethics committee.

Table 1

Reaction times (RT), error rates and exclusion rates for experiments 1-4

Experiment (eccentricity)	Total *N* (final *N*)	Mean age (SD) years	Percentage (%) of females	Number of righthanded participants	Excluded due to mean error rate +2SD
1 (3.3°)	22 (21)	26.3 (±5.8)	72.3	22	1
2 (4.3°)	21 (20)	24.1 (±4.6)	90.5	19	1
3 (5.5°)	20 (20)	24.1 (±4.1)	60	17	0
4 (6.5°)	22 (21)	25.4 (±4)	77.3	19	1

Note. Final *N* = number of participants after data exclusion. Mean age, percentage of females and number of righthanded participants prior to data exclusion.

Task and procedure

We applied the experimental design by Böckler et al. (2015; 2014) and varied the distance of the stimuli from the central fixation cross between groups. Participants were placed 80cm away from a 69-cm diagonal (59.77cm × 33.62cm) monitor (screen resolution:

81.6 dpi; 1920 × 1080 pixels) with their two index fingers placed on the response keys ("S" and "H") of a standard Qwertz keyboard and their head fixated in a chin rest. Each trial began with four images of the same female face (each face 139 × 171pixels, 3 × 3.7°VA), two with direct gaze and two with their gaze averted by 45°, located around a central fixation cross (Figure 1). Small figure-8 symbols (0.3 × 0.5°VA) were positioned on the forehead of each face (2.2 × 3.4°VA) and replaced by one target ("S"/"H") and three distractor letters ("E"/"U") after 1500 ms. Simultaneous to target presentation, one direct face changed to averted and one averted face changed to direct (inducing apparent motion, e.g. Wertheimer, 1912) while the other two faces remained unchanged (no motion). There was a total of 384 trials in each experiment, with 96 trials per condition. The identity and position of target and distraction letters appeared equally often in all possible combinations. Prior to and during the task, participants were explicitly instructed to fixate on the central cross throughout the experiment and to respond as quickly and accurately as possible to the target letter by pressing the respective response key. Hence, paying attention to gaze direction or motion was not necessary for task completion. Importantly, while the factors gaze and motion were varied randomly within participants, the distance of the four faces to the central fixation cross changed between experiments. Hence, for any participant, faces were presented at the same distance throughout the experiment. For experiment 2, the distance from the inner edge of target letter to the center of the screen (DTC) was the same as in the previous studies from our research group (4.3° VA, see Böckler et al., (2015; 2014; Breil, Huestegge, et al., 2021), whereas in the other experiments, the distance decreased by 25% (3.3° VA; experiment 1), or increased by 25% (5.5° VA; experiment 3) or 50% (6.5° VA; experiment 4; Figure 2). In each experiment, all possible combinations of position and identity of target and distraction letter as well as of gaze direction and motion appeared equally often across all 384 trials. PsychoPy (psychopy.org) was used for stimulus presentation and response recording. R toolboxes were used for data formatting and data analyses.

Data analysis

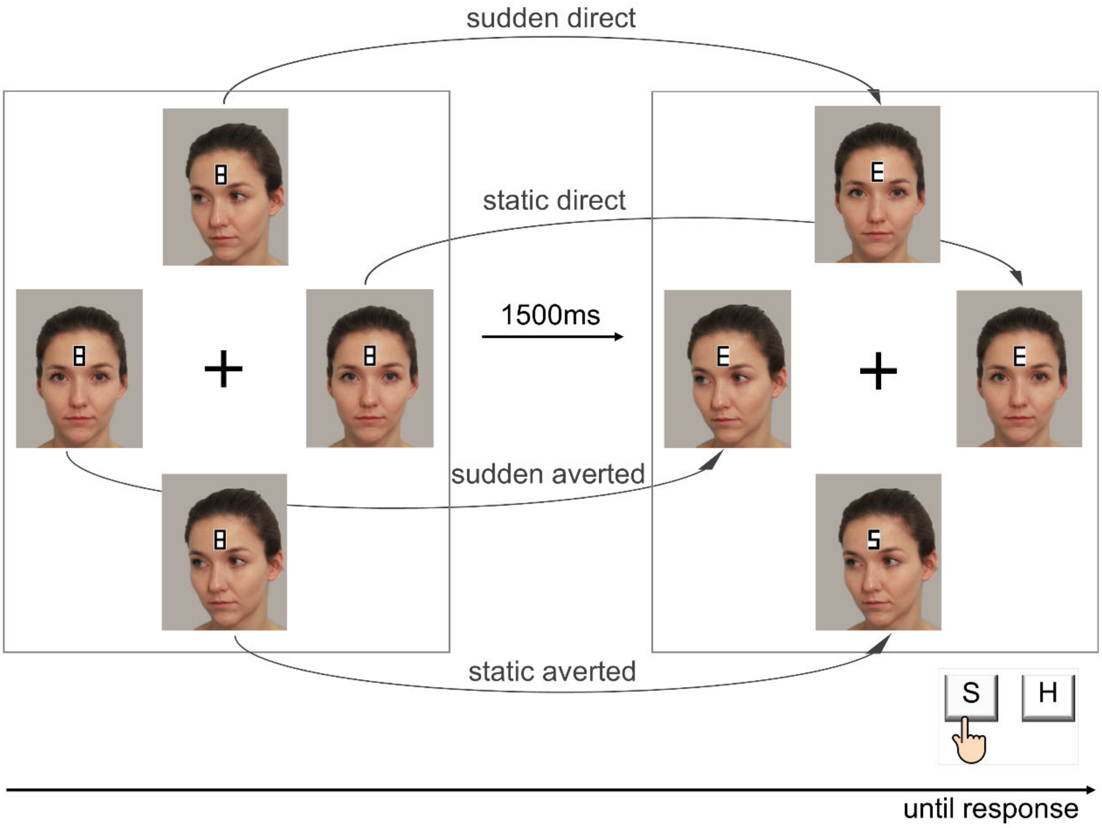
We defined reaction time (RT) as the time window from target onset until key press. For data trimming, participants with error rates +2SD of the global mean were removed. This was the case for 1 participant in each of experiments 1, 2 and 4. RTs of error trials and RTs exceeding ±2SD of the participant's mean in each condition were excluded from further analysis. Table 2 contains the exclusion rates for each experiment.

For each experiment, the remaining data sets were submitted to two 2×2 repeated measures ANOVAs with the within-subject factors gaze (direct, averted) and motion (static, sudden), using mean RTs and error rates as dependent variables. Note that the factor gaze direction refers to the orientation of the face when the target/distractor display was presented, not to the orientation presented at the beginning of the trial. We applied two-sided t-tests with Bonferroni correction for multiple testing to resolve interaction effects.

To confirm null results, we additionally drew on Bayesian statistics. Non-directional Bayes factors (BFs) were calculated as f (data | H_0) / f (data | H_1) with a prior distribution value of 1 for all non-significant main effects of gaze and motion on RTs of experiments 1-4. Significant interaction effects were further assessed with Bayesian analyses of variance and subsequent calculation of BF in order to compare the model containing the interaction term to the model restricted to main effects. For BF > 3, we interpreted BFs as evidence for the null hypothesis and BF < 1/3 as evidence for the alternative hypothesis (Rouder et al., 2009).

Figure 1

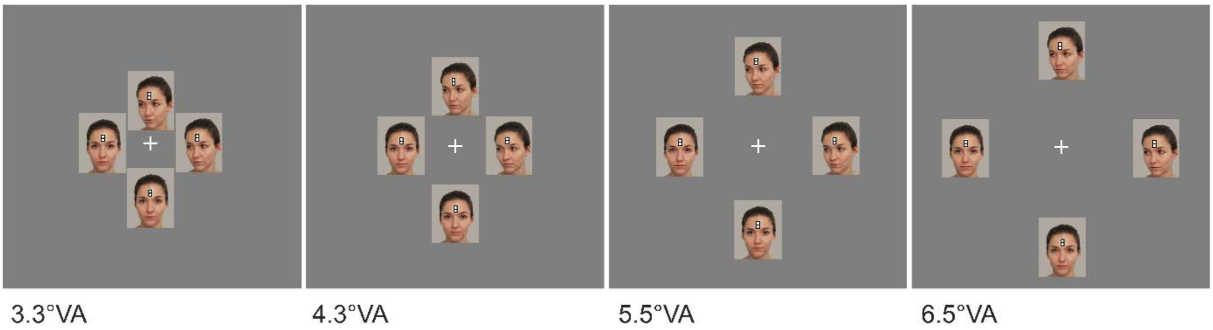
Sample trial sequence of experiment 2



Note. At the beginning of each trial, number-8-figures were presented at the forehead of all four faces. After 1500 ms, they were replaced by one target and three distractor letters while, at the same time, one direct stimulus changed to averted and one averted stimulus changed to direct. The other two stimuli remained unchanged, resulting in the four experimental conditions sudden direct, static direct, sudden averted, static averted. Participants were required to react to the target letter by pressing the corresponding button on the keyboard as fast and accurate as possible. The original distance of 4.3°VA from Böckler et al. (2014) was adopted for experiment 2 as visualized in Figure 1 and was varied for the remaining experiments as shown in Figure 2. All parts of this figure are original. Photographs were taken by the authors and printed with permission.

Figure 2

Distances of stimuli to center for experiments 1-4



Note. The original distance of 4.3°VA from Böckler et al. (2014) was applied to experiment 2 and decreased by 25% for experiment 1 (3.3°VA), or increased by 25% or 50% for experiment 3 (5.5°VA) and 4 (6.5°VA), respectively. Please note that the figure is an approximate depiction and does not show the size of face stimuli in relation to the real screen size. All parts of this figure are original. Photographs were taken by the authors and printed with permission.

Results

The data sets generated in this study are available in the Open Science Framework (DOI: 10.17605/OSF.IO/FXTE7). Table 2 contains the mean RTs and error rates for each experiment. Figure 3 visualizes the direct gaze and sudden onset motion effects by DTC, calculated as the mean RT for direct gaze minus for averted gaze stimuli and the mean RT for sudden minus for static stimuli, respectively.

Experiment 1: 3.3°VA

The mean RT was 903 ms (SD = 253). The main effect of gaze was significant with faster responses to faces that depicted direct gaze ($F(1,20) = 11.71$, $p = .003$, $\eta^2_p = .37$). In contrast, the main effect of motion was not significant ($p = .696$, $BF_{01} = 5.56$), nor was the gaze × motion interaction ($p = .827$).

Table 2

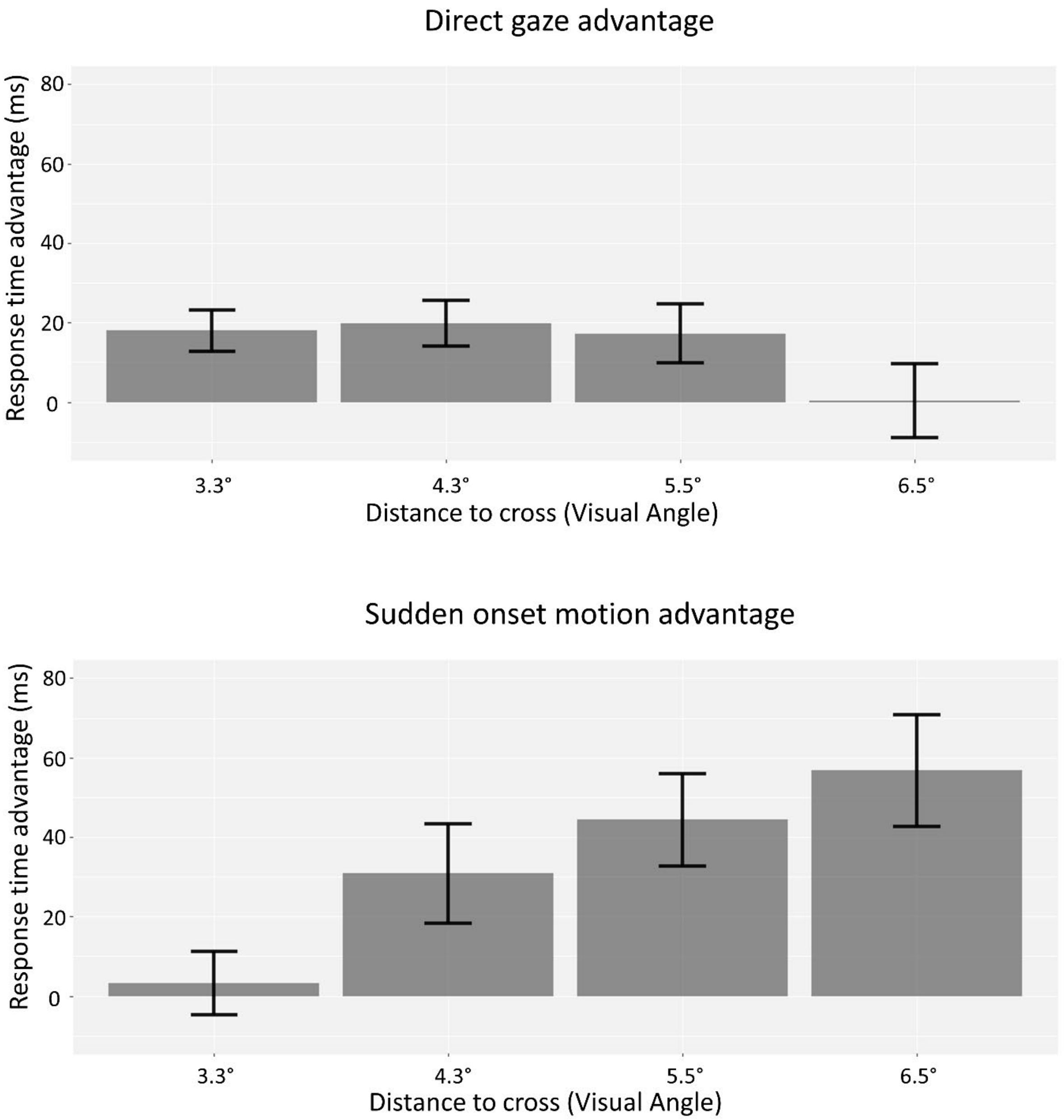
Reaction times (RT), error rates and exclusion rates for experiments 1-4

Experiment	Condition										Exclusion rate
			sudden direct		static direct		sudden averted		static averted		
	RT (SE)	ER (SE)	RT (SE)	ER (SE)	RT (SE)	ER (SE)	RT (SE)	ER (SE)	RT (SE)	ER (SE)	
1 (3.3°)	903 (±0.002)	4.2 (±0.002)	893 (±22)	2.53 (±.67)	895 (±23)	1.74 (±.38)	910 (±22)	2.18 (±.48)	914 (±22)	2.33 (±.64)	8.12
2 (4.3°)	1005 (±0.003)	4.7 (±0.002)	982 (±28)	3.18 (±.49)	1037 (±31)	2.93 (±.50)	1016 (±27)	4.58 (±.60)	1033 (±28)	4.58 (±.73)	8.02
3 (5.5°)	1083 (±0.004)	3.3 (±0.002)	1054 (±28)	3.39 (±.43)	1093 (±31)	2.60 (±.47)	1066 (±27)	3.17 (.50)	1116 (±29)	3.85 (±.44)	6.41
4 (6.5°)	1131 (±0.004)	3.9 (±0.002)	1103 (20)	3.22 (±.82)	1155 (±29)	2.68 (±.54)	1099 (±22)	2.98 (±.53)	1160 (±33)	4.17 (±1.01)	7.14

Note. RT = mean reaction times, ER = mean error rates, SE = standard error. Mean RTs of correct responses in ms; error and exclusion rates in %.

Figure 3

Direct gaze and sudden onset motion advantage effects across distances to the center



Note. The direct gaze advantage is calculated as the mean RT for averted gaze minus the mean RT for gaze gaze stimuli. The sudden onset motion advantage is calculated as the mean RT for static minus the mean RT for sudden stimuli. Hence, positive values indicate a direct gaze or a sudden onset motion advantage, respectively. Experiment 1 = 3.3°VA, experiment 2 = 4.3°VA, experiment 3 = 5.5°VA, experiment 4 = 6.5°VA. Error bars represent standard errors. All parts of this figure are original.

Error analyses yielded no significant effects (gaze: $p = .666$; motion: $p = .212$; gaze × motion: $p = .245$). Hence, the interpretation of our RT results is not compromised by any speed-accuracy trade-offs.

Experiment 2: 4.3°VA

On average, participants responded after 1005 ms (SD = 293) in experiment 2. There were significant main effects of gaze ($F(1,19) = 12.05$, $p = .003$, $\eta^2_p = .39$) and motion ($F(1,19) = 6.15$, $p = .023$, $\eta^2_p = .24$) due to faster responses to faces that depicted direct gaze and to moving stimuli. The gaze × motion interaction effect was significant ($F(1,19) = 5.423$, $p = .031$, $\eta^2_p = .22$) because RTs were faster for direct gaze than averted gaze with moving stimuli ($t(19) = -3.967$, $p < .001$) but not with static stimuli ($p = .916$). An additional Bayesian analysis of variance with subsequent likelihood comparison yielded $BF_{10} = 0.379$ for the model containing the interaction term.

Error analyses revealed a significant effect of gaze with higher accuracy for direct gaze stimuli ($F(1,19) = 16.6$, $p = .001$, $\eta^2_p = .47$), but no other significant effects (motion: $p = .719$; gaze × motion: $p = .605$). As participants responded both faster and more accurate to direct compared to averted gaze, these findings give no indication of speed-accuracy trade-offs.

Experiment 3: 5.5°VA

The average response time in experiment 3 was 1083 ms (SD = 337) with main effects of gaze ($F(1,19) = 5.43$, $p = .031$, $\eta^2_p = .22$) and motion ($F(1,19) = 14.61$, $p = .001$, $\eta^2_p = .43$) but no interaction effect ($p = .532$). As in experiment 2, participants reacted faster to direct gaze and to moving stimuli.

The main effects of gaze and motion on error rate were not significant (gaze: $p = .171$; motion: $p = .867$), indicating that our findings were not due to a speed-accuracy trade-off. The gaze × motion interaction was significant ($F(1,19) = 5.29$, $p = .033$, $\eta^2_p = .22$) due to higher accuracy for direct gaze than to averted gaze stimuli that were static ($t(19) = -3.093$, $p = .01$) but no difference in the sudden condition ($p > .999$). Bayesian analysis of variance with comparison of likelihoods revealed $BF_{10} = 2.237$ for the model that contained an interaction term.

Experiment 4: 6.5°VA

At the greatest eccentricity, participants produced responses after a mean time of 1131 ms (SD=324). Responses were facilitated by sudden onset motion, as evident by a significant main effect of motion ($F(1,20) = 16.15$, $p < .001$, $\eta^2_p = .45$). In contrast, there was no significant main effect of gaze ($p = .975$; $BF_{01} = 14.16$). The interaction effect on RTs was not significant ($p = .666$) nor were any effects of the analyses on error rates (gaze: $p = .178$; motion: $p = .301$; gaze × motion: $p = .121$).

Discussion

The present study systematically manipulated presentation eccentricity of human faces and addressed its effect on attention capture by direct gaze and sudden onset facial motion. In four experiments, photographic faces were presented at different distances to a central fixation, spanning 3.3°, 4.3°, 5.5° and 6.5° of the visual field while stimulus size was held constant. As expected based on the lower spatial resolution abilities in the periphery, we found that the direct gaze effect was present across the lower distances but decreased as a function of eccentricity. Crucially, we determined that this effect was relatively stable until 5.5° of eccentricity, but completely collapsed at 6.5° (experiment 4). In contrast, the motion effect was present at medium and greater eccentricities, but absent at the smallest angle (3.3°; experiment 1), confirming the finding of a higher motion sensitivity in the periphery based on rod density. These findings replicate and extend previous research suggesting that attention capture by direct gaze and by sudden facial onset motion are based on parallel but independent processing channels (Böckler, van der Wel, et al., 2014) and are in line with the notion of a functional specialization of foveal and peripheral vision.

Across all gaze and motion conditions, we found a processing benefit for stimuli that were presented more centrally. The finding of a general distance effect is not surprising as it has been reported repeatedly and reliably in previous studies (e.g. Carrasco et al., 1998; Gruber et al., 2014). As a baseline, we replicated the finding that direct gaze presented at 4.3° eccentricity captures attention (Böckler, van der Wel, et al., 2014; Boyer & Wang, 2018, 2018; Breil, Huestegge, et al., 2021). We furthermore showed that this effect could still be observed when the original distance was varied by ±25% (3.3°, experiment 1; 5.5°, experiment 3, respectively), providing further evidence for the generalizability of the direct gaze effect across different eccentricities. Critically, when stimuli were presented beyond foveal limits (at 6.5°, experiment 4), the direct gaze effect collapsed. This pattern fits related lines of research that find gaze cueing effects when the stimulus is presented at 5.0°, but no longer at 7.5° (Yokoyama & Takeda, 2019) and that judgements of gaze direction are reliable up to 4 or 5° (Loomis et al., 2008; Palanica & Itier, 2017). In contrast, head orientation is still accurately perceived at much higher eccentricities, such as 90°(Loomis et al., 2008). Taken together, these findings emphasize the importance of spatial resolution abilities in the periphery for attention capture by direct gaze and enables us to quantify its reach: Beyond an angle of 5.5°, direct gaze appears to be no longer salient enough to capture attention.

At ±3° eccentricity, head orientation starts to bias perceptions of gaze direction and this bias increases along with eccentricity (Palanica & Itier, 2014). With respect to our paradigm, this interesting finding gives rise to the question of whether the direct gaze effect would still hold when gaze and head direction diverge. Future studies should independently manipulate gaze direction and head orientation to investigate modulation of the direct gaze effect by head orientation. Ideally, this should be probed at different eccentricities, such as in the present study and augmented with eye tracking to ensure central fixation. Another aspect that could be addressed in future studies involves the central-to-peripheral compression of projections from the retina to the striate cortex (cortical magnification factor). In order to unequivocally assess the influence of eccentricity on the effects of direct gaze and motion on attention, we held the angular size of stimuli constant across all experiments of this study. However, perceptual acuity decreases from central to peripheral vision and the area of the visual cortex that is filled by a stimulus gets smaller the further away from the fovea it is presented (Anstis, 1998). Hence, constant visual acuity can be achieved by increasing the target size according to perceptual eccentricity (Anstis, 1974). We found in our study that, when physical stimulus size is invariant, attention capture by direct gaze behaves in a step-like rather than a gradual way across the central-to-peripheral gradient of the visual field. Future studies could investigate whether increased stimulus size can compensate for large eccentricity. For example, it could be tested whether the direct gaze effect can be preserved in the further periphery by parametrically manipulating the angular size of stimuli according to the distance to the center.

In contrast to our findings regarding the direct gaze effect, the motion effect was powerful at the higher eccentricities (4.3-6.5°), but absent close to the center (3.3°). This absence of any statistically significant effect of sudden facial motion onset at 3.3° may come as a surprise. The efficient detection of sudden movements in the periphery is an important mechanism to redirect foveal attention that could help to facilitate quick reactions to objects and events in the environment (Burnat, 2015; Kowler, 2011). This notion has been substantiated by the finding that motion-sensitive areas of the dorsal stream respond to attentional enhancement in the periphery (Bressler et al., 2013) and that the peripheral visual field is specifically tuned to high velocities (Kitterle, 1986). However, it has been noted before that there is no evidence that the periphery is better capable of detecting motion than the central visual field (McKee & Nakayama, 1984; To et al., 2011). In this light, our findings may speak to the idea that, rather than capturing attention only when presented at higher eccentricities, motion cues are simply less salient than gaze cues when both are presented near center. Close to fixation, processing of facial features that depend on highly resolved spatial processing may override processing of even quite salient dynamic features based on motion. It should be noted, however, that absolute and general statements on sensitivity thresholds are not necessarily conclusive. Previous research suggests an intricate modulation of foveal and peripheral sensitivity to motion by context effects. In particular, the relative thresholds of motion perception depend on the task at hand and on parameters of the stimuli, such as luminance, contrast, size and velocity (see Sekuler et al., 2002 for an overview). Albeit the present study provides promising initial findings, future studies are certainly needed in order to systematically assess the boundary conditions of attention capture by motion cues by investigating various stimulus materials at a broad range of eccentricities.

Unexpectedly, we found two significant interaction effects: At 4.3° eccentricity (experiment 2), responses to sudden direct stimuli were faster compared to sudden averted stimuli, whereas no significant difference was found for static stimuli. Furthermore, at 5.5° (experiment 3), participants produced more errors in response to static averted compared to static direct stimuli while error rates were consistent in the sudden condition. These findings contradict results from earlier experiments inside and outside of our group that have never reported interaction effects between gaze and motion cues with this paradigm, particularly at 4.3° (Böckler et al., 2015; Böckler, van der Wel, et al., 2014; Boyer & Wang, 2018), thereby emphasizing the independency of processing of gaze cues and sudden facial motion. Why exactly our results are different from these earlier findings cannot be fully explained with the present data. We did not find any consistent pattern of interactions across the various eccentricities, which leads us to the assumption that these results probably reflect some random, unsystematic fluctuations in the data that we do not want to overinterpret at this point. This notion is supported by Bayesian analyses of variance, giving no clear evidence for reliable interaction effects in these experiments (all $BF_{10} > 1/3$). Nevertheless, further research is certainly necessary to investigate the reliability of these particular interaction effects.

Taken together, our findings corroborate the expected dissociation of gaze and motion cues, in line with the well-established functional specialization of central and peripheral vision. In previous studies, fMRI responses to visual stimuli indicated eccentricity-dependent effects of attention (Bressler et al., 2013; Roberts et al., 2007). In particular, central vision has been linked to attentional enhancement of early visual, ventral and lateral occipital cortical areas, indicating a role of endogenous attention for fine detail perception. In contrast, responses in areas of the dorsal cortical network were enhanced by attention to peripheral objects. The dorsal stream is particularly involved in action and object allocation and comprises areas that are important for subsequent eye position and attention shifts (Bressler et al., 2013; Corbetta & Shulman, 2002). Our present findings fit this line of research by showing that effects of attention capture by direct gaze and motion cues both depend on eccentricity, but in markedly different ways. Eye gaze is thought to be one of the most important cues during social interactions (see Emery, 2000; Kleinke, 1986 for reviews). The accurate perception of gaze direction is crucial for perception of emotion and attention (McCrackin & Itier, 2019), to understand what others are thinking (Baron-Cohen & Cross, 1992b), and to communicate with them (Cook, 1977). When interacting with another person, that person is usually placed in front of us, hence centrally fixated and processed with high precision (via the ventral pathway). Even though in these situations our main focus of attention lays on the person that we are facing, it is crucial that we can become aware of unexpected events in our peripheral environment in order to quickly initiate an adequate action (via the dorsal pathway). In other words, we attend to direct gaze with central vision while reacting to sudden onset motion in the periphery.

Limitations

We are aware that the interpretation of null effects is difficult and that, in the present study, it is further limited by the sample size (e.g., Brysbaert, 2019). Note, however, that the present data support our conclusions in multiple ways. For one, numerical differences between the respective conditions were negligible. In particular, the difference between RTs to sudden versus static stimuli in experiment 1 amounted to 3 ms and responses to direct versus averted gaze in experiment 4 were virtually identical, with a difference in RTs of only 0.5 ms. This equivalence in RTs is further supported by Bayesian statistics. The results of this additional analysis were clearly favoring the null hypothesis for the motion effect in experiment 1 and the gaze effect in experiment 4, respectively. Nonetheless, replication studies with larger samples are needed before stronger conclusions can be drawn.

Conclusion

In sum, our results corroborate and extend previous findings regarding the processing of facial gaze and motion cues in the periphery (Böckler et al., 2015; Böckler, van der Wel, et al., 2014; Boyer & Wang, 2018; Breil, Huestegge, et al., 2021). In particular, we were able to specify how exactly attention is captured by direct gaze and sudden facial motion onset at various eccentricities. On the one hand, our data reflected the expected difference between motion cues, which can easily be resolved in the periphery, and direct gaze information, which strongly depends on spatial resolution abilities in the periphery. On the other hand, we were able to specify the spatial extent of the direct gaze effect and to show a substantial attenuation of the motion capture effect near the center of fixation. On a more general level, our results further emphasize the importance of taking specific stimulus types and properties into account when studying perception and attention in the periphery. In sum, the current study highlights the sophisticated adaption of the human attentional and visual system, allowing us to flexibly and efficiently navigate in our (social) environment.

Acknowledgements

This work was supported by the Emmy Noether Programme of the German Research Foundation (grant number BO 4962/1-1).

MANUSCRIPT 3

DON'T LOOK AT ME LIKE THAT: INTEGRATION OF GAZE DIRECTION AND FACIAL EXPRESSION

Christina Breil[1], Tim Raettig[1], Roxana Pittig[1], Robrecht P.R.D. van der Wel[3], Timothy Welsh[4], Anne Böckler[1,2]

[1]Department of Psychology, Julius-Maximilians-University of Wuerzburg, Germany

[2]Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[3]Department of Psychology, Rutgers University

[4]Faculty of Kinesiology & Physical Education, University of Toronto

This manuscript has been submitted to the *Journal of Experimental Psychology: Human Perception and Performance.*

Abstract

Efficient decoding of facial expressions and gaze direction supports reactions to social environments. Although both cues are processed fast and accurately, when and how these cues are integrated is still debated. The present study investigated the temporal integration of gaze and emotion cues by concurrently manipulating them in a target detection task. Participants responded to letters that were randomly presented on one of four faces that started neutral but adopted either an approach- or an avoidance-oriented emotion expression (Experiment 1a: angry/fearful; Experiment 1b: happy/disgusted). Further, two faces initially showed direct gaze and two initially showed averted gaze. Simultaneous with target presentation, two faces changed gaze direction (from averted to direct and vice versa). Although angry and fearful expressions diminished any effects of gaze direction (Experiment 1a), a direct gaze advantage was found for happy faces and an averted gaze advantage for disgusted faces (Experiment 1b). In Experiment 2, we tracked eye movements in addition to performance and, again, found evidence for an approach/avoidance-congruency advantage for happy and disgusted faces both in performance and gaze behavior. Analyses of gaze behavior suggested early integration of gaze and emotion information (starting at 300 ms) that is driven by emotion-specific attention patterns.

Public significance statement

This study highlights flexible and emotion-specific temporal aspects of gaze and face information integration processes that are involved in social perception and attention. The findings strengthen the approach/avoidance congruency hypothesis with an integration of gaze and face information at early processing stages.

Keywords: Social cognition, visual attention, gaze, facial expressions, emotions

Introduction

As social beings, people naturally pay attention to the faces of others. Prioritized decoding of facial signals is crucial for people to successfully navigate their social environment and to react to others around them in an appropriate way (Hessels, 2020). One of the most meaningful cues in this context is the direction of another person's gaze (Böckler, van der Wel, et al., 2014; Emery, 2000; Kleinke, 1986; Schilbach, 2015b; Tomasello & Carpenter, 2007). Without effort, people detect the gaze of others, follow these gaze cues to external locations and process the objects at these gazed-at locations with greater efficiency than objects at non-gazed-at locations (Driver et al., 1999; Farroni et al., 2004; Friesen et al., 2004; Frischen et al., 2007; Pfeiffer et al., 2013; Ristic et al., 2002). Another salient cue is when another person's gaze is directed at the individual. Direct gaze has profound effects on the subsequent processing of (social) information, and it is a critical determinant of human behavior (Conty et al., 2016; Emery, 2000; J. K. Hietanen, 2018; Senju & Hasegawa, 2005; Senju & Johnson, 2009b; Stein et al., 2011). Investigating the effects of direct gaze and the integration of this cue with other (social) cues is key to a deeper understanding of how social information is conveyed and processed, for instance. One such important cue is the facial expression.

In recent years, researchers in various fields of cognition have dedicated effort to understanding the influences of gaze direction on the perception of facial expressions. For instance, a range of studies indicate that angry faces are processed more efficiently when the gaze of these faces is oriented towards the observer, whereas fearful faces are processed more efficiently when the faces look into the periphery (Adams & Kleck, 2003, 2005; N'Diaye et al., 2009; Sander et al., 2007). Researchers have reasoned that the perception of emotional signals is enhanced when the signals match the expresser's approach-avoidance behavioral tendency (Adams & Kleck, 2003, 2005). Such congruency between gaze direction and facial expression could provide evolutionarily meaningful information by helping the observer to detect threats in the environment, thus enabling quick and appropriate reactions. For instance, an angry person looking at another person can efficiently signal that the observer needs to flee or fight. Similarly, the averted gaze of a fearful person may signal that the observer should look away from the individual expressing fear and towards the threat (e.g., the tiger hiding in the bushes).

The traditional model from Adams and Kleck (2003, 2005) assumes an integration of gaze direction and emotion expression at early processing stages (*early integration*). Other researchers, however, found a generally faster recognition performance for faces with direct gaze without modulatory effects (Bindemann et al., 2008). It was hence assumed that the model by Adam's and Kleck could account for the integration of gaze direction and emotion expression at later processing stages (*late integration*). Concluding evidence for these two possibilities is lacking, and recent studies suggest a more flexible and context-dependent interaction for the processing of gaze direction and facial expression that crucially depends on the task instructions and the relevance of the face stimulus for the observer (Ricciardelli et al., 2016). This idea is supported by event-related potential (ERP) and imaging studies suggesting that top-down influences and stimulus-driven effects interact in the deployment of attentional resources when processing faces, and that this interplay flexibly varies across the visual processing stream (Haxby & Gobbini, 2011; Schindler & Bublatzky, 2020). For example, task-driven prioritization (top-down) and emotional stimulus content (bottom-up) have been found to interact in visual word processing, and the nature of interaction changed across processing stages, suggesting shared and distinct mechanisms of emotion and cognition (Schindler & Kissler, 2016). Importantly, the time course and location of neural activation during facial perception are emotion-specific, indicating that different emotions engage distinct brain areas at different time points in visual processing (Itier & Neath-Tavares, 2017; Neath & Itier, 2015; Neath-Tavares & Itier, 2016; Schupp et al., 2004). Note that, to the best of our knowledge, no eye tracking study has yet examined this question.

Currently, the precise temporal dynamics of the integration of gaze direction and facial expression are still a matter of debate. To better understand this process, the present study employed a modified version of a target detection task from Böckler et al. (2015; 2014). In this task, each trial started with a display of four faces and a fixation cross, with two faces gazing directly at the observer and two faces gazing into the periphery. Simultaneous with target and distractor presentation on the foreheads of the faces, two faces remained static, while one direct face changed to averted and one averted face changed to direct. This approach yields a 2×2 factorial design with gaze direction (direct, averted) and motion (present, absent) as within-subject factors. Typically, target detection is facilitated by direct gaze and by sudden onset motion with no interaction between the two effects (e.g. Böckler et al., 2014, 2015; Boyer & Wang, 2018). Critically, previous studies used neutral faces throughout each trial. In the present study, however, each trial started with four neutral faces that changed their facial expression to a specific emotion at the exact same time that gaze direction changed and targets were presented. This set-up enabled the exploration of whether or not the direct gaze effect is modulated by simultaneously appearing emotion expression, which would speak to an integration of both cues at early processing stages. In contrast, if the data reveal that prioritization by direct gaze is not different across all emotions, then one could reason that emotion expression is not integrated with gaze information at early processing stages when these cues appear simultaneously. Experiments 1a and 1b of the present study address this question, with each experiment presenting one approach-oriented emotion (Experiment 1a: angry; Experiment 1b: happy) and one avoidance-oriented emotion (Experiment 1a: fearful, Experiment 1b: disgusted). Each of these two emotions appeared in half of the 364 trials, with the order of trials being fully randomized. If emotion expressions influence the processing of gaze cues at an early stage, we would expect to find a difference in RTs to the targets on faces displaying an approach- or an avoidance-oriented emotion. Specifically, the direct gaze advantage should be present for approach-oriented expressions (happy, angry) and reduced or even reversed in avoidance-oriented expressions (disgusted, fearful; see Adams & Kleck, 2005). If, on the contrary, these cues are not integrated at an early stage, there should be no difference in the pattern of RTs across emotion conditions.

To further investigate the time course of integration, Experiment 2 employed the same design and tracked participants' gaze behavior. This eye tracking was done only for the specific emotion expressions for which a modulation of the gaze effect was found in Experiment 1, namely happy/disgusted. We collected markers of gaze behavior and additionally explored fixation patterns over time to identify the window in which gaze direction is particularly salient and whether this period changes as a function of emotion expression. If processing of gaze cues is influenced by emotion expression at very early stages, differences in visual fixation patterns to approach- compared to avoidance-oriented expression should be observed shortly after target presentation. However, if the integration takes place at later stages, this should be reflected in differences between fixation patterns that occur at later time windows.

Experiment 1

Methods

For all experiments, we report how the sample sizes were determined. We also report all manipulations and measures that were collected, and all data exclusions (Simmons et al., 2012).

Participants

The number of participants was determined using G*power3 (Faul et al., 2007b), assuming 80% power and an α of .05 with a medium effect size based on previous studies (Böckler, van der Wel, et al., 2014). Overall, 72 participants were tested and randomly assigned to one of the two emotion expression combinations (Experiment 1a: angry/fearful; Experiment 1b: happy/disgusted). This procedure resulted in 38 participants in Experiment 1a (angry/fearful). After participant-wise data exclusions due to unusually high error rates (+2SD above the global mean), the final analysis included 32 participants (27 females, 31 right-handed; mean age = 23.1±3 years). There were 34 participants that completed Experiment 1b (happy/disgusted). After data exclusion, the final analysis included 31 participants (26 females, 29 right-handed; mean age = 23.1±2.92 years). Sensitivity analyses with G*power suggest that we can detect an effect size of f=.10 with this sample size for a 2x2 interaction in a within-subject ANOVA. All participants gave written informed consent and were compensated with 7€ or course credit. The study complied with the ethical standards of the 1964 Declaration of Helsinki regarding the treatment of human participants and was approved by the local ethics committee.

Stimuli

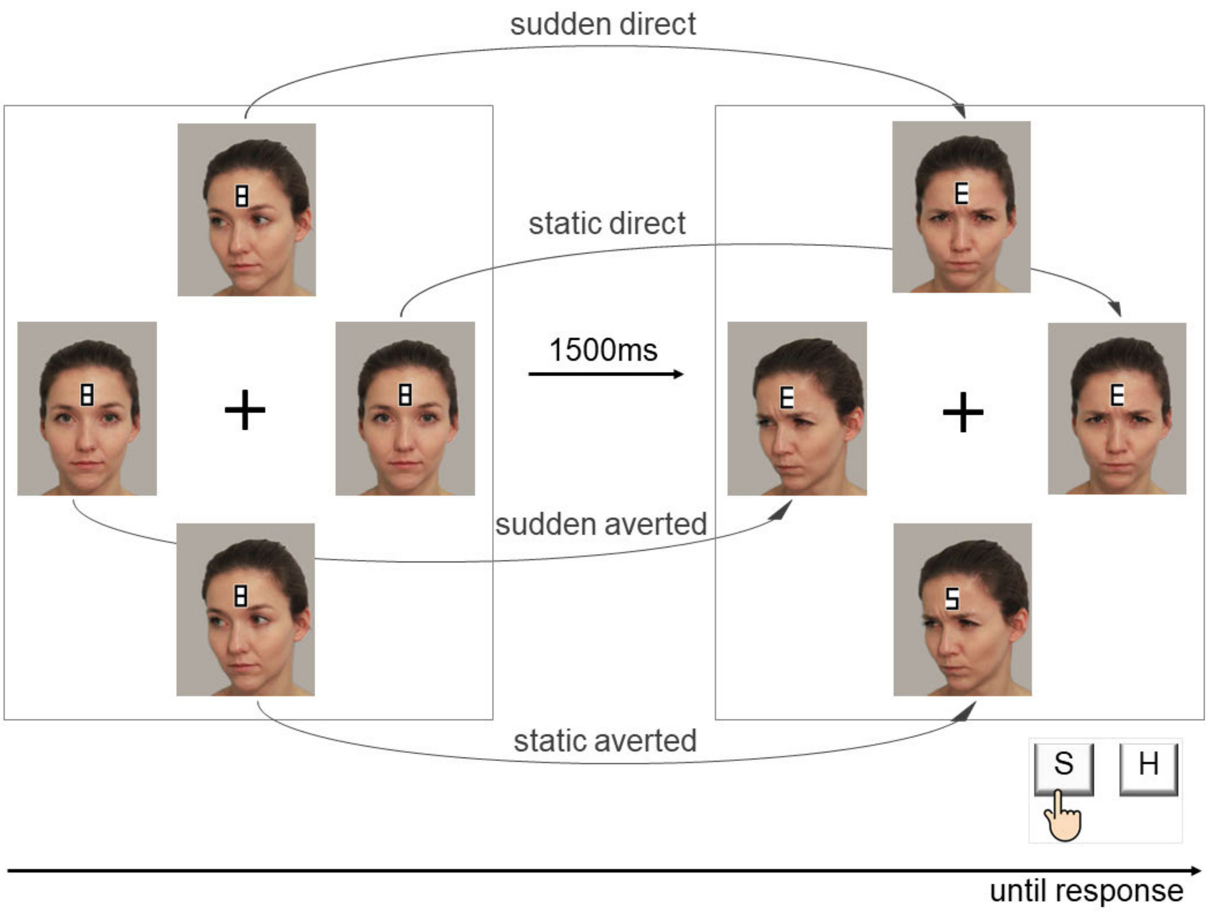
All stimuli depicted the same person with different facial expressions, while adopting either direct head and gaze orientations or head deviations of 30° with averted gaze. These stimuli were validated in 50 participants (38 females, mean age = 26.4±8.3 years), with high rates of correct classifications as indicated by binomial tests (proportions ≥ .82; all *p*s < .001; neutral: 97%; angry: 95%; fearful: 84%; happy: 100%; disgusted: 95%). Observed patterns of confusion between two emotion expressions (particularly between fearful and disgusted) were not relevant for the present study and are hence not discussed in detail (but see Calvo & Nummenmaa, 2016 for a review with matching results).

Task and procedure

The experimental design was based on previous studies by Böckler et al. (2015; 2014). Participants sat 80cm away from a 43-cm TFT monitor and instructed to place their right and left index fingers on the response keys (“S” and “H”) on a standard keyboard. In each trial, participants saw two subsequent displays, each consisting of four images of the same face (200×250pixels, 3.8×4.7 visual angle (VA)) positioned around a central fixation cross (see Figure 1). In the first display, all faces had a neutral expression, with two of the faces depicting direct gaze and the other two with averted gaze. Each face had small figures of the number “8” positioned on their forehead. After 1500 ms, three important changes happened simultaneously. First, 2 line segments disappeared so that the number-8 figures were replaced by three distractor letters (“E”/”U”) and one target letter (“S”/“H”) to which the participants were required to respond by key press. Second, all four faces changed their expression from neutral to the same emotion (angry or fearful in Experiment 1a; happy or disgusted in Experiment 1b). Finally, two of the faces changed their orientation: one direct face suddenly changed to averted gaze (*sudden-averted*) and one averted face suddenly changed to direct gaze (*sudden-direct*). The other two faces remained static (*static-direct; static-averted*). A sample trial sequence is displayed in Figure 1.

Figure 1

Example of a Sudden-angry Trial Sequence in Experiment 1a



Note. Each approach- and avoidance-oriented emotion expression was presented in a random order in 50% of the trials in each experiment. Experiment 1a: angry/fearful; Experiment 1b and 2: happy/disgusted. Facial expressions were always neutral in display 1 and changed to an emotion expression after 1500 ms. Simultaneously, the target (letter “S” or “H”) appeared randomly on one of the four faces, while two faces changed direction (one form averted to direct and one form direct to averted), inducing apparent motion.

In Experiment 1a, 50% of the total of 384 trials involved *neutral-to-angry* changes of emotion expression and 50% were *neutral-to-fearful* trials. In Experiment 1b, the same ratio was applied to *neutral-to-happy* versus *neutral-to-disgusted* trials. Emotion expressions were presented in randomized order in each experiment and appeared equally often in all possible combinations of gaze direction, motion, target identity (S/H), distractor identity (E/U) and location of the target and distractor letters.

Data analysis

Response time (RT) was defined as the time window from target onset until one of the response keys was pressed. All RTs of response error trials (incorrectly identifying “S” when the correct target was “H”, and vice versa) were excluded from further analysis, as were RTs exceeding ±2SD of the participant’s mean in each condition. Error and exclusion rates as well as mean RTs for each condition and emotion are displayed in Table 1.

For each of Experiments 1a and 1b, the remaining RTs and error rates were analyzed by conducting two 2×2×2 repeated measures ANOVAs with the within-subject factors gaze direction (direct, averted), motion (static, sudden) and emotion expression (Experiment 1a: angry, fearful; Experiment 1b: happy, disgusted). Two-tailed *t*-tests with Bonferroni correction for multiple testing were applied to resolve interaction effects. Partial η^2 and Cohen’s *dz* are reported as measures of effect size.

Table 1

Response Times (RT), Error Rates and Exclusion Rates for Experiments 1 and 2

Experiment	Emotion	Gaze direction and Motion										Exclusion rate
		Total		sudden direct		static direct		sudden averted		static averted		
		Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	Mean RT (SD)	Mean error rate (SD)	
1a	angry	1087 (±151)	2.05 (±2.22)	1063 (±137)	2.08 (±2.15)	1108 (±144)	2.55 (±2.20)	1086 (±162)	1.95 (±2.34)	1091 (±165)	1.61 (±2.20)	12.3
	fearful	1113 (±161)	2.81 (±2.48)	1104 (±162)	2.42 (±2.41)	1137 (±138)	3.76 (±2.55)	1098 (±176)	2.49 (±2.37)	1112 (±172)	2.55 (±2.45)	
1b	happy	1054 (±133)	3.74 (±3.64)	1026 (±119)	3.52 (±3.68)	1058 (±129)	4.04 (±3.99)	1051 (±156)	3.52 (±2.67)	1081 (±126)	3.91 (±4.16)	8.9
	disgusted	1.52 (±135)	3.22 (±3.75)	1062 (±146)	3.06 (±2.90)	1058 (±110)	3.32 (±4.33)	1024 (±131)	3.58 (±4.14)	1059 (±152)	2.93 (±3.62)	

2	happy	951.8 (±169.0)	2.99 (±3.52)	929 (±166)	3.33 (±3.53)	953 (±158)	2.92 (±3.73)	950 (±179)	2.85 (±3.81)	976 (±173)	2.85 (±3.02)	6.8
	disgusted	949.7 (±168.2)	3.06 (±3.18)	942 (±185)	2.71 (±3.10)	971 (±158)	3.26 (±2.98)	926 (±159)	3.26 (±3.18)	960 (±170)	2.99 (±3.45)	

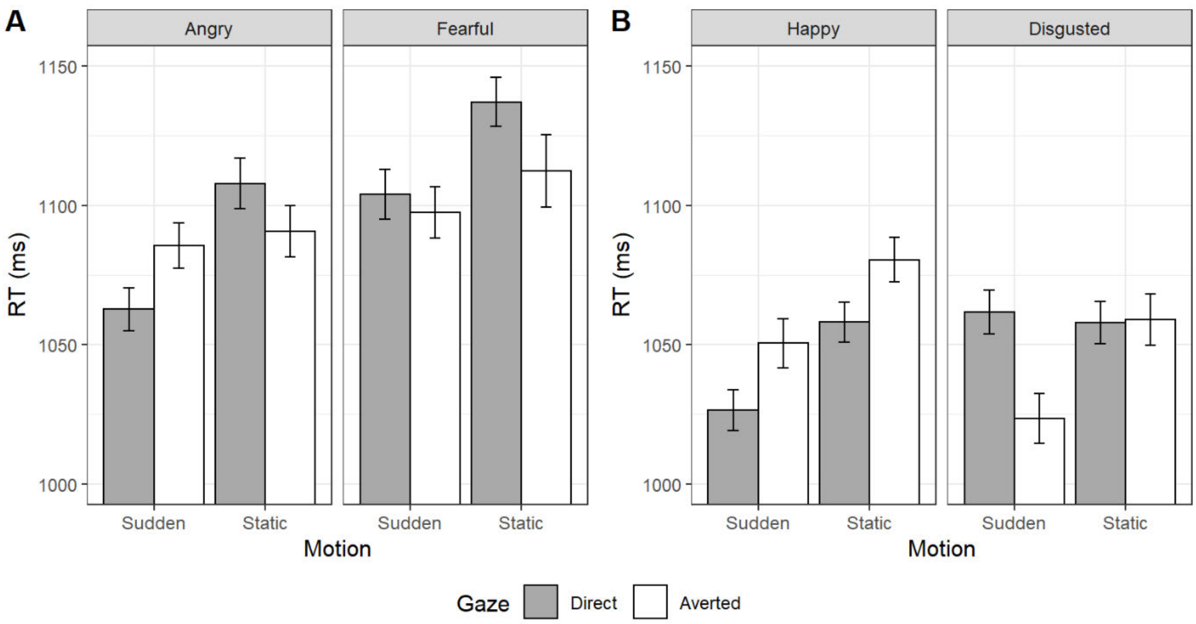
Note. Mean RTs of correct responses in ms; error and exclusion rates in %

Results

The following data sets are available on the Open Science Framework (DOI: 10.17605/OSF.IO/HFD58). Although none of the experiments were preregistered, the experimental design, all conditions, measures and hypotheses of this study were specified in the grant application that financed the study (respective sections can be found in Supplement S1). The mean RTs for all experimental conditions of Experiment 1 are visualized in Figure 2.

Figure 2

Mean RTs for each Emotion Expression, Gaze Direction and Motion Condition in Experiment 1



Note. Panel A: Experiment 1a, angry/fearful; Panel B: Experiment 1b, happy/disgusted. Error bars represent SE.

Experiment 1a: Angry/fearful

Consistent with previous studies, the main effect of motion (F(1,30) = 17.86, p < .001, η_p^2 = .37) was significant, with shorter RTs to targets on sudden stimuli than on static stimuli. In addition, the main effect of emotion expression (F(1,30) = 16.05, p < .001, η_p^2 = .35) was significant, with shorter RTs to faces with an angry expression compared to a fearful expression. In contrast to previous studies, there was no significant main effect of gaze direction ($F(1,30) = 0.44$, $p = .514$, $\eta_p^2 < .001$).

The gaze direction×motion interaction was significant ($F(1,30) = 7.677$, $p = .010$, $\eta_p^2 = .20$) but was not robust enough to survive post-hoc testing. In particular, there was a 21 ms RT advantage for averted gaze over direct gaze for static stimuli that approached conventional levels of statistical significance ($t(30) = 1.76$, $p = .089$, $dz = .253$) whereas there was a smaller but not significant 8 ms difference of gaze direction in sudden onset motion trials ($t(30) = 0.82$, $p = .421$, $dz = .109$). Importantly, there was no significant interaction between gaze direction and emotion expression ($F(1, 30) = 2.368$, $p = .134$, $\eta_p^2 = .07$), hence, the size of the direct gaze effect did not differ significantly between approach-oriented (angry) and avoidance-oriented (fearful) expressions. No other effects were significant (motion×emotion expression: $F(1, 30) = 0.006$, $p = .940$, $\eta_p^2 < .001$; gaze direction×motion×emotion expression: $F(1, 30) = 0.645$, $p = .428$, $\eta_p^2 = .02$).

When analyzing error rates, the main effect of emotion expression was significant because participants produced more errors in response to fearful faces ($F(1,30) = 13.27$, $p = .001$, $\eta_p^2 = .31$). No other main effects were found (gaze direction: $p = .100$, $\eta_p^2 = .09$, motion: $p = .151$, $\eta_p^2 = .07$). There was also a gaze direction×motion interaction ($F(1,30) = 4.794$, $p = .036$, $\eta_p^2 = .14$) due to more errors in direct compared to averted gaze trials for static stimuli ($t(30) = 2.751$, $p = .01$, $dz = .344$) while gaze direction had no effect on error rates in sudden onset motion trials ($t(30) = 0.08$, $p = .936$, $dz = .011$). No other effects were significant (gaze direction×emotion expression: $F(1, 30) = 0.005$, $p = .946$, $\eta_p^2 < .01$; motion×emotion expression: $F(1, 30) = 1.175$, $p = .287$, $\eta_p^2 = .04$; gaze direction×motion×emotion expression: $F(1, 30) = 0.138$, $p = .713$, $\eta_p^2 < .001$). Overall, the result pattern for error rates mirrors RT results, indicating that there was no speed-accuracy trade-off.

Taken together, these findings reveal a general processing benefit for angry faces, which is in line with the literature (Calvo & Nummenmaa, 2008; M. Williams et al., 2005). The results also replicate the beneficial effect of sudden onset motion on stimulus processing for both angry and fearful faces (see, for instance, Böckler et al., 2014). By contrast, direct gaze failed to induce the typical prioritization effect when faces showed angry or fearful emotion expressions. Hence, while fearful and angry expressions did not differentially modulate the direct gaze effect, these expressions seemed to override the direct gaze effect altogether when appearing simultaneously with gaze shifts.

Experiment 1b: Happy/disgusted

The analysis of RTs revealed a main effect of motion due to shorter RTs to suddenly moving stimuli compared to static stimuli ($F(1,31) = 17.07$, $p < .001$, $\eta_p^2 = .36$). Again, the main effect of gaze direction was not significant ($F(1,31) = 0.19$, $p = .688$, $\eta_p^2 = .01$) and neither was the main effect of emotion expression ($F(1,31) = 0.38$, $p = .541$, $\eta_p^2 = .01$). Importantly, however, there was a significant gaze direction×emotion expression interaction ($F(1,31) = 14.97$, $p = .001$, $\eta_p^2 = .33$). Participants had shorter RTs when responding to targets on direct gaze faces (compared to averted gaze faces) when the expression was happy ($t(31) = 3.08$, $p = .004$, $dz = .362$; direct gaze advantage: 23ms), but had shorter RTs to averted (compared to direct) gaze faces when these faces expressed disgust ($t(31) = 2.42$, $p = .02$, $dz = .247$; averted gaze advantage: 18ms). No other effects were significant (gaze direction×motion: $F(1, 30) = 3.763$, $p = .062$, $\eta_p^2 = .11$; motion×emotion expression: $F(1, 30) = 1.602$, $p = .215$, $\eta_p^2 = .05$; gaze direction×motion×emotion expression: $F(1, 30) = 1.537$, $p = .224$, $\eta_p^2 = .05$).

The error analysis revealed no significant effects (gaze direction: $F(1, 30) < 0.001$, $p > .999$, $\eta_p^2 < .001$; motion: $F(1, 30) = 0.113$, $p = .739$, $\eta_p^2 < .01$; emotion expression: $F(1, 30) = 2.952$, $p = .096$, $\eta_p^2 = .09$; gaze direction×motion: $F(1, 30) = 0.912$, $p = .347$, $\eta_p^2 = .03$; gaze direction×emotion expression: $F(1, 30) = 0.052$, $p = .820$, $\eta_p^2 < .001$; motion×emotion expression: $F(1, 30) = 0.960$, $p = .335$, $\eta_p^2 = .03$; gaze direction×motion×emotion expression: $F(1, 30) = 0.255$, $p = .617$, $\eta_p^2 = .01$). These results indicate that a speed-accuracy trade-off was unlikely.

Similar to Experiment 1a, the findings of Experiment 1b replicate the beneficial effect of sudden onset motion. In addition, when happy and disgusted facial expressions were employed, the direct gaze effect was modulated by emotion expression in line with the hypothesis by Adams and Kleck (2003, 2005). Specifically, direct gaze was beneficial when an approach-oriented emotion expression was displayed (happy) and was reversed for the avoidance-oriented expression (disgust). These findings suggest that, at least in some instances, gaze direction and emotion expression can be integrated even when they appear simultaneously.

Discussion

Experiment 1 probed the integration of two sources of social information, namely gaze direction and emotion expression. In both Experiment 1a and 1b, approach- or avoidance-oriented emotion expressions (Experiment 1a: angry or fearful; Experiment 1b: happy or disgusted) were presented simultaneously with gaze information (direct or averted) to investigate whether and how emotion expression shapes the direct gaze effect. The results indicated emotion-specific processing of gaze information. Specifically, facial expressions of anger and fear (Experiment 1a) obviated the previously established effect of gaze direction. While angry faces showed a general processing advantage (Calvo & Nummenmaa, 2008; M. Williams et al., 2005), fear and anger did not exert differential effects on gaze processing. It seems that emotion expressions in Experiment 1a were prioritized over gaze information when appearing simultaneously to gaze shifts. One possible explanation for this finding is that the information that is critical for both cues is provided in (gaze direction) or very close to (emotion expression) the eyes of the presented faces. For instance, the expression of anger crucially depends on the musculus corrugator supercilii, which are located slightly above and between the eyes (Du et al., 2014). This spatial closeness could lead to a prioritization of emotion over gaze cues. An alternative explanation is that anger and fear are particularly strong and action-related emotions that could override gaze information, especially when they are presented in a sudden fashion and simultaneous with gaze shifts (Ortony & Turner, 1990; Schupp et al., 2004; Vuilleumier, 2002b). This hypothesis is supported by the overall shorter RTs for angry/fearful (Experiment 1a: 1009 ms) compared to happy/disgusted expressions (Experiment 1b: 1051 ms), suggesting that the strong and threat-related emotions in Experiment 1a had a particularly powerful effect on attention.

By contrast, happy and disgusted expressions (Experiment 1b) differentially shaped the direct gaze effect. In line with the gaze congruency hypothesis by Adams and Kleck (2003, 2005), direct gaze (which is approach-oriented) was beneficial with approach-oriented happy facial expressions. In contrast, avoidance-oriented disgusted facial expressions facilitated responses for averted gaze (which also can signal avoidance). This modulation of gaze effects by emotion expression suggests that these two facial cues can be integrated early in the course of visual processing.

Taken together, these findings strengthen the idea of a flexible integration of gaze and other face information with emotion-specific processing. While anger and fear seemed to override any effects of gaze cues on performance, happiness and disgust were integrated with gaze direction before a response was made, even when the cues appeared simultaneously. Possible explanations for this pattern of results are addressed in the general discussion, after examining additional evidence concerning the time course of visual attention in Experiment 2.

Experiment 2

To better understand the processes that take place during the integration of gaze direction and emotion expression, Experiment 1b (happy versus disgusted) was repeated while tracking participants' eyes. It was expected that this study would replicate and extend behavioral results from Experiment 1. Specifically, it was expected that fixations would be earlier and longer for *direct* gaze on *happy* faces but for *averted* gaze on *disgusted* faces.

Furthermore, we were interested in the time course of visual attention in face processing to specify the time window of the integration process of gaze direction and emotion expression. To that end, we analyzed the probability of fixating a particular stimulus as a function of time after target presentation, gaze direction, motion, and emotion expression.

Methods

Participants

Thirty participants (19 females, 29 right-handed, mean age = 21.2±6.4 years) took part in Experiment 2 in exchange for 10€ or course credit. All participants provided informed consent prior to the experiment. According to G*Power (Faul et al., 2007b), this sample size is sufficiently large to potentially detect small effect sizes. Based on the findings of Experiment 1b (medium effect sizes), this sample is sufficient to detect a gaze direction×emotion expression interaction effect.

Task and procedure

Participants were seated in front of a 69-cm monitor and placed their head on a chin rest. The index fingers of the left and right hand rested on the response keys ("S" and "H") on a standard keyboard. Participants completed the same task as in Experiment 1b, reacting to target letters ("S" and "H") that were presented randomly on one of four happy or disgusted faces (each 3.8×4.7°VA), while their eyes were tracked with an EyeLink 1000 Plus eye tracker (SR Research, Ontario, Canada). A four-point calibration was performed prior to each test block and again after every tenth trial.

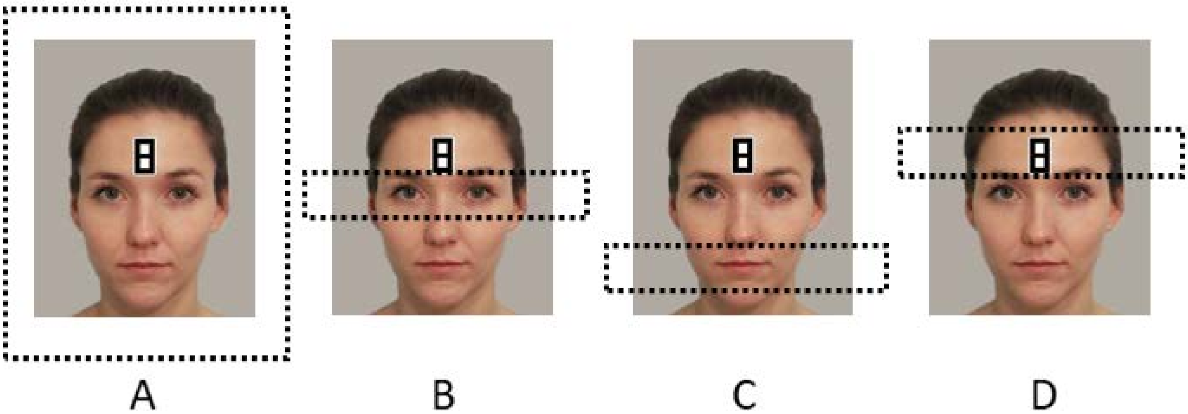
Data analysis

The analysis of manual responses was consistent with the analysis protocol of Experiment 1. Eye tracking data were analyzed with the corresponding EyeLink software (Data Viewer version 3.2) and with customized R scripts (Barthelme, 2019; Ferguson, 2018;

R Core Team, 2021). Four rectangular areas of interest (AOIs) were created for each of the four stimulus positions (Figure 3): one for the whole picture (204×201 pixels) plus one for each of three subregions (204×38 pixels each) – eyes, mouth and target/distractor letter. Three dependent variables were central to our investigation of gaze behavior at the trial level: latency (in ms) - time after target presentation until a whole picture AOI at any stimulus position was fixated; first fixations (in percent) - the proportion of trials in which a particular AOI at any stimulus position was the first AOI to be fixated[1]; and finally, relative dwell time (in percent) - the cumulative duration of fixations in a particular AOI at a particular stimulus position, divided by the cumulative duration of fixations in any AOI at any stimulus position[2].

Figure 3

Eye tracking Areas of Interest (AOIs) in Experiment 2



Note. AOIs are represented by dashed lines. A = whole face; B = eyes; C = mouth; D = targets/distractors.

---

[1] Any saccades after target presentation that landed outside of the defined AOIs were disregarded here. First fixations were calculated separately for whole stimulus AOIs and subregions.

[2] Relative dwell time was calculated separately for whole stimulus AOIs and subregions.

Similar to manual responses, a 2×2×2 repeated measures ANOVA with the within-subject factors gaze direction (direct, averted), motion (static, sudden) and emotion expression (happy, disgusted) was conducted for each dependent variable. For relative dwell times and first fixations, separate ANOVAs were conducted for different AOIs. Two-sided *t*-tests with Bonferroni correction for multiple testing were applied to examine interaction effects.

To investigate the time course underlying the integration of visual emotion expression and gaze information, an analysis was performed on the sample level. For each participant and for each measured time point after target presentation (i.e., once per ms), the number of trials in which the whole picture AOIs were fixated was counted. This approach was completed separately for each AOI and for all combinations of the factors stimulus position, gaze direction (at stimulus position), motion (at stimulus position) and emotion expression. We then divided that number by the total number of trials with the same factor specification. For example, participant PB2126 performed 56 trials in which the topmost face exhibited averted gaze, sudden motion and disgust. On millisecond 345 after target presentation, participant PB2126 fixated the corresponding AOI in 18 out of these 56 trials, yielding a proportion of 0.32. When averaged across participants and stimulus positions, the corresponding value can be interpreted as a measure of the probability of looking at a picture with a given combination of facial features at a particular point in time. This fourth DV (probability) was used for both plotting millisecond-by-millisecond time courses of gaze behavior, and in a window analysis where we partitioned the first 1000 ms of each trial into four 250-ms bins. Specifically, we calculated 2×2×2×4 repeated measures ANOVAs with the within-subject factors gaze direction, motion and emotion expression, and time window (0-250, 250-500, 500-750, 750-1000 ms).

Results and discussion

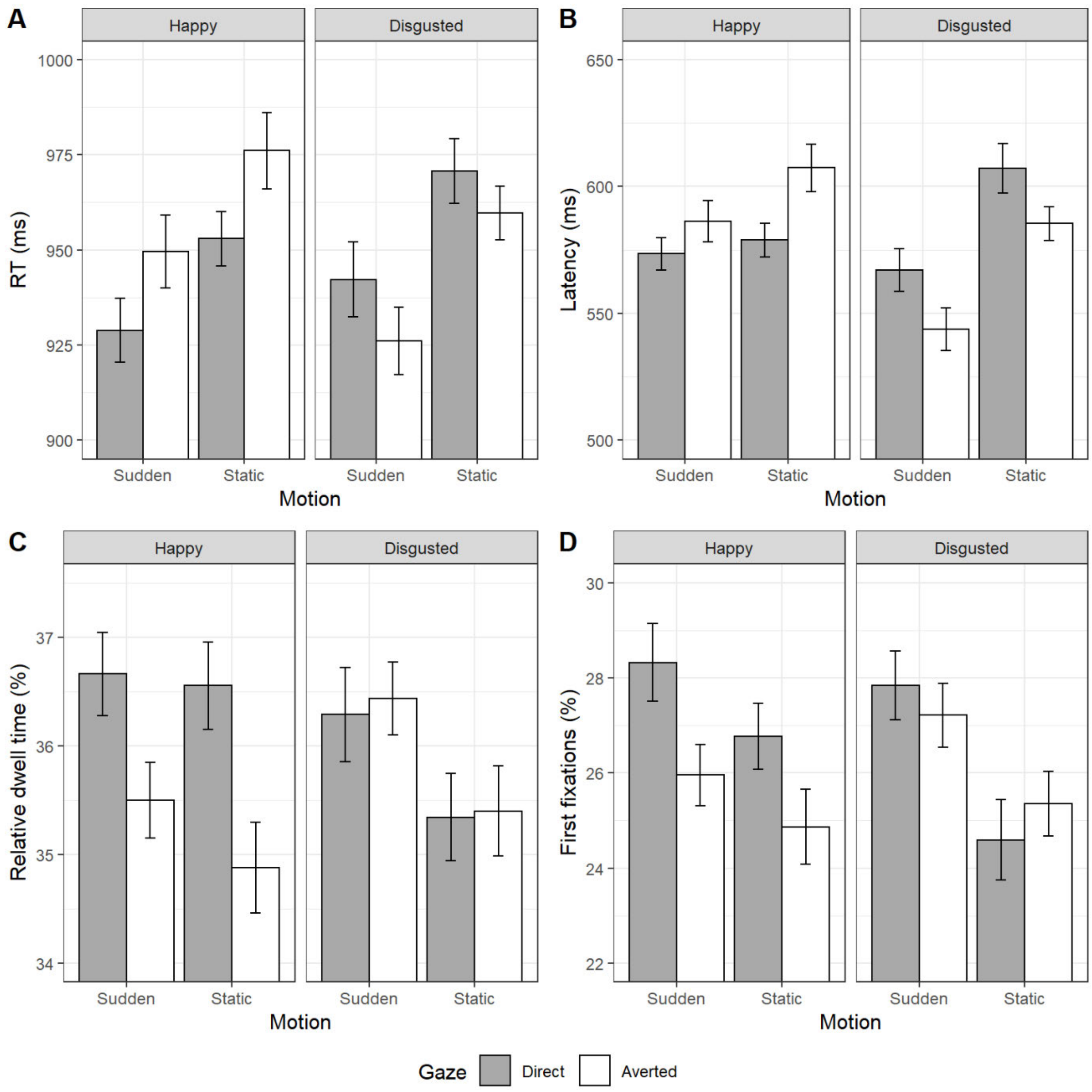
The following analyses are available under (https://tim-raettig.shinyapps.io/datex_gaze/).

Manual behavior

Table 1 shows error and exclusion rates as well as mean RTs for each emotion expression, gaze direction and motion condition. Mean RTs per condition are visualized in Figure 4 (Panel A).

Figure 4

Behavioral Measures for each Emotion Expression, Gaze Direction and Motion Condition in Experiment 2



Note. Panel A: manual behavior; Panels B-D: gaze behavior. Error bars represent SE.

Performance (RTs and error rates)

Effects on RTs largely mirrored those of Experiment 1b: there was a main effect of motion with shorter RTs to sudden than to static stimuli (F(1,29) = 21.60, p < .001, η^2_p = .427), and no main effect for gaze direction (F(1,29) = .39, p = .54, η^2_p = .013). Critically, the gaze direction×emotion expression interaction effect was again significant (F(1,29) = 10.42, p = .003, η^2_p = .264). For happy faces, responses were significantly shorter when the faces depicted direct over averted gaze (t(29) = 2.46, p = .04, dz = .449; direct gaze advantage: 22 ms). Though averted gaze was numerically advantageous in disgusted faces, no significant effect of gaze direction was found for disgusted faces (t(29) = 1.64, p = .23, dz = .298; averted gaze advantage: 14 ms). No other effects were significant (emotion expression: F(1,29) = 0.141, p = .710; η_p^2 = .005; gaze direction×motion: F(1,29) = 0.096, p = .759; η_p^2 = .003; motion×emotion expression: F(1,29) = 0.175, p = .679, η_p^2 = .006; gaze direction×motion×emotion expression: F(1,29) = 0.011, p = .917, η_p^2 < .001).

Error analysis yielded no significant effects (gaze direction: F(1,29) = 0.047, p = .830, η_p^2 = 002; motion: F(1,29) = 0.009, p = .925, η_p^2 < .001; emotion expression: F(1,29) = 0.043, p = .837, η_p^2 = 001; gaze direction×motion: F(1,29) = 0.145, p = .707, η_p^2 = 005; gaze direction×emotion expression: F(1,29) = 0.533, p = .471, η_p^2 = .018; motion×emotion expression: F(1,29) = 0.185, p = .670, η_p^2 = .006; gaze direction×motion×emotion expression: F(1,29) = 0.726, p = .401, η_p^2 = .024).

Gaze behavior

Descriptive statistics on gaze behavior are presented in Table 2 and visualized in Figure 4 (Panels B, C and D).

Table 2

Mean Fixation Patterns in Experiment 2.

Area of interest	Emotion expression	Gaze direction	Motion	Latency (mean (SD) ms)	First fixation (mean (SD) %)	Relative dwell time (mean (SD) %)
Whole picture	Happy	Direct	Sudden	573 (97)	28.3 (5.8)	36.7 (3.9)
			Static	579 (87)	26.8 (5.0)	36.6 (3.9)
		Averted	Sudden	586 (107)	26.0 (4.9)	35.5 (3.8)
			Static	607 (92)	24.9 (4.0)	34.9 (3.4
	Disgusted	Direct	Sudden	567 (95)	27.8 (5.5)	36.6 (3.9)
			Static	607 (104)	24.6 (4.7)	35.3 (3.6)
		Averted	Sudden	544 (87)	27.2 (4.7)	36.4 (3.9)
			Static	585 (107)	25.4 (4.3)	35.4 (3.6)
Eyes	Happy	Direct	Sudden		10.9 (4.7)	15.0 (5.2)
			Static		11 (4.2)	15.9 (4.7)
		Averted	Sudden		10.4 (4.7)	14.2 (4.6)

			Static	9.5 (4.2)	14.0 (4.6)
	Disgusted	Direct	Sudden	9.8 (4.2)	13.4 (4.3)
			Static	9.6 (4.8)	13.9 (4.9)
		Averted	Sudden	10.9 (3.9)	14.9 (4.8)
			Static	10.1 (4.0)	14.6 (4.6)
Mouth	Happy	Direct	Sudden	3.2 (3.7)	3.6 (3.4)
			Static	3.0 (2.8)	3.7 (3.3
		Averted	Sudden	2.6 (2.8)	3.7 (3.2)
			Static	3.3 (3.3)	4.0 (4.0)
	Disgusted	Direct	Sudden	3.3 (3.0)	3.8 (3.1)
			Static	2.5 (2.8)	3.2 (3.6)
		Averted	Sudden	3.6 (3.6)	4.4 (3.6)
			Static	3.5 (3.2)	4.1 (3.2)
Target/distractor letter	Happy	Direct	Sudden	13.5 (6.9)	17.9 (6.8)
			Static	12.2 (5.0)	17.6 (6.3)
		Averted	Sudden	12.6 (5.1)	17.9 (6.8)

		Static	11.4 (5.0)	16.9 (6.1)
Disgusted	Direct	Sudden	13.7 (6.3)	18.8 (6.7)
		Static	11.2 (5.2)	17.5 (5.9)
	Averted	Sudden	13.0 (5.4)	17.6 (7.0)
		Static	12.2 (5.5)	17.3 (6.4)

Whole picture. ***Latency.*** Similar to RTs, the main effects of motion (F(1,29) = 25.527, p < .001, η^2_p = .468) and emotion expression (F(1,29) = 5.847, p = .022, η^2_p = .168) were significant, while there was no main effect of gaze direction (F(1,29) = 0.041, p = .841, η^2_p = .001). Specifically, we found earlier fixations for faces that had suddenly moved simultaneous to target presentation (compared to static faces) and to faces with a disgusted expression (compared to a happy expression).

Critically, the gaze direction×emotion expression interaction effect was significant (F(1,29) = 13.897, p = .001, η^2_p = .324) because participants were faster to make a saccade to happy stimuli when they depicted direct (versus averted) gaze (t(29) = 3.229, p = .006, dz = .59; direct gaze advantage: 21 ms). In contrast, disgusted faces were fixated earlier when they showed averted gaze (t(29) = 2.698, p = .024, dz = .493; averted gaze advantage: 23 ms). No other effects were significant (gaze direction×motion: F(1,29) = 0.550, p = .464, η^2_p = 0.019; motion×emotion expression: F(1,29) = 3.445, p = .074, η^2_p = .106; gaze direction×motion×emotion expression: F(1,29) = 0.348, p = .560, η^2_p = .012).

First fixations. For the whole picture area, there was a main effect of motion (F(1,29) = 9.384, p = .005, η^2_p = .244), with relatively more first fixations on sudden motion stimuli. The factors gaze direction and emotion expression did not yield significant effects (gaze direction: F(1,29) = 3.782, p = .062, η^2_p = .115; emotion expression: F(1,29) = 0.504, p = .483, η^2_p = .017).

The gaze direction×emotion expression interaction effect was significant (F(1,29) = 4.732, p = .038, η^2_p = .140) due to significantly more first fixations on happy faces with direct compared to averted gaze (t(29) = 3.139, p = .008, dz = 0.573; direct gaze advantage: 2.1%). There was no significant difference for faces with a disgusted expression (p = .999). No other effects were significant (gaze direction×motion: F(1,29) = 0.719, p = .403, η^2_p = .024; motion×emotion expression: F(1,29) = 1.174, p = .287, η^2_p = .039; gaze direction×motion×emotion expression: F(1,29) = 0.229, p = .636, η^2_p = .008).

Relative dwell time. We found significant main effects of gaze direction (*F*(1,29) = 8.930, *p* = .006, η^2_p = .235) and motion (*F*(1,29) = 4.265, *p* = .048, η^2_p = .128), due to more time being spent on faces that established direct gaze (versus averted gaze) and that moved simultaneous to target presentation (versus remained static). Similar to latencies and first fixations of the whole picture and the eye regions, there was a significant gaze direction×emotion expression interaction effect (*F*(1,29) = 5.982, *p* = .021, η^2_p = .171), such that happy faces that established direct gaze were fixated longer compared to happy faces with averted gaze (*t*(29) = 3.667, *p* = .002, *dz* = 0.669; direct gaze advantage: 14.0%). In contrast, no significant difference in gaze conditions was found for disgusted faces (*t*(29) = 0.271, *p* = .999, *dz* = 0.049). No other effects were significant (emotion expression: *F*(1,29) = 0.029, *p* = .866, η^2_p = .001; gaze direction×motion: *F*(1,29) = 0.235, *p* = .631, η^2_p = .008; motion×emotion expression: *F*(1,29) = 0.953, *p* = .337, η^2_p = .032; gaze direction×motion×emotion expression: *F*(1,29) = 0.200, *p* = .658, η^2_p = .007).

Eyes. ***First fixations.*** We found no main effects for the eye region (gaze direction: *F*(1,29) = 0.138, *p* = .713, η^2_p = .005; emotion expression: *F*(1,29) = 1.227, *p* = .277, η^2_p = .041; motion: *F*(1,29) = 1.214, *p* = .280, η^2_p = .040). However, consistent with the results for the whole picture, there was a significant gaze direction×emotion expression interaction effect for the percentage of first fixations on the eye area (*F*(1,29) = 6.626, *p* = .015, η^2_p = .186). Though the direct gaze advantage was numerically present for happy faces (11%) and numerically reversed for disgusted faces (8%), none of the Bonferroni-corrected post-hoc pairwise comparisons became significant (happy: *t*(29) = 1.876, p = .142, *dz* = 0.342; disgusted: *t*(29) = 1.879, *p* = .140, *dz* = 0.343). No other effects were significant (gaze direction×motion: *F*(1,29) = 1.460, *p* = .237, η^2_p = .048; motion×emotion expression: *F*(1,29) = 0.033, *p* = .858, η^2_p = .001; gaze direction×motion×emotion expression: *F*(1,29) = 0.070, *p* = .793, η^2_p = .002).

Relative dwell time. The gaze direction×emotion interaction effect was also significant for the relative time spent on the eye regions ($F(1,29) = 7.836$, $p = .009$, $\eta^2_p = .213$). Participants fixated longer on the eyes of disgusted faces with averted gaze than direct gaze ($t(29) = 2.476$, $p = .038$, $dz = .452$; averted gaze advantage: 13%) whereas no significant difference was found for happy faces ($t(29) = 2.016$, $p = .102$, $dz = 0.368$). No other effects were significant (gaze direction: $F(1,29) = 0.158$, $p = .694$, $\eta^2_p = .005$; motion: $F(1,29) = 0.006$, $p = .939$, $\eta^2_p < .001$; emotion expression: $F(1,29) = 1.916$, $p = .177$, $\eta^2_p = .062$; gaze direction×motion: $F(1,29) = 0.701$, $p = .409$, $\eta^2_p = .024$; motion×emotion expression: $F(1,29) = 0.191$, $p = .665$, $\eta^2_p = .007$; gaze direction×motion×emotion expression: $F(1,29) = 0.258$, $p = .616$, $\eta^2_p = .009$).

Mouth. ***First fixations.*** There were no significant effects for the percentage of first fixations on the mouth area (gaze direction: $F(1,29) = 1.561$, $p = .222$, $\eta^2_p = .051$; motion: $F(1,29) = 0.151$, $p = .700$, $\eta^2_p = .005$; emotion expression: $F(1,29) = 1.942$, $p = .174$, $\eta^2_p = .063$; gaze direction×motion: $F(1,29) = 2.690$, $p = .112$, $\eta^2_p = .085$; gaze direction×emotion expression: $F(1,29) = 2.847$, $p = .102$, $\eta^2_p = .089$; motion×emotion expression: $F(1,29) = 2.869$, $p = .101$, $\eta^2_p = .090$; gaze direction×motion×emotion expression: $F(1,29) = 0.012$, $p = .915$, $\eta^2_p < .001$).

Relative dwell time. In contrast to the analysis on first fixations, we found a main effect of gaze direction due to more time being spent on the mouth regions of averted gaze stimuli ($F(1,29) = 5.330$, $p = .028$, $\eta^2_p = .155$). There was also a significant gaze direction×emotion expression interaction effect ($F(1,29) = 4.425$, $p = .044$, $\eta^2_p = .132$), as relatively more time was spent on the mouth region of disgusted faces with averted as compared to direct gaze ($t(29) = 2.893$, $p = .014$, $dz = 0.528$; averted gaze advantage: 8%), but no significant difference between gaze directions for happy faces ($t(29) = 0.857$, $p = .798$, $dz = .156$). No other effects were significant (motion: $F(1,29) = 0.321$, $p = .575$, $\eta^2_p = .011$; emotion expression: $F(1,29) = 0.933$, $p = .342$, $\eta^2_p = .031$; gaze direction×motion: $F(1,29) =$

0.286, p = .597, η^2_p = .010; motion×emotion expression: $F(1,29)$ = 3.928, p = .057, η^2_p = .119; gaze direction×motion×emotion expression: $F(1,29)$ = 0.005, p = .942, η^2_p < .001).

Targets/distractors. *First fixations.* There was a main effect of motion ($F(1,29)$ = 12.741, p = .001, η^2_p = .305), with relatively more first fixations on targets/distractors for sudden motion stimuli. No other effects were significant (gaze direction: $F(1,29)$ = 0.584, p = .451, η^2_p = .020; emotion expression: $F(1,29)$ = 0.125, p = .727, η^2_p = .004; gaze direction×motion: $F(1,29)$ = 1.381, p = .249, η^2_p = .045; gaze direction×emotion expression: $F(1,29)$ = 1.387, p = .249, η^2_p = .046; motion×emotion expression: $F(1,29)$ = 0.264, p = .611, η^2_p = .009; gaze direction×motion×emotion expression: $F(1,29)$ = 1.314, p = .261, η^2_p = .043).

Relative dwell time. There were no significant effects for the time spent on target/distractor letter regions (gaze direction: $F(1,29)$ = 4.058, p = .053, η^2_p = .123; motion: $F(1,29)$ = 0.565, p = .089, η^2_p = .019; emotion expression: $F(1,29)$ = 0.565, p = .458, η^2_p = .019; gaze direction×motion: $F(1,29)$ = 0.014, p = .908, η^2_p < .001; motion×emotion expression: $F(1,29)$ = 0.100, p = .755, η^2_p = .003; gaze direction×emotion expression: $F(1,29)$ = 0.180, p = .674, η^2_p = .006; gaze direction×motion×emotion expression: $F(1,29)$ = 1.100, p = .303, η^2_p = .037).

Time course analysis.

To investigate the temporal evolution of gaze behavior for the whole picture AOI, the first 1000 ms of the RT period (after target presentation) was split into four windows of 250 ms. We then completed a new repeated-measures ANOVA with the DV probability and the within-subject factors gaze direction, motion, emotion expression and time window (0-250, 250-500, 500-750, 750-1000 ms).

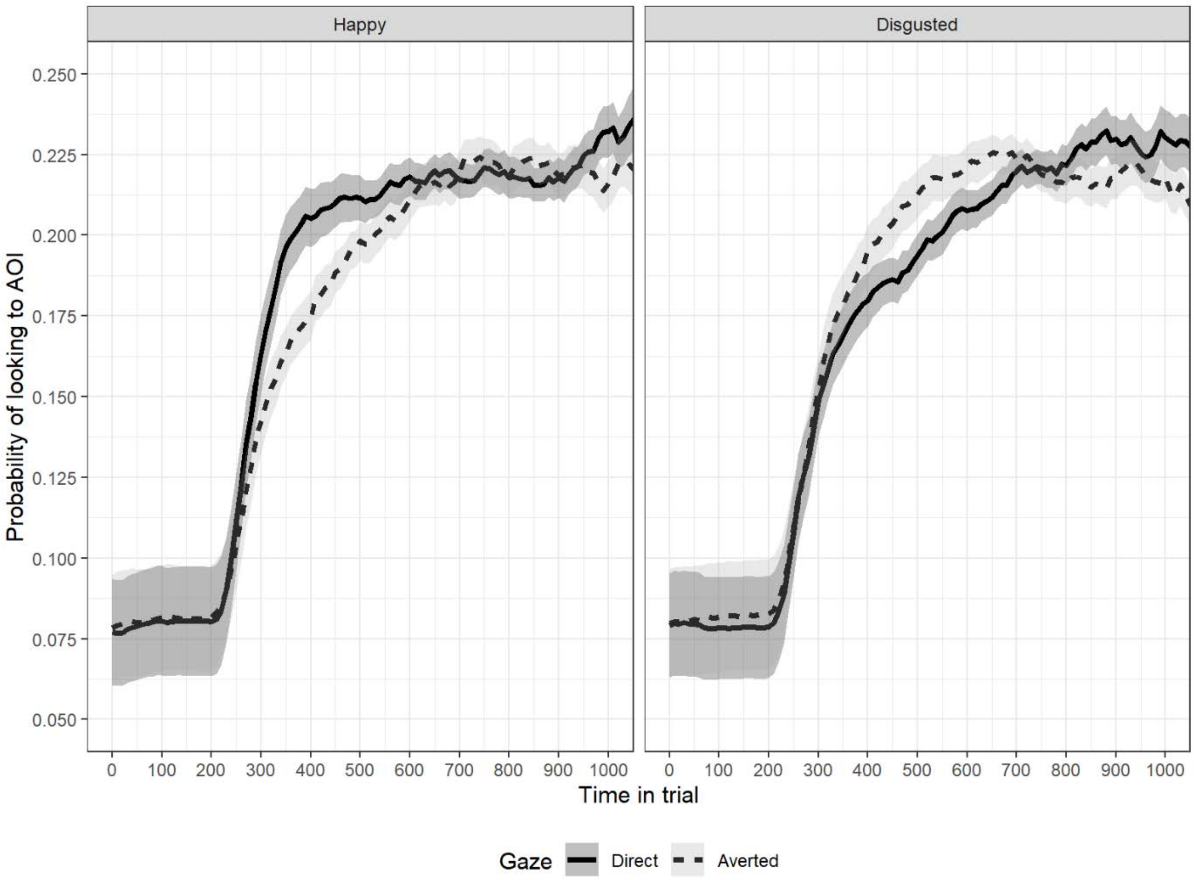
Several expected effects were observed, including the main effect of motion ($F(1,29)$ = 4.81, p = .037, η^2_p = .142) due to overall higher probabilities of picture fixation for sudden compared to static stimuli (sudden: 17.3; static: 16.9). In addition, there was a significant main effect of time window ($F(3,87)$ = 56.98, p < .001, η^2_p = .663) with higher probabilities of fixations in later time windows (0-250 ms: 0.082; 250-500 ms: 0.173; 500-750 ms: 0.212; 750-1000 ms: 0.217). The factors gaze direction and emotion expression did not reveal significant main effects (gaze direction: $F(1,29) = 0.311$, $p = .581$ $\eta_p^2 = .011$; emotion expression: : $F(1,29) = 0.420$, $p = .522$, $\eta_p^2 = .014$). A significant motion×time window interaction was observed ($F(3,87) = 14.70$, $p < .001$, $\eta^2_p = .336$) due to significant differences between the motion conditions only occurring in the time window of 250-500 ms ($t(29) = 6.05$, $p < .001$, $dz = 1.104$) but not in the other time windows (0-250: $t(29) = 1.764$, $p = .352$, $dz = 0.322$; 500-750: $t(29) = 0.352$, $p > .999$; $dz = .064$; 750-1000: $t(29) = 1.531$, $p > .999$, $dz = .279$). Between 250 and 500 ms, stimuli with sudden onset motion simultaneous to target presentation had a higher probability of fixation.

Critically, the gaze direction×emotion expression interaction effect was significant ($F(1,29) = 4.90$, $p = .035$, $\eta^2_p = .145$), due to a higher probability of picture fixation of happy faces with direct compared to averted gaze (17.4 vs. 16.8), but a higher probability of fixating disgusted faces with averted compared to direct gaze (17.2 vs. 16.9). This interaction was further qualified by a three-way interaction of gaze direction×emotion expression×time window ($F(3,87) = 9.22$, $p < .001$, $\eta^2_p = .241$). Specifically, happy faces with direct compared to averted gaze were more likely to be fixated between 250-500 ms ($t(29) = 5.37$, $p < .001$, $dz = .981$), but in no other time window (0-250: $t(29) = 0.479$, $p > .999$, $dz = .088$; 500-750: $t(29) = -0.849$, $p = .806$, $dz = .155$; 750-100: $t(29) = 0.522$, $p > .999$, $dz = .095$). In contrast, although disgusted faces with averted gaze were more likely to be fixated compared to direct gaze, these differences approached but not cross conventional levels of statistical significance: 250-500 ms ($t(29) = 2.27$, $p = .062$, $dz = .414$), between 500-750 ms ($t(29) = 2.10$, $p = .088$, $dz = .384$) and between 750-1000 ms ($t(29) = 2.29$, $p = .058$, $dz = .418$), 0-250 ms ($t(29) = 0.767$, $p = .898$, $dz = .14$). No other effects were significant (gaze direction×motion: $F(1,29) = 0.651$, $p = .426$, $\eta^2_p = .022$; gaze direction×time window: $F(3,87) = 1.685$, $p = .176$, $\eta^2_p = .055$; motion×emotion expression: $F(1,29) = 1.208$, $p = .281$, $\eta^2_p = .040$; emotion expression×time window: $F(3,87) = 1.251$, $p = .296$, $\eta^2_p = .041$; gaze direction×motion×emotion expression: $F(1,29) = 1.303$, $p = .263$, $\eta^2_p = .043$; gaze direction×motion×time window: $F(3,87) = 1.402$, $p = .248$, $\eta^2_p = .046$; motion×emotion expression×time window: $F(3,87) = 1.038$, $p = .380$, $\eta^2_p = .035$; gaze direction×motion×emotion expression×time window: $F(3,87) = 1.229$, $p = .304$, $\eta^2_p = .041$).

Visual inspection of the timelines (Figure 5) supported the statistical findings of the window analysis, but allowed us to more precisely pinpoint start and duration of the integration of facial emotional and gaze information. Specifically, direct and averted gaze were differentiated earlier for happy faces (around 275 ms) than for disgusted faces (around 375 ms), with the effect lasting around 225 ms for both.

Figure 5

Time Course of Gaze Behavior



Note. Probability of looking at a whole picture AOI with a given combination of facial features (emotion expression and gaze direction, averaged across motion conditions) at a particular point in time.

Discussion

Experiment 2 explored the temporal dynamics of visual attention to social cues. The findings clearly support the idea of an early integration of gaze direction and emotion expression. In line with results from Experiment 1b, the results revealed a benefit for direct over averted gaze in happy faces, and no such effect, or even an averted gaze advantage in some measures, for disgusted faces. These effects were present both at the level of manual responses and at the level of gaze behavior. Time course analysis suggests that the integration process of gaze and face information begins between 275 - 375 ms after stimulus onset, and lasts around 225 ms. These findings speak to a response facilitation by congruent gaze and emotion expression information and an *early integration* of these cues in accordance with the approach/avoidance congruency hypothesis by Adams and Kleck (2003, 2005).

Effects of gaze direction were more reliable for happy faces. A statistically significant direct gaze effect for happy faces was observed in the RTs as well as in all fixation measures on whole faces. In contrast, the results of the statistical analyses for disgusted faces were less consistent. A statistically significant averted gaze advantage was found for disgusted faces in the latency of whole face fixations and relative dwell times on the eye and mouth areas. Although no statistically significant differences were present for manual responses or for the percentage of first fixations in any AOI, numerical trends were observable in the hypothesized direction (see Figures 4 and 5; Table 2). Taken together, while the direct gaze advantage was stable for faces exhibiting an approach-oriented emotion expression, the effect disappeared or was even reversed for faces with an avoidance-oriented expression.

The mouth is a highly salient area for both happy and disgusted faces and one of the most informative facial regions for the detection of these emotions (Calvo et al., 2014; Calvo & Nummenmaa, 2008; Dimberg et al., 2000; Du et al., 2014; Vrana, 1993). The results revealed higher relative dwell times on the mouth region of disgusted faces when the face depicted averted compared to direct gaze. In contrast, no significant difference was found between gaze directions for happy faces. This finding might suggest that the mouth of faces with a disgusted expression is especially salient with averted gaze whereas a smiling mouth could attract attention regardless of gaze direction. Finally, there were no effects for target/distractor regions, indicating that this region was equally important in all conditions.

General discussion

Direct gaze can indicate that another person is interested in approaching the observer with at least two possibilities: The gaze cue can be an attempt to initiate a friendly and cooperative social interaction or it can signal aggression or disaffirmation. Likewise, averted gaze can either suggest that the other person is avoiding the observer or that they are more interested in an object or event in the environment (that could be meaningful to us as well). Emotion expressions provide important information that can help one to distinguish between these possibilities. In the present study, it was investigated whether and when gaze and emotion cues are integrated. The results indicate that this process occurs early in visual processing and is emotion-specific.

The results of Experiment 1 indicate that the expressions of anger and fear diminished the direct gaze effect, while the processing advantage for motion onset cues remained intact. In contrast, expressions of happiness and disgust differentially modulated the effect of gaze information. As predicted by the approach/avoidance congruency hypothesis by Adams and Kleck (2003, 2005), responses to happy faces (approach-oriented) were facilitated by direct gaze. In contrast, responses to disgusted faces (avoidance-oriented) were facilitated by averted gaze. Results of Experiment 2 mirrored the effects for happy and disgusted faces at the level of behavior and visual attention. Timeline analyses of fixation patterns suggest an integration of gaze and expression information starting around 300 ms after stimulus onset when both cues are presented simultaneously. Hence, the findings from Experiment 1b and 2 are in line with the approach/avoidance congruency hypothesis by Adams and Kleck (2003, 2005) and speak to an early and differential integration of gaze direction and emotion expression.

In Experiment 1, it was found that only some approach- versus avoidance-oriented emotion expressions, namely happiness and disgust, differentially modulated the direct gaze effect while others, in particular anger and fear, annulled it altogether. This latter finding seems to contradict the model from Adams and Kleck (2003, 2005), which predicts a modulation of the gaze effect in accordance with approach versus avoidance for all four emotions. There are at least three possible explanations for these findings. First, as mentioned earlier in the discussion of Experiment 1, expressions of anger and fear centralize around the eyes while happiness and disgust do not. Each emotion expression is the result of specific facial muscle activation patterns (e.g. Du et al., 2014). For happy and disgusted emotion expressions, the most diagnostic areas are the mouth with the *zygomaticus major* and the nose with the *levator labii*, respectively. In contrast, anger and fear are most prominently displayed by muscles around the eyes (the *corrugator supercilii*). The resulting spatial overlap with gaze cues could lead to a prioritization of emotion expression over gaze direction. A second possible explanation stems from findings on visual search that indicate that some emotion expressions rely more strongly on feature-based processing, and are easier and less costly to process than others. This notion particularly applies for happy faces where the smiling mouth represents a prominent feature that is displayed in a consistent fashion across expressers. As a result, the recognition of happy faces is faster and more accurate compared to, for example, angry and fearful faces (Calvo & Nummenmaa, 2008). Finally, anger and fear are negative, highly aversive and action-related emotion expressions (Ortony & Turner, 1990; Schupp et al., 2004; Vuilleumier, 2002b). As such, they could attract attention more strongly than happy and disgusted faces – and also more strongly than direct gaze cues. This explanation fits with findings from ERP-studies that suggest that emotional information of faces is first extracted at early processing stages, e.g. around 170 ms after stimulus presentation (N170; (for a review, see Schindler & Bublatzky, 2020). At later stages (e.g. EPN around 200-300 ms after stimulus onset), the emotion expression can compete with concurrent goals of the task, for example the detection of a target. At this stage, emotion cues can benefit from low-level information that signal biological relevance, such as gaze cues. Findings on modulation of attention by emotions are particularly inconsistent when faces serve as distractors in the task, such as in our study. Under these circumstances, effects on early ERP-components (e.g. P1, N170) are more pronounced for angry, fearful and happy faces (compared to other commonly investigated emotion expressions), suggesting a hierarchy of expressions at early processing stages (Hinojosa et al., 2015; Schindler & Bublatzky, 2020). In contrast, during passive viewing, a modulation of the N170 by emotion expression is observed more frequently, suggesting that active task demands compete for attentional resources with emotional decoding. At later stages, such as the EPN, top-down and bottom-up processes interact and threat-related emotions profit the most from task relevance. With respect to the present study, it could be hypothesized that non-threatening emotions, such as happiness and disgust, did not dominate in visual processing to a degree to corroborate the processing of gaze direction, but, at the same time, were prominent enough to be integrated with this information. In contrast, biologically relevant emotions, such as anger and fear, may have been prioritized and thereby pulled visual attention entirely away from co-occurring gaze information. Future studies can further test the boundary conditions that shape when and how emotion expressions interact with gaze information. Similarly, investigating these effects on an electrophysiological level could shed light on underlying perceptual and attentional processes.

The analyses of visual fixation patterns further substantiated the idea of emotion-specific integration processes of gaze and face information. In particular, we found overlapping but distinct time windows of interaction effects between gaze direction and emotion expression for happy versus disgusted faces. This process seemed to start earlier for happy compared to disgusted faces, but had the same duration for both emotions.

Furthermore, throughout Experiments 1 and 2, the integration of gaze and face information was reflected in more pronounced and consistent effects for happy as compared to disgusted faces. Specifically, it was found that a reliable direct gaze advantage for happy faces appears on almost all dependent variables, whereas disgusted faces produced absent or even reversed effects of gaze direction (i.e., an averted gaze advantage). This pattern matches previous studies that report stronger effects for happy compared to disgusted expressions on attention (Calvo & Nummenmaa, 2008). It could be that happy faces are more easily and reliably recognized, while the processing of disgusted faces and their subsequent integration with gaze cues is more challenging. Critically, however, the typical direct gaze advantage did not emerge for any of the measures for disgusted faces, indicating that this avoidance-oriented emotion expression reliably diminished attention capture by direct gaze.

For both happy and disgusted expressions, the present findings speak to an early integration of gaze and face information in accordance with the approach/avoidance congruency hypothesis. Even though differences were observed in the exact time course of events, eye tracking data suggest that the integration process led to reliably detectable behavioral results already around 300 ms after stimulus onset for both emotions. After this time point, happy-direct and disgusted-averted faces attracted more attention, as evident in earlier, longer and more frequent fixations compared to happy-averted and disgusted-direct stimuli. Hence, these results suggest that visual attention was driven by combined effects of gaze and face cues. In particular, while direct gaze was more salient on happy faces, attention to averted gaze faces increased for disgusted expressions in more narrow regions, namely the eyes and the mouth.

Although the present study revealed an interaction between gaze cues and emotion expressions, there was no evidence that gaze and motion cues interacted. This pattern is in line with previous studies with this paradigm (Böckler et al., 2015; Böckler, van der Wel, et al., 2014; Boyer & Wang, 2018; Breil, Huestegge, et al., 2021; van der Wel et al., 2018).

Similarly, the effect of sudden onset motion was not modulated by emotion expression in any of the experiments, neither on the level of performance nor in measures of gaze behavior. Hence, sudden onset motion captured attention and led to a processing advantage for targets appearing on moving stimuli, but the extent of this attention capture effect was independent of emotional context. This differential influence of emotion expression on the effects of sudden onset motion and gaze direction further supports the notion that gaze and motion cues are susceptible to distinct modulators and rely on parallel and independent processing channels.

The present study did not include surprised and sad expressions because their approach-avoidance tendencies are not as clear-cut. Unfortunately, particularly these expressions (as well as disgust) have received only little attention in previous research on the effects of emotion expression on attention (e.g. see Schindler & Bublatzky, 2020). Future studies should expand the focus from emotions that produce the most reliable modulations (i.e. anger, fear and happiness) to emotions that produce more subtle effects, such as sadness and surprise. Ideally, these studies should include a range of expressers, i.e. stimulus face identities, that are rated and matched for intensity. This was not the case in the present study in which we employed the same stimulus identity throughout each experiment. Note, however, that we found notable effects for the emotions that typically are rated as less intense (happiness, disgust).

Finally, it is likely that different implementations of emotion expression might differentially influence effects of gaze direction. While the present study employed a within-subject design and altered emotion expressions simultaneously with sudden onset direct/averted gaze and target presentation, other studies from our lab used a between-subject implementation of emotion expressions in that a given participant only saw faces with one and the same emotion expression throughout the experiment and throughout every trial (Pittig, van der Wel, Welsh, & Böckler, submitted). Results were also in line with the approach-avoidance-congruency hypothesis, in that the approach-oriented expression of anger increased the direct gaze advantage compared to the avoidance-oriented expression of fear, for which no direct gaze effect was observed. In this other study, however, no modulating effects were revealed for the less threat-related emotion expressions of happiness and disgust. This difference suggests that only strong and threat-related expressions like anger and fear are powerful enough to shape the direct gaze advantage in a setup where participants could get used to an expression throughout the task. By contrast, the expressions of happiness and disgust seem to exert their influence particularly when they temporally co-occur with relevant gaze information. Studies are warranted that further address the impact of design-choice on the processing and integration of basic social information.

Conclusion

Emotion expression and gaze direction are two meaningful social cues to infer the attentional and intentional states of other people (Frijda, 1986; Hartikainen et al., 2000; Lundqvist & Ohman, 2005; Öhman et al., 2001; Schilbach, 2015b; Tomasello et al., 2005b). Our study suggests that both cues interact in a flexible fashion. When temporally co-occurring with gaze information, threat-related cues, such as facial expressions of anger and fear, were prioritized over gaze information and diminished the relatively stable processing advantage for direct gaze. In contrast, expressions of happiness and disgust differentially modulated the direct gaze effect, suggesting an integration of these cues in early stages of visual processing. Starting around 300 ms after stimulus onset, happy faces received more attention when they depicted direct gaze, while disgusted faces were better detected with averted gaze. Taken together, our findings suggest that gaze and face information can be integrated early and occur in an emotion-specific manner.

Acknowledgements

This work was supported by the Emmy Noether Programme of the German Research Foundation (grant number BO 4962/1-1).

MANUSCRIPT 4

LOOK AWAY TO LISTEN: THE INTERPLAY OF EMOTIONA CONTEXT AND EYE CONTACT IN VIDEO CONVERSATIONS

Christina Breil[1], Anne Böckler[1,2]

[1]Department of Psychology, Leibniz-University Hannover, Hannover, Germany

[2]Max-Planck-Institute for Human Cognitive and Brain Science, Leipzig, Germany

This article is published in *Visual Cognition* (2021) with DOI 10.1080/13506285.2021.1908470.

Abstract

Eye gaze is a fundamental element of social interaction. We investigated the role of gaze direction during video conversations between friends, colleagues or strangers. Participants watched short video cuts of a target person engaging in direct gaze, averted gaze or a mixture of both (*gaze direction*) while listening to another, invisible, person recounting a neutral or negative autobiographical episode (*emotional context*). Subsequently, participants rated the target person on empathy, perspective taking and trustworthiness and indicated how close they perceived the relationship between conversation partners. We found that participants rated the target person and the interaction less favorable when the target's gaze was averted. Critically, these effects of gaze direction were modulated by emotional context: When narrations were negative, (partly) averted gaze had a less negative impact on participants evaluations. Hence, gaze direction is not perceived and interpreted in isolation, but in context.

Keywords: social interaction, communication, direct gaze, gaze behavior, empathy, video conversations

Introduction

Wishing your family a good day, then heading to a conference before meeting your colleagues for lunch: No matter whether or not this describes your typical morning, it is most likely that you will interact with other people already within the first few hours of your day. Interpersonal communication is an indispensable part of our life and there is much more to it than mere exchange of information. We communicate in various ways and the different media we make use of vary in their potential to deliver rich and comprehensive information (Daft & Lengel, 1986). Written forms of communication, such as letters, e-mails or instant messages, are limited to explicit in-text information, whereas verbal interaction additionally allows for implicit messages through intonation, speaking rate and speech volume. Face to face conversations are the richest, giving further room for non-verbal cues such as posture, gesture, mimic expression and gaze (Bavelas & Gerwing, 2011). While some of these cues can also be effective in video conversations, others, such as gaze, seem more challenging to preserve.

A person's gaze is a meaningful signal that can deliver information about the person's attentional, intentional and emotional state (Schilbach, 2015b; Tomasello et al., 2005b). As such, it has been in the focus of research for decades. Gaze that is directed towards us is perceived as a sign of the other person's interest in us (Cook, 1977) and enhances bodily self-awareness (Baltazar et al., 2014), self-referential processing (J. O. Hietanen & Hietanen, 2017), arousal (Conty et al., 2010) and activity in approach-related brain systems (J. K. Hietanen, Leppänen, Peltola, et al., 2008). Eye contact also leads us to ascribe the other a sophisticated mind (Khalid et al., 2016) and higher trustworthiness (Kaisler & Leder, 2016b). Paradoxically, people who exhibit longer durations of direct gaze are perceived as smarter and more likeable, even though intelligence and personality traits are unrelated to gaze behavior (Harrison et al., 2018; Kuzmanovic et al., 2009; Wheeler et al., 1979).

Though discrete periods of mutual gaze are typically very short, they sum up to about a third of the length of a conversation (Argyle & Ingham, 1972; Kendon, 1967b, 1969) and their effects are manifold and powerful. As a sign of mutual engagement, eye contact is used interchangeably with physical contact in primates (Bard et al., 2005) and constitutes a sign of friendship in human children from six years of age (Nurmsoo et al., 2012). In adults, mutual gaze communicates closeness (Cui et al., 2019) and intimacy (Scherer & Schiff, 1973). Even total strangers are perceived as closer to us in terms of both outward appearances and personality when they directly look at us (Zhou et al., 2018). People feel disconnected from others when strangers fail to make eye contact with them (Wesselmann et al., 2012) and an avoidance of eye contact threatens the needs of belonging, control, meaningful existence and self-esteem (Böckler, Hömke, et al., 2014; K. D. Williams et al., 1998). On the other hand, we tend to avoid other’s gaze when we are ashamed or embarrassed (Exline et al., 1965), and sometimes the norm even requires us to look away. In these situations, direct gaze can be used or perceived as a signal of anger and hostility (Goffman, 2008). Clearly, the effects of direct gaze are far-reaching and likely context dependent.

During a conversation, people look at their partner up to 70% of the time, and they do so more while listening than while talking (Kendon, 1967b, 1969). In face-to-face conversations, direct gaze plays a regulative role and is frequently used to signal a transition from speaking to listening or back (Ho et al., 2015). In video conferences, however, many of the natural functions of direct gaze are distorted. Nevertheless, these tools are becoming ever more popular. In 2019, 59% of employees in the United States reported to use video communication at work and almost 50% stated an increase compared to two years earlier (Lifesize, 2019). During the global SARS-COV2 pandemic that began in late 2019, many of us have made the painful experience of how quickly we can be thrown back onto exclusive virtual interaction. In the Netherlands, the home office share of working hours has increased from about 10% to 60% in late March 2020 (Von Gaudecker et al., 2020) while the buyer activity of web and video conferencing technology has increased by 500% within weeks (Sadler, 2020).

In sum, eye gaze serves important social functions (Baron-Cohen, 1997) and there is evidence that individuals efficiently make use of other people's gaze direction to infer information about their mental state, trustworthiness and the closeness of relationship. However, a critical shortcoming of the majority of previous studies on this topic is the exclusive use of static materials and the neglect of social context variables. Yet, in order to better understand social interaction processes, we need to investigate the interpersonal effects of gaze behavior within more naturalistic and contextually diverse settings. Concerning online and video-based conversations, for example, the effects that gaze distortions may have on person perception and the communicative process remain largely unclear. The present study was designed as a first step to overcome these limitations by testing the effects of eye contact (versus averted gaze) during the observation of an everyday, ecologically valid yet controlled social setting: an online and video-based dyadic conversation. Critically, we also investigated the impact of context variables by manipulating the type of dyadic relationship as well as the emotional valence of the conversation.

To this end, we employed a between-participants design and presented short cuts of video conversations between two people who were allegedly either friends, colleagues or strangers. In these videos, participants could see only one conversation partner (the "target person") and could hear but not see the other conversation partner (the "narrator"). In order to fully engage participants into the social interaction, they were explicitly and repeatedly instructed to step into the shoes of the narrator while watching each video. Critically, the gaze direction of the target person (direct, downwards-averted or mixed) and the emotional context of the conversations (neutral or negative narrations) were manipulated within participants. After each video, participants rated how they perceived the (1) empathy, (2) perspective taking and (3) trustworthiness of the target person, and the (4) closeness of the relationship between the two conversation partners.

Importantly, participants were told that they were witnessing conversations that had been previously held and recorded via video conference tools. A critical difference to real world conversations is that the common applications do not support mutual gaze due to an offset between the camera and the other's eyes on the monitor. We predicted that, despite this commonly known mismatch, a lack of direct gaze in a video call would lead to perceptions of reduced empathy, perspective taking and trustworthiness of the partner as well as to the impression of a less intimate relationship between the conversation partners. In addition, we hypothesized that the role of gaze direction would be less pronounced in negative contexts because, in these situations, gaze avoidance can serve as an aid for regulating one's own emotions (Kendon, 1967b) and a means to give the other space. Hence, we predicted a negative effect of averted gaze on target person evaluation in neutral contexts, but a reduction of this negative effect in emotionally negative conversations. Finally, we assumed that these effects would be stronger in less intimate relationships (e.g., between strangers), because external cues should have a stronger impact on person perception when there is missing knowledge about the other person.

Methods

Participants

We collected data of 54 participants for each relationship condition (friends, colleagues or strangers), resulting in 162 participants in total. All participants of the condition "friends" and twelve participants (22%) of the condition "colleagues" were students that were recruited via SONA systems and compensated with course credit, or volunteers who participated without financial compensation. The remaining participants of the conditions "colleagues" and "strangers" were recruited via Prolific (www.prolific.co) [09/2020] and financially compensated. The number of participants per experiment was a priori determined using G*power3 (Faul et al., 2007b), assuming 80% power and an α of .05 with a small effect size (f = .1) and by considering that the number of participants per experiment had to be a multitude of 18 (see Task and procedure section). Results showed that a total sample of 162 participants with three equal sized groups of N = 54 was required.

There were 29 females and 25 males in the “friends” condition (mean age 31.35, SD = 15.69), 20 females and 34 males for “colleagues” (mean age 27, SD = 9.66) and 24 females and 30 males for “strangers” (mean age 25.3, SD = 6.57). Participants in all experiments were fluent in German. The present study is compliant with the ethical standards of the 1964 Declaration of Helsinki regarding the treatment of human participants in research and was approved by the local ethics committee.

Task and procedure

We report how the sample size was determined, all manipulations and measures that were collected and all data exclusions (Simmons et al., 2012). The data of all experiments are available at DOI: 10.17605/OSF.IO/X7W5A. We used the online software PsyToolkit (Stoet, 2017b) for programming. Participants had access to the online study through their SONA systems account, Prolific or by directly following the link that we shared on our homepage and in social media groups. An HTTP cookie was placed on the device of each participant to prevent repeat participation. The study was accessible both on computers or laptops and on mobile devices, though the use of the former was encouraged to ensure a high video resolution. We also pointed out at the beginning of the experiment that a stable internet connection, a quiet environment and functioning headphones or speakers would be necessary to participate in the study. The participants first completed a set of demographic questions before the instruction screen was presented. We told the participants that they were about to see and listen to short sequences of different conversations during a video call. The participants would perceive these conversations out of the perspective of the interaction partner that they could hear but not see and were asked to empathize with this person during the video. The instructions were similar for all three groups of participants, except for the description of the relationship between the conversation partners: They were described as (lose or close) friends, as (lose or close) colleagues or as strangers. The detailed instructions of each relationship condition can be found in Appendix A.

All participants completed one training trial and 18 test trials. Each trial consisted of a short video (15s) that displayed one of six persons (the "target person") that was either male or female in either early, middle or late adulthood. In every video, the head and torso of the person was presented in front of a light neutral background (Figure 1). The person remained mostly static, did not speak and kept a neutral expression throughout the video, blinking naturally and looking either directly into the camera (*direct gaze* condition), gazing downwards (*averted gaze* condition) or switching between direct and averted gaze every three to four seconds (*mixed gaze* condition with equal parts of direct and averted gaze). Three videos of each of the six persons resulted in a total of 18 videos. These videos were paired with 18 audio files playing back an autobiographical episode that was recounted in German language by one of nine "narrators" who was not visible in the video. For each narrator, there was one neutral story and one story of negative valence, thus entailing experiences of regret, loss or disappointment. Two neutral and two negative example stories in English translation can be found in Appendix B. Crossing all video and audio files resulted in 324 video-audio combinations that were partitioned into 18 groups of 18 trials, each group containing each video and each audio file exactly once in a unique combination. Each group of trials was presented to an equal number of participants in each of the three relationship conditions; assignment was performed in pseudorandomized order.

Figure 1

Screenshots of example videos from the direct and averted gaze conditions

Note. The mixed gaze condition is formed by alternating direct and averted gaze direction.

After each video, participants completed the same set of German questions in each trial, namely: (1) *How empathic was the listener?* (2) *How much did the listener put him- or herself in the position of the narrator?* (3) *How much would you trust the listener?* (4) *How close do you think is the relationship between the narrator and the listener after this conversation?*. Each question was to be answered on a 6-point Likert-scale ranging from *not at all* to *a lot*. Altogether, it took about 20 minutes to complete the experiment.

At the end of the experiment, we asked the participants if they had any assumptions about the goal of the study and whether they encountered any technical difficulties with the video or audio playback. Furthermore, we included two comprehension questions to make sure that the instructions were understood and carried out correctly. Only the data sets of participants who gave correct answers to the comprehension questions and who reported no technical difficulties were included in the analysis. For balancing purposes, we collected data until all 18 trial lists in all three experiments were filled as intended with three participants per list.

Design and Analyses

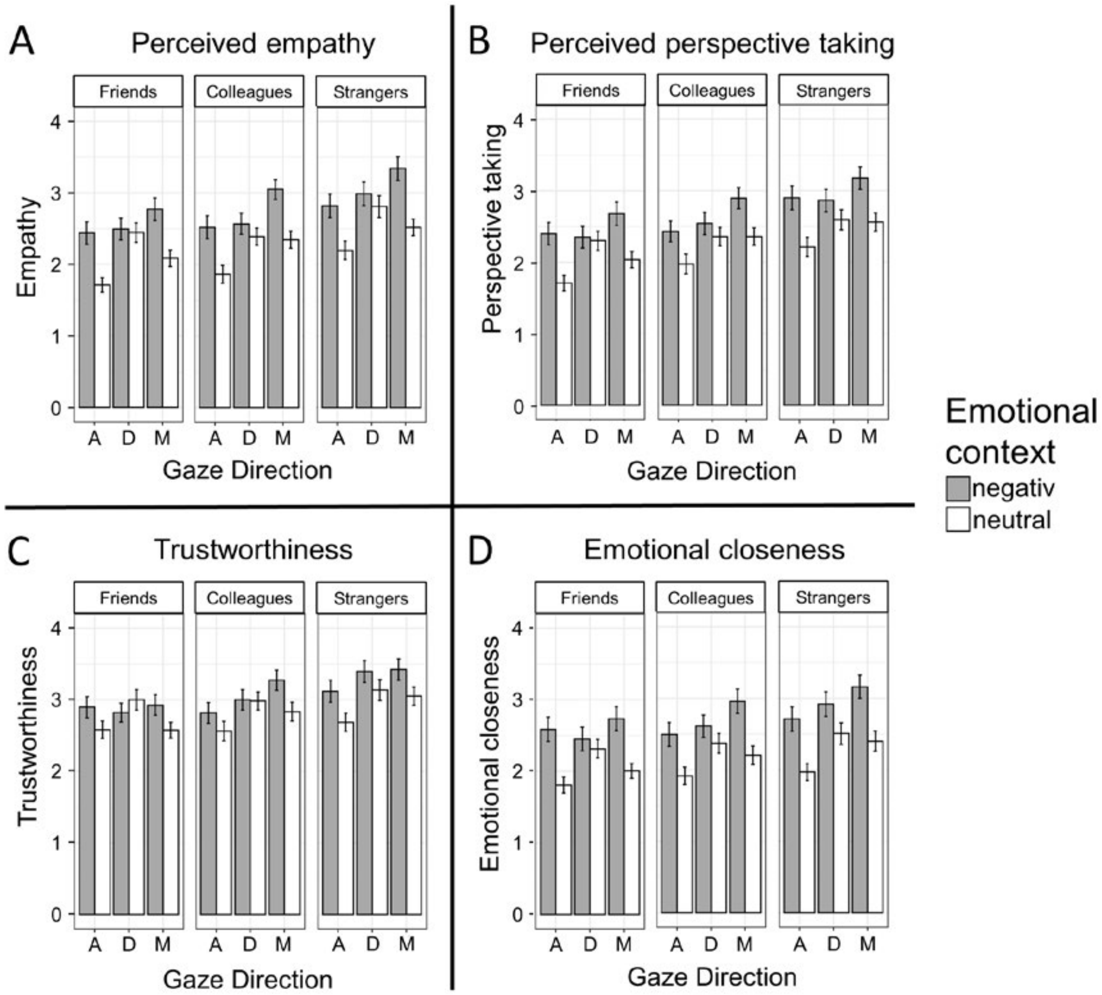
The study followed a 3 × 2 × 3 factorial design with the within-subject factors *gaze direction* (direct, averted, mixed) and *emotional context* (neutral, negative) and the between-subjects factor *relationship* (friends, colleagues, strangers). The rating data of each of the four questions were submitted to a 3 × 2 × 3 mixed effects analysis of variance (ANOVA). Post-hoc t-tests with Bonferroni correction for multiple testing were used to resolve within-subject interaction effects across the three relationship-groups (friends, colleagues and strangers). Detected differences between the groups (relationship conditions) were further investigated with individual two-way ANOVAs with the within-subject factors *gaze direction* (direct, averted, mixed) and *emotional context* (neutral, negative). We report generalized η^2 and Cohen's *d* as effect sizes.

Results

The mean ratings per condition are visualized in Figure 2 and reported in Table C1 in Appendix C.

Figure 2

Mean ratings for each relationship condition.



Note: Mean ratings per condition on a 6-point scale with separate graphs for each relationship condition and with gaze direction on the x-axis (A = averted gaze, D = direct gaze, M = mixture of averted and direct gaze). *Panel A*: Mean ratings of perceived empathy of the person in the video. *Panel B*: Mean ratings of perceived perspective taking of the person in the video. *Panel C*: Mean trustworthiness ratings of the person in the video. *Panel D*: Mean ratings of emotional closeness of the conversation partners.

Empathy

Across conditions, the participants rated the empathic responding of the target persons with 2.52 out of 6 points (SD = 1.10). The mean ratings of empathy per condition are visualized in panel A of Figure 2.

Gaze direction had a significant effect on perceived empathy ($F(2,318) = 30.132$, $p < .001$, $\eta_{gen}^2 = .032$). Post hoc *t*-tests indicated that the target persons were perceived as more empathic when direct or a mixed form of gaze was displayed (direct vs. mixed: $p = .707$; direct vs. averted: $t(323) = 5.429$, $p = .001$, $d = 0.328$; mixed vs. averted: $t(323) = 9.14$, $p < .001$, $d = 0.392$). The main effect of emotional context was significant due to higher empathy ratings when a negative compared to a neutral story was told ($F(1,159) = 117.008$, $p < .001$, $\eta_{gen}^2 = .059$). There was also a significant main effect of relationship ($F(2,159) = 4.367$, $p = .014$, $\eta_{gen}^2 = .033$). Post hoc comparisons indicated that perceived empathy was higher when participants observed strangers compared to friends or colleagues (friends vs. strangers: $t(323) = -4.947$, $p < .001$, $d = -0.412$; colleagues vs. strangers: $t(323) = -3.835$, $p < .001$, $d = -0.294$) but not significantly different between colleagues and friends ($p = .417$).

Critically, the gaze direction × emotional context interaction effect was significant ($F(2,318) = 24.840$, $p < .001$, $\eta_{gen}^2 = .017$). Post hoc tests indicated that emotional context had no effect on empathy ratings when the target person constantly established direct gaze ($p = .192$). In contrast, ratings were lower during neutral compared to negative stories when the target person had averted gaze ($t(161) = 9.420$, $p < .001$, $d = 0.642$) or mixed gaze ($t(161) = 10.412$, $p < .001$, $d = 0.721$). Note that, during emotional videos, ratings of empathy were even higher for mixed than for direct gaze ($t(161) = -4.4973$, $p < .001$, $d = -0.322$).

No other effects were significant (relationship × gaze direction: $p = .480$; relationship × emotional context: $p = .106$; relationship × gaze direction × emotional context: $p = .772$).

Perspective-taking

The global mean rating of perceived perspective-taking of the target person in the videos was 2.46 (SD = 1.10). The mean ratings per condition are visualized in panel B of Figure 2.

As for empathy, there were significant main effects for gaze direction ($F(2,318) = 17.210$, $p < .001$, $\eta_{gen}^2 = .019$), emotional context ($F(1,159) = 86.101$, $p < .001$, $\eta_{gen}^2 = .047$) and relationship ($F(2,159) = 4.310$, $p = .015$, $\eta_{gen}^2 = .034$). Post hoc tests revealed that ratings were higher when direct or a mixed form of gaze were pursued by the person in the video (direct vs. mixed: $p = .152$; direct vs. averted: $t(323) = 3.523$, $p = .002$, $d = 0.210$; mixed vs. averted: $t(323) = 7.913$, $p < .001$, $d = 0.316$) as well as when he or she was listening to a negative story. Target persons were also rated as more prone to perspective-taking when they were allegedly unacquainted to the speaker, but ratings were not reliably different between friends and colleagues (friends vs. colleagues: $p = .129$; friends vs. strangers: $t(323) = -5.178$, $p < .001$, $d = -0.433$; colleagues vs. strangers: $t(323) = -3.405$, $p = .002$, $d = -0.268$).

Again, our results indicate that the effect of gaze direction during a conversation depended on the emotional content of the narration as evident by a significant gaze direction × emotional context interaction effect ($F(2,318) = 19.020$, $p < .001$, $\eta_{gen}^2 = .010$). Post hoc tests indicated that averted or mixed gaze appeared to be accepted during negative conversations but led to lower ratings of perspective-taking when a neutral story was told (averted: $t(161) = 8.524$, $p < .001$, $d = 0.573$; mixed: $t(161) = 8.926$, $p < .001$, $d = 0.576$). No reliable effect was found for direct gaze ($p = .104$).

No other effects were significant (relationship × gaze direction: $p = .806$; relationship × emotional context: $p = .589$; relationship × gaze direction × emotional context: $p = .397$).

Trustworthiness

On average, the mean trustworthiness of the people that were seen during the videos was 2.95 (SD = 1.04). How much, on average, the participants would have trusted the persons in the videos is visualized in panel C of Figure 2.

We found a main effect of gaze direction ($F(2,318) = 18.704$, $p < .001$, $\eta_{gen}^2 = .015$) with higher ratings of trustworthiness when the person in the video was looking into the camera all the time or most of the time (t-tests: direct vs. mixed: $p = .999$; direct vs. averted: $t(323) = 5.231$, $p < .001$, $d = 0.270$; mixed vs. averted: $t(323) = 5.972$, $p < .001$, $d = 0.231$). The effect of emotional context was significant, indicating that trustworthiness was enhanced by negative story content ($F(1,159) = 36.383$, $p < .001$, $\eta_{gen}^2 = .015$).

As for empathy and perspective-taking, a significant gaze direction × emotional context interaction effect indicated that the trustworthiness of the target person depended on an interplay between gaze direction and emotional context ($F(2,318) = 10.462$, $p < .001$, $\eta_{gen}^2 = .006$). The trustworthiness of the target person was lower for neutral compared to negative narrations when they looked down sometimes (mixed: $t(161) = 6.414$, $p < .001$, $d = 0.386$) or all the time (averted: $t(161) = 5.371$, $p < .001$, $d = 0.327$), whereas no significant difference was found for direct gaze patterns ($p = .999$).

There was also a significant relationship × gaze interaction effect ($F(2,318) = 10.462$, $p < .001$, $\eta_{gen}^2 = .004$). Separate two-way ANOVAs for each relationship group with the within-subject factors gaze direction and emotional context revealed that the effect of gaze direction on trustworthiness was significant in all conditions (friends: $F(2,106) = 3.289$, $p = .041$, $\eta_{gen}^2 = .007$; colleagues: $F(2,106) = 12.433$, $p < .001$, $\eta_{gen}^2 = .025$; strangers: $F(2,106) = 8.096$, $p = .001$, $\eta_{gen}^2 = .024$). In order to further test for differences in gaze effects between groups, we performed the initial three-way ANOVA with the within-subject factors gaze direction and emotional context and the between-subjects factor relationship again, entering the data of only two conditions at a time. The relationship × gaze direction interaction effect collapsed when the data of the condition “friends” was dropped (*p* = .760), but remained significant between friends and colleagues ($F(2,318) = 5.383$, $p = .005$, $\eta_{gen}^2 = .005$) as well as between friends and strangers ($F(2,318) = 3.316$, $p = .038$, $\eta_{gen}^2 = .004$), indicating that the effect of gaze direction on trustworthiness ratings was smaller for friends compared to colleagues and strangers.

No other effects were significant (relationship: *p* = .122; relationship × emotional context: *p* = .170; relationship × gaze direction × emotional context: *p* = .212).

Emotional closeness

The emotional closeness between the narrator and the target person was rated with 2.45 (SD = 1.14). The data per condition is visualized in panel D of Figure 2.

As for all other ratings, there was a significant main effect of gaze direction ($F(2.318) = 18.923$, $p < .001$, $\eta_{gen}^2 = .018$). Ratings of emotional closeness were lower for averted gaze (direct vs. averted: $t(323) = 4.13$, $p < .001$, $d = 0.251$; mixed vs. averted: $t(323) = 7.158$, $p < .001$, $d = 0.292$), but were not significantly different between direct and mixed gaze (*p* = .999). The main effect of emotional context was significant ($F(1,159) = 105.099$, $p < .001$, $\eta_{gen}^2 = .065$), indicating that emotional closeness was perceived as higher during negative stories.

The interaction effect of gaze direction × emotional context was once more significant ($F(2,318) = 17.113$, $p < .001$, $\eta_{gen}^2 = .010$). Post hoc tests revealed that, during neutral stories, emotional closeness ratings were highest for direct and lowest for averted gaze (direct vs. averted: $t(161) = 6.7963$, $p < .001$, $d = 0.5322$; direct vs. mixed: $t(161) = 3.0128$, $p < .001$, $d = 0.198$; averted vs. mixed: $t(161) = -5.3119$, $p < .001$, $d = -0.343$). However, during negative stories, ratings were not significantly different for direct and averted gaze (p > .999) but increased for mixed gaze (direct vs. mixed: $t(161) = - 3.5631$, $p < .001$, $d = -0.238$; averted vs. mixed: $t(161) = -4.9063$, $p < .001$, $d = -0.289$).

The rating of emotional closeness was not affected by the relationship of the conversation partners (relationship: p = .213; relationship × gaze direction: p = .388; relationship × emotional context: p = .680; relationship × gaze direction × emotional context: p = .519).

Discussion

Though human interaction and personal contact are the backbone of our lives as social creatures, an increasing amount of interpersonal communication takes place online, supported by video communication software. While some visual signals that are crucial in face-to-face interaction can be preserved, gaze direction in particular is difficult to validly capture during video conversations. In the present study, we probed the role of gaze direction in different scenarios of video-based social interactions. Specifically, we prepared video conversations between friends, colleagues or strangers in which the target person exerted direct, downwards-averted or a mixed form of gaze while listening to a narration with neutral or negative emotional content (Figure 1). Participants were asked to watch the videos from the perspective of the invisible narrators during the conversation in order to subsequently rate their impression of the displayed target person in terms of empathy, perspective taking and trustworthiness as well as the emotional closeness of their relationship.

Our results clearly indicate that gaze direction during a conversation has profound effects on the impression one makes on one's conversation partner with (partly) direct gaze being more favorable. In all three relationship conditions, participants attributed more empathy and perspective taking to persons who (at least occasionally) gazed into the camera during a video call. Indeed, this might be a valid inference: Previous studies suggest that individuals who fixate the eyes of others longer also score higher in a standardized questionnaire of empathy and perspective taking (Cowan et al., 2014; Martínez-Velázquez et al., 2020). Other studies suggest that scores on these questionnaires seem to in fact be positively related to actual emotion recognition performance (Israelashvili et al., 2019). In a similar vein, the person who was seen in the video appeared more trustworthy when they frequently employed direct gaze. This pattern fits previous studies reporting that people who make eye contact are trusted more (Bekkering & Shim, 2006) and that people who gaze less are more likely to be perceived as deceptive (Zuckerman et al., 1981). Ironically, precisely because of this commonly held assumption, people deliberately make more eye contact when they are trying to deceive (Burns & Kintz, 1976; Mann et al., 2013). Finally, the relationship between the conversation partners was rated as closer when eye contact was established (occasionally). Since we look more at other people if we like them (Argyle & Dean, 1965), frequent eye contact could indeed be a reliable indicator for emotional closeness. However, we also ascribe positive attributes to people who frequently look at us: They appear more familiar to us (Taubert et al., 2017) and are perceived as closer and more similar to ourselves (Zhou et al., 2018).

Taken together, our findings on effects of gaze in social interactions are well in line with previous studies that all seem to lead to the same conclusion: looking at one's conversation partner signals interest and affiliation while looking away constitutes a sign that the partner is trying to hide something or finding the conversation strenuous (Bekkering & Shim, 2006; Kendon, 1969). In line with that, people who permanently looked downwards in our study reliably suffered losses at all four ratings.

Critically however, the nature of the effects of gaze direction depended fundamentally on the emotional context of the conversation, hence, on the story the target person was allegedly listening to. While participants reliably responded with lower ratings to averted gaze during neutral stories, avoiding eye contact while negative story content was shared was widely accepted. In the latter case, ratings were largely similar to those after constant direct gaze in all outcome measures, from empathy to relationship closeness. Interestingly, however, the highest ratings were given when the target person switched between direct and averted gaze during a negative story. In these arousing and potentially delicate situations, gaze avoidance could have been interpreted as a sign of high compassion and the resulting need to down-regulate one's own emotions (Kendon, 1967b) or giving the person who is recounting a difficult personal episode some space. By contrast, during neutral stories, frequent or permanent gaze avoidance might have been considered a sign of boredom, indifference or even deception.

Note that despite the striking difference in ratings, the videos that we showed during neutral and during negative stories were exactly the same. Even though participants believed to witness snippets of an allegedly longer video conference, the videos were not really related to the stories but randomly paired and counterbalanced across participants. Furthermore, these video-audio-combinations were the same in all three relationship conditions. The only thing changing between experiments was the description of the relationship between the conversation partners.

Overall, ratings were largely unaffected by the alleged relationship between the conversants with two exceptions: First, ratings of perceived empathy and perspective taking were highest when the participants were told that the conversation partners were unacquainted and lower when they were purportedly friends or colleagues. The mere fact that people who do not know each other share private matters in a video conversation might have been taken as indication for high levels of empathy and perspective-taking in the target person, irrespective of gaze direction or story matter. Second, the effect of gaze direction on trustworthiness was stronger when the conversation partners were labelled as colleagues or strangers rather than friends. Hence, when witnessing a conversation between less well-known partners, participants may have relied more on gaze cues to infer whether the other could be trusted. In contrast, when friendship was (believed to be) established already, mutual trust was taken for granted and gaze direction became less important as an indicator of trustworthiness. Interestingly, the size of this effect did not vary in parallel with the level of acquaintance of the conversation partners. Instead, the impact of gaze direction on trustworthiness ratings was similar between colleagues and strangers but significantly smaller for friends. We are aware that the different recruitment strategies in our study constitute a potential confound with the relationship manipulation and, hence, we only cautiously interpret these between-group differences. Note, however, that the effects do not mirror differences in recruitment strategy. Ratings of empathy and perspective taking of colleagues were different to those of strangers, but the same as those of friends, even though the cohorts of the conditions "colleagues" and "strangers" were more similar to each other. Furthermore, the largest gaze effect was found for conversations between colleagues, which is the one condition in which both recruitment strategies were combined.

One limitation of our study is the exclusive use of subjective measures. While they provide the best estimation of individual perception and the consistency of results across ratings and relationship conditions speaks for the reliability of our findings, it would be interesting to test whether the present effects of gaze on subjective impression translate onto actual interpersonal behavior. Ideally, future studies could incorporate a broader range of story content to investigate how gaze effects are influenced by positive context.

In sum, we find that eye contact is an important tool during social interactions to signal attention, care and positive intentions. Violating the convention of establishing eye contact during a conversation *can* make a bad impression, but in certain cases it does not. Averting gaze while listening to a negative narration seems to be widely accepted and can, in correct dosage, even communicate social understanding and trustworthiness. Our data fit and extent previous findings that direct gaze is an important, yet context-dependent cue and that gaze behavior during a conversation is used as a heuristic to infer personal and interpersonal information. Furthermore, we demonstrate that gaze direction, even though it is a basic social signal with immediate effects on attentional and socio-cognitive processes (e.g. Senju & Johnson, 2009), is not perceived and interpreted in isolation. Rather, effects of gaze direction seem to be susceptible to context variables and top-down modulation processes. This notion is in line with and extends previous studies probing the context sensitivity and boundary conditions of gaze cues (e.g. Böckler et al., 2015; J. Hietanen & Yrttimaa, 2005; for an overview, see Frischen et al., 2007) and has important implications for social attention and interaction, because it enables basic social signals such as gaze direction to be flexibly and adequately interpreted. Depending on the affective and cognitive states of (potential) interaction partners, their nonverbal (and potentially also verbal) behavior is interpreted differentially. Further investigating the degree and the ways in which basic social signals and context information are integrated during interaction will help gain a better understanding of how we perceive and judge others in different situations and, as a consequence, how we behave towards them.

In this light, the fact that most video-communication applications do not support eye contact during video conferences appears especially problematic. Humans are extremely skilled at perceiving direct gaze and are very sensitive to when their partner looks away from the camera. Already small deviations of less than 10° are sufficient to lose the perception of eye contact (M. Chen, 2002). However, in usual video conference set-ups, the camera is placed above the monitor and, hence, a person who is actually looking at the eyes of the conversation partner appears to be looking downwards (Bekkering & Shim, 2006; Loh & Redd, 2008). At least the perception of eye contact can be replicated by gazing into the camera instead of watching the video conversation partner on the monitor. However, this comes with the drawback of not being able to read the other’s face or their reaction. Therefore, an important recent trend has been the development of gaze correction techniques that enable eye contact during video conferences within milliseconds (Kononenko & Lempitsky, 2015; Wolf et al., 2010; Wood et al., 2018). Critically, our new finding of a modulation of gaze effects by emotional context suggests that there are situations in which masking averted gaze might be quite unfavorable and counterproductive. What we really need is a solution that mirrors actual gaze behavior without concealing or distorting any aspect of it.

Acknowledgements

This work was supported by the German Research Foundation under Grant number BO 4962/1-1. We thank two anonymous reviewers for their constructive feedback that helped us to strengthen our manuscript.

MANUSCRIPT 5

GOLDEN GAZES: GAZE DIRECTION AND EMOTIONAL CONTEXT PROMOTE PROSOCIAL BEHAVIOR BY MODULATION PERCEPIONS OF OTHER'S EMPATHY AND PERSPECTIVE-TAKING

Christina Breil[1], Leticia Retore Micheli[1], Anne Böckler[1,2]

[1]Julius-Maximilians-University of Würzburg, Würzburg, Germany

[2]Max-Planck-Institute for Human Cognitive and Brain Science, Leipzig, Germany

This manuscript is unpublished.

Abstract

Prosocial behavior is a fundamental element of cooperation in large societies. We investigate the impact of gaze behavior and emotional context on perceptions of other people's social cognition and on prosocial decision making. In two experiments, participants witness short pre-recorded conversations between two people and subsequently rate their impression of empathy and perspective-taking of one of the conversation partners (the target person). Subsequently, participants play the Trust Game (experiment 1) or the Dictator Game (experiment 2) with this person. We replicate findings that occasional gaze avoidance, especially during emotionally negative conversations, increases perceived social understanding of the target person. We extend these results by showing that individuals who are perceived as highly empathic and taking perspective are ultimately treated with more trust and generosity in subsequent strategic and non-strategic economic games. We conclude that non-verbal cues in social interactions can serve as an indication of another's reputation, thereby promoting indirect reciprocity to stabilize cooperation in large societies.

Keywords: social interaction, gaze behavior, empathy, perspective-taking, trust, prosociality

Introduction

Lending a hand to assist friends, helping strangers on the street by giving directions, or even donating money, humans are remarkably prosocial. However, acting in ways that benefit others often requires individuals to exert effort (Lockwood et al., 2017) and to bear financial and moral costs (Crockett et al., 2014; Engel, 2011). Thus, individuals usually have to be selective in deciding who they want to help or cooperate with. Evolutionary psychology predicts that individuals are more altruistic towards close family members and friends to enable cooperation in hunter-gatherer societies (W. D. Hamilton, 1964). But how is the decision about with whom to cooperate made in an ever-growing society where we increasingly interact with unrelated and unknown individuals? Social interactions are extremely rich in information and individuals may rely on a variety of verbal and non-verbal cues to make interpersonal decisions (Behrens et al., 2020). Here, we investigate the role of subtle context cues in promoting prosocial behavior.

Gaze direction – where people look during social interactions – is a non-verbal cue which has been associated with increased prosocial behavior. Partners in an economic game who are allowed to look at each other cooperate more than dyads of individuals who cannot establish mutual gaze (Behrens et al., 2020). Likewise, children share more with others who occasionally look at them (Wu et al., 2018). One potential explanation for the influence of gaze on prosocial behavior in interactive dyads is that eye contact enables the transmission of emotional states. Previous research has shown that when individuals look at each other, their physiological states (for example as measured by pupil dilation) tend to synchronize (Fawcett et al., 2016), suggesting that individuals are to a certain level aware and able to mimic others' internal states.

Individuals may also decide with whom to cooperate by observing others interact. Nevertheless, the influence of gaze direction on prosocial behavior has not yet been investigated in third-party contexts. Understanding how the non-verbal demeanor of individuals during a social interaction might affect the behavior of observers can be relevant considering that individuals are often exposed to the interactions of other people and even judge or choose to intervene in these situations. For example, individuals are more inclined to be prosocial towards those who have just done a good deed (Wedekind & Milinski, 2000) and to reward those who were fair with others (Almenberg et al., 2011). Given that verbal behavior of third-parties can affect the behavior of observers, one might expect a similar role from non-verbal behavior. Accordingly, the gaze of individuals during social interactions can provide important information, ultimately influencing how they are perceived by others. Social perception, in turn, can play an important role in determining how nicely others are treated. For instance, the attribution of characteristics such as “warmth” to individuals has been shown to increase prosocial behavior towards them (Jenkins et al., 2018).

Similarly, individuals who frequently engage in eye contact during conversations might be perceived as more capable of emotional connection and more willing to tend to their conversation partners’ needs. In fact, gaze direction has been shown to efficiently signal one’s attention, interest and intentions (Schilbach, 2015a; Tomasello et al., 2005a). Previous research has shown that faces showing direct gaze are considered more trustworthy (Breil & Böckler, 2021; Kaisler & Leder, 2016a). In addition, in a conversation, listeners who frequently engage in direct gaze as opposed to avoiding eye contact are judged by third-parties to be more empathic and to have a better understanding of the speakers’ mental states (Breil & Böckler, 2021). As previous literature has demonstrated a positive association between social understanding traits and prosocial behavior (FeldmanHall et al., 2015; Leiberg et al., 2011), eye contact might be considered an important signal of one’s prosocial inclinations towards others by modulating perceptions of one’s social cognition. As such, it may be reciprocated with generous behavior.

Interestingly, the influence of gaze direction on social perception appears to be dependent on yet another non-verbal cue: The emotional context of the conversation (Breil & Böckler, 2021). In this study, participants witnessed short pre-recorded conversations between two people: a speaker, who was invisible in the video, and a listener, who responded only with gaze behavior to the speaker's story. While reduced eye contact during emotionally neutral conversations was perceived by observers as a signal of listeners' low empathizing and perspective-taking, this was not the case during emotionally negative conversations. Listeners who avoided the speakers' gaze while hearing an emotionally negative story did not receive lower ratings of social understanding than those who engaged in more frequent eye contact (Breil & Böckler, 2021). Thus, the influence of listeners' gaze direction on the prosocial behavior of third-parties might be modulated by the emotional context of the social interaction.

Across two experiments (one preregistered) we investigate whether the gaze direction of listeners during an emotionally negative or neutral conversation influences the prosocial behavior they are met with subsequently. Specifically, we expect that the listeners' gaze direction as well as the emotional context of the conversation might have an effect on how prosocial participants are towards the listener in a subsequent interaction. In experiment 1, we employ the Trust Game (Berg et al., 1995b) to test whether gaze and emotional context influence observers' trust in the listener. In experiment 2, we aim to expand the findings to another domain of prosocial behavior, namely giving behavior in the Dictator Game. In addition, we expect that the effect of these non-verbal cues on participants' prosocial behavior towards the listener is mediated by their perceptions of the listeners' capacity to empathize with and take the perspective of the speaker.

Experiment 1

Methods

Participants

We recruited 162 participants (M_{age} = 27.8 years, *SD* = 8.4, 41% female, 59% male) from an online panel (https://prolific.co/). Only participants who were fluent in German, did not have hearing impairments and disposed of a functioning audio-visual system in their devices were invited to take part in the experiment. In addition, because the experiment involved deception, only participants who had previously expressed willingness to take part in experiments with deception via Prolific were invited to participate in the study. The experiment lasted for approximately 15 minutes and participants received a flat fee of 2.40€ for their participation.

The number of participants was determined a priori using G*Power (Faul et al., 2007a). We based our power calculation on the results of a previous experiment with a similar study design showing small effect sizes of gaze behavior and emotional context on trust perceptions (Breil & Böckler, 2021). We assumed 80% of power and an α of .05 to detect a small effect size (f = .1). We also considered that the number of participants needed to be a multiple of 54 (see *Procedure* section).

Procedure

The experimental design was adapted from (Breil & Böckler, 2021) and the experiment was programmed using the online software PsyToolkit (Stoet, 2017a). Participants were encouraged to use a computer or a laptop to complete the task in order to ensure a proper visualization of the video stimuli. Participants were also instructed to ensure they had a stable internet connection and working headphones or speakers before starting the experiment.

Upon consenting to take part in the experiment and answering demographic questions, participants were told that they would witness six trials of short video-based encounters between two other people, namely a speaker and a listener (henceforth also referred to as *target person*). In each episode, participants could hear (but not see) the speaker narrating an autobiographical episode and see a video-record of the target person, remaining mostly static and allegedly listening to the narration. Participants were instructed to take the perspective of the speaker while observing each encounter, after which they were asked to answer two questions regarding their perceptions of the target person and subsequently have the opportunity to interact in an economic game with this person. To ensure a correct understanding of the task, after reading the instructions, participants were asked to answer four multiple-choice comprehension questions. If an incorrect response was provided, participants saw a pop-up message explaining why their answer was incorrect and were given a second chance to answer correctly. Only participants who answered all questions correctly were allowed to move forward with the experiment.

In total, participants completed one training trial and six test trials. In each trial, participants could hear a different speaker recounting either a neutral or a negative autobiographical episode. The audio narrations used in the experiment were taken from the validated EmpaToM task (Kanske et al., 2015a). While neutral stories mostly revolved around mundane events (e.g., daily routine after coming home from work), negative stories evoked experiences of losses or disappointment (e.g., sick family member) and have been shown to elicit considerably more negative affect than neutral stories (Breil, Kanske, et al., 2021; Kanske et al., 2015a). An audio transcript exemplifying a negative autobiographical episode used in the experiment can be found at the project folder in OSF (https://osf.io/pjzh7/).

Each narration was paired with a short video (~15s) displaying the target person. In every trial, participants could see the head and torso of a target person in front of a light neutral background. The target person did not speak and kept a neutral facial expression while remaining mostly static for the duration of the video. In each trial, participants saw a different target person who would either constantly look directly into the camera (*direct gaze* condition), constantly look downwards (*averted gaze* condition) or intermittently switch between direct and averted gaze for the duration of the video (*mixed gaze* condition). Videos used in the experiment were taken from (Breil & Böckler, 2021) and an example can be found at OSF.

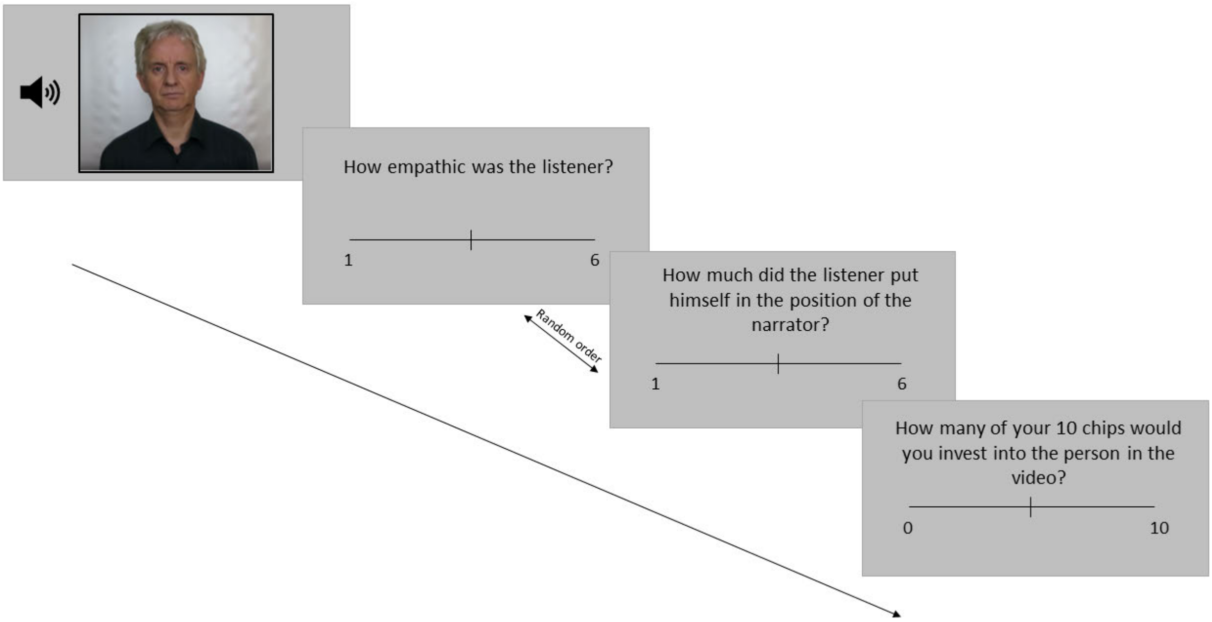
To prevent characteristics of the target person (e.g., gender, age) or the specific content of the autobiographical episode from influencing participants' perceptions and decisions, 18 audios from nine different speakers (each recounting one neutral and one negative story) and 18 videos from six different target persons (one video portraying direct gaze, one mixed gaze and one averted gaze) were counterbalanced so that each audio and video were paired equally often across participants. This resulted in 54 stimuli lists containing six trials each. Participants were randomly assigned to one stimuli list in a way that they were exposed only once to a given speaker and target person across the six conditions (direct-neutral, direct-negative, averted-neutral, averted-negative, mixed-neutral, mixed-negative). For balancing purposes, data was collected until each list was completed by three participants.

In each trial, after observing the encounter, participants were asked to answer two questions regarding their perceptions of the target person. Specifically, participants were asked to answer one item measuring how empathic the target person was perceived during the encounter ("how empathic was the listener?") and one item measuring how much the participant thought that the target person was able to take the perspective of the speaker ("how much did the listener put him or herself in the position of the narrator?"). The order of presentation of both questions was randomized and each question was answered on a 6-point Likert scale ranging from "not at all" to "a lot" (Figure 1)

Next, to investigate how gaze and emotional context play a role on prosocial behavior, participants adopted the role of the investor in a one-shot Trust Game (TG) with the target person as the trustee in each trial. They were endowed with 10 chips and asked how much (if any) they would like to invest in the target person. Participants could enter their answers on a scale ranging from 0 to 10 chips with increments of .25 chips. They were told that the amount of chips sent to the target person would be tripled and that the target person would have the chance to return a freely selected amount of this increased pot of money back to them. Importantly, similar to the participant, the target person did not need to send any chips to their partner in the Trust Game.

Figure 1

Trial outline of experiment 1



Note. Participants witnessed an encounter in which they heard a speaker narrating either a neutral or negative autobiographical episode to a target person who displayed either direct, averted or mixed gaze in a video. After each encounter, participants were asked to rate the target person in terms of empathy and perspective-taking displayed during the conversation. Participants then played a one-shot Trust Game with the target person and were asked to decide how many of their 10 chips they would like to send to the target person. After imputing their responses, a new trial began.

Participants were informed that, at the end of the experiment, one trial would be randomly selected and they would be paid according to the sum of chips they kept for themselves plus the amount of chips the target person returned to them in that round (with 1 chip = 0.2€). Unbeknownst to participants, the target person was not part of the experiment and thus was not informed of the participants' decision in the TG. All participants were paid the full corresponding amount of 10 chips (2€) in addition to the flat fee.

At the end of the experiment, participants were asked whether they (i) believed the instructions of the experiment, (ii) had any assumptions about the goals of the study and (iii) encountered any technical issues with the video and audio playback. Participants were then debriefed and thanked.

Data Analysis

We employed a 2 (emotional context: negative or neutral) × 3 (gaze: direct, averted or mixed) within-subjects design. All analyses were conducted using the R software for statistical analyses (R Core Team, 2013). We first analysed the effects of emotional context and gaze direction on participants' perceptions of the target person's ability to empathize with the speaker and take their perspective using a 2 × 3 repeated measures analyses of variance (ANOVA). The same analysis was conducted to investigate the effect of emotional context and gaze direction on participants' willingness to trust the target person in the TG. Bonferroni correction for multiple testing was applied to subsequent pairwise comparison tests. For all analyses, we report generalized η^2 or Cohen's *d* as effect sizes. Non-significant interaction effects were further investigated with Bayesian ANOVAs and interpreted as evidence for H1 when BF10 > 3s (Jarosz & Wiley, 2014). Finally, to investigate the psychological mechanisms underlying a potential effect of emotional context and/or gaze direction on participants' trust behavior, we conducted a multiple mediation analysis using the package *lavaan* in R (Rosseel, 2012).

Results and Discussion

Empathy ratings

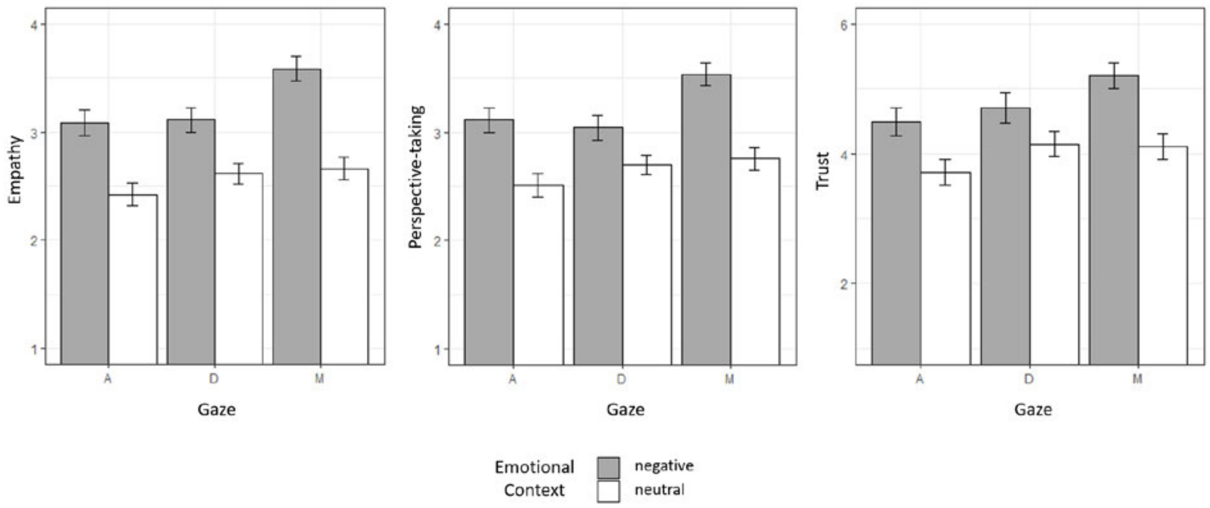
ANOVA results showed a main effect of gaze direction on perceived empathy of the target person (F(2,322) = 7.69, p = .001, η^2_{gen} = .012). Post-hoc t-tests revealed that target persons displaying mixed gaze were considered more empathic than those displaying averted gaze (t(323) = 4.01, p < .001, d = .254; M_{mixed} = 3.13, SD = 1.38, $M_{averted}$ = 2.76, SD = 2.84) and direct gaze (t(323) = 2.67, p = .024, d = .183; M_{direct} = 2.87, SD = 1.36). There were no significant differences between empathy perception of target persons displaying direct and averted gaze (t(323) = 1.17, p = .727, d = .078). We also observed a main effect of emotional context on perceived empathy of the target person (F(1,161) = 56.8, p = .001, η^2_{gen} = .059), such that those who were confided a negative story were rated as more empathic than those who were told a neutral story by the speaker ($M_{negative}$ = 3.26, SD = 1.47; $M_{neutral}$ = 2.57, SD = 1.30). The interaction between gaze and emotional context was not significant (F(2,322) = 2.79, p = .063, η^2_{gen} = .004). This is supported by a Bayesian ANOVA giving substantial evidence for H0 (BF10 = 0.255). The mean empathy ratings per condition are shown in Figure 2, left panel).

Perspective-taking ratings

The mean perspective-taking ratings per condition can be seen in Figure 2, middle panel. ANOVA results revealed a similar pattern compared to empathy ratings. We found a main effect of gaze direction on perceived perspective-taking of the target person (F(2,322) = 6.98, p = .001, η^2_{gen} = .012). Post-hoc t-tests showed that target persons displaying mixed gaze were rated higher in perspective-taking than those displaying averted (t(323) = 3.8, p < .001, d = .240; M_{mixed} = 3.15, SD = 1.30, $M_{averted}$ = 2.81, SD = 1.43) and direct gaze (t(323) = 2.98, p = .003, d = .204; M_{direct} = 2.87, SD = 1.32). There were no significant differences between target persons displaying averted and direct gaze (t(323) = -.67, p > .99, d = .044). We also found a main effect of emotional context, indicating that those who listened to negative stories were rated higher in perspective-taking than those who listened to neutral stories ($F(1,161) = 45.06$, $p < .001$, $\eta^2_{gen} = .044$; $M_{negative} = 3.23$, SD = 1.41; $M_{neutral} = 2.65$, SD = 3.88).

Figure 2

Results per gaze condition and emotional context (Experiment 1)



Note. The figure displays how participants rated the target person regarding their capacity to empathize with the speaker (left panel) and take their perspective (middle panel) on a 6-point scale. Higher ratings indicate higher capacity for empathy and perspective taking. The right panel shows the average chips (from 0 to 10) participants invested in the target person in the Trust Game. The x-axis shows the gaze conditions (A: averted gaze; D: direct gaze; M: mixed gaze).

Differently from the empathy ratings, however, we found a significant interaction between gaze direction and emotional context on perspective-taking ratings ($F(2,322) = 3.36$, $p = .036$, $\eta^2_{gen} = .004$). Post-hoc tests indicated that differences between mixed and averted gaze, as well as between mixed and direct gaze were only significant when the target person was told a negative story by the speaker (mixed vs. averted gaze: $t(161) = 3.13$, $p = .012$, $d = .341$; mixed vs. direct gaze: $t(161) = 3.4$, $p = .005$, $d = .327$), whereas no differences were found between the different gaze conditions when the target person heard a neutral story (mixed vs. averted gaze: $t(161) = 2.18$, $p = .18$, $d = .180$; mixed vs. direct gaze: $t(161) = .5$, $p > .99$, $d = .045$). Differences between direct and averted gaze were non-significant in both emotional contexts (Neutral: $t(161) = -1.61$, $p = .11$, $d = .148$; Negative: $t(161) = .48$, $p > .99$, $d = .046$).

Investment in the TG

The mean investment on the target person per condition in the TG is displayed in Figure 2, right panel. Across conditions, participants invested 4.4 (SD = 2.63) chips in the target person. We found a main effect of gaze direction on the amount sent to the target person in the TG ($F(2,322) = 5.92$, $p = .003$, $\eta^2_{gen} = .007$). Post-hoc *t*-tests showed that target persons displaying mixed gaze were trusted more with a higher investment in the TG than target persons displaying averted gaze ($t(323) = 3.78$, $p < .001$, $d = .210$; $M_{mixed} = 4.66$, SD = 2.52, $M_{averted} = 4.1$, SD = 2.66). No differences in investment were found between mixed and direct gaze ($t(323) = 1.49$, $p = .41$, $d = .087$; $M_{direct} = 4.43$, SD = 2.715) and direct and averted gaze ($t(323) = 1.99$, $p = .14$, $d = .119$). Similar to the results for empathy and perspective-taking ratings, we also found a main effect of emotional context indicating that target persons who heard a negative story were trusted more in the TG than those hearing a neutral story ($F(1,161) = 26.66$, $p < .001$, $\eta^2_{gen} = .023$; $M_{negative} = 4.8$, SD = 2.75; $M_{neutral} = 3.99$, SD = 2.51). The interaction between gaze direction and emotional context was not significant ($F(2,322) = 1.67$, $p = .19$, $\eta^2_{gen} = .002$). As for empathy, Bayesian analysis strongly indicates that the data is more likely under H0 (BF = 0.08).

Multiple mediation analysis

Next, we tested whether perceptions of the target persons' social affect and social cognition could underlie the effect of gaze direction and emotional context on trust behavior. Perceptions of empathy and perspective-taking were significantly correlated even after controlling for gaze direction and emotional context ($r = .793$, $p < .001$), suggesting that these two perceptions may affect each other (Hayes, 2017). Hence, we conducted a multiple mediation analyses with xxx bootstrapping samples with gaze direction (coded with averted gaze = 1 (baseline), mixed gaze = 2) and emotional context (coded with neutral = 0 (baseline, negative = 1) as independent variables, investment in the TG as the dependent variable, empathy as the first mediator and perspective-taking as the second mediator. Importantly, despite the significant correlation between empathy and perspective-taking perceptions, there was no indication of multicollinearity issues. All variance inflation factors were low (below 3) and tolerance levels were high (above 0.3). Given that we only found significant differences in trust behavior between averted and mixed gaze, we only considered these two gaze conditions in this analysis.

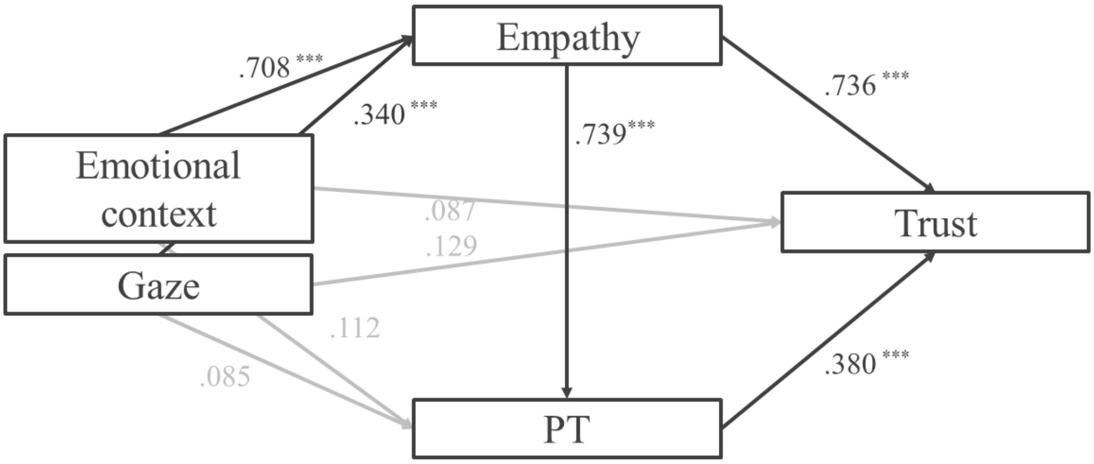
Results are summarized in Figure 3. We found that the indirect effect of gaze direction and emotional context on trust behavior through empathy (b = .77, SE = .16, $p < .001$) as well as through perspective-taking was significant (b = .36, SE = .10, $p < .001$).

In addition, the direct effect of gaze direction and emotional context on trust was not significant (b = .22, SE = .26, $p = .40$; total effect: b = 1.34, SE = .28, $p < .001$). Our results suggest that the effects of gaze and emotional context on trust are mediated by perceptions of the target person's ability for social understanding. Specifically, target persons who displayed mixed gaze or who listened to negative stories were considered more empathic and more capable of perspective-taking and such perceptions led to higher trust in the TG.

We note that we considered empathy as a first mediator. However, results are the same if the order of the mediators is inverted. Finally, excluding participants (N = 16) who did not believe the instructions of the experiment (i.e., that they were playing the TG with another participant) led to the same results, except that the interaction between gaze direction and emotional context on perspective taking perceptions was no longer significant ($p = .18$).

Figure 3

Multiple mediation model displaying the direct and indirect effects of gaze and emotional context on trust behavior



Note. Mixed gaze as well as negative emotional context significantly increased participants' perceptions of the target person capacity for empathy, which in turn also led to higher ratings of perspective taking and higher investments in the TG.

Altogether, results of experiment 1 are partially in line with a previous study on the effects of gaze behavior and emotional context on perceptions of the target person (Breil & Böckler, 2021). Differently than previously reported, we did not find significant differences between direct and averted gaze for perceptions of empathy and perspective-taking, while we did find differences between direct and mixed gaze. This is supported by Bayesian analyses that indicate that the results are much more likely under the model that includes only the two main effects of gaze and emotional context but not their interaction. Regarding trust behavior, results are partially in line with trust perceptions reported in (Breil & Böckler, 2021). With a behavioral measure of trust, we replicated and extended the differences previously observed on trust perceptions between emotional contexts and between mixed and averted gaze. However, we did not find differences in trust behavior between direct and averted gaze, nor a significant interaction between gaze direction and emotional context (which is further supported by Bayesian statistics). Our results extend previous findings by showing that emotional context and gaze direction may affect not only trust perceptions of the target person, but also participants' willingness to place their trust in this person in an incentivized interaction. Moreover, our results show that gaze direction and emotional context ultimately influence prosocial behavior of observers by affecting their perceptions of the target person's social affect and social cognition. Those who are deemed more skilled in social understanding are more trusted in a subsequent social interaction.

One possible interpretation of this result is that individuals observing a social interaction might consider social affect and cognition skills as a signal of one's reputation for attending to the needs of others. As a result, those who are considered more socially competent are generally met with greater generosity. An alternative explanation, however, is that results of experiment 1 are due to participants' strategic behavior. Participants might have expected that individuals with greater social understanding skills would be more prosocial and thus more likely to reciprocate their trust in the TG. This assumption is not unreasonable considering that previous studies have shown a positive correlation between social understanding and prosocial behavior (FeldmanHall et al., 2015; Leiberg et al., 2011). As a consequence, participants might have invested more in individuals who they deemed more socially competent.

Experiment 2 aims to disentangle these two competing explanations and gain better understanding of the psychological motivations underlying the observed result of higher prosocial behavior towards individuals perceived as socially competent. To do so, we investigate whether perceptions of others' social affect and social cognition also affect prosocial behavior in a non-strategic setting. We preregistered these two competing hypotheses as well as an analysis plan at OSF (https://osf.io/8ey4f).

Experiment 2

Methods

Participants

We recruited 162 participants (M_{age} = 28.65 years, *SD* = 9.71, 46% female, 51.5% male, 2.5% diverse) from the same online panel as in experiment 1. We used the same inclusion criteria as in experiment 1 and only participants who (i) were fluent in German, (ii) did not have hearing issues, (iii) had a functioning audiovisual system in their devices and (iv) were willing to participate in experiments involving deception were invited to take part in the experiment. Individuals who participated in experiment 1 could not participate in experiment 2. The experiment lasted for approximately 15 minutes and participants received a flat fee of 2.40 € for their participation. Sample size was determined a priori using Psychometrica (https://www.psychometrica.de/effect_size.html). We assumed 80% of power and an α of .05 to detect a small effect size as reported in experiment 1 (*f* = .1). As before, we took into account that the number of participants should be a multiple of 54 when determining the sample size (see *Procedure* section).

Procedure

We followed the same study design and procedure of experiment 1 with a few differences. First, instead of playing a TG with the target person, participants were asked to play a Dictator Game (DG) where they would be in the role of the dictator and could freely decide how to allocate chips between themselves and the target person. The use of the DG is crucial to the goals of experiment 2, as in this game the target person is passive. Thus, participants' decisions are devoid of a strategic component (List, 2007), as their payoff does not depend on how the target person reacts to their decisions. Participants were endowed with 10 chips in each trial and asked how they would like to allocate these chips between themselves and the target person. Participants were informed that, at the end of the experiment, one trial would be randomly selected and they would be paid according to how many chips they decided to keep for themselves (1 chip = 2.40€). Like in experiment 1, the target person was not part of the experiment and thus was not informed of the participants' decision in the DG.

Second, when playing the DG, participants were asked to input their responses in a text box (instead of a scale as in the TG in experiment 1). This option allowed more granularity in participants' responses, as they could enter any rational number between 0 and 10 chips. Third, and finally, the experiment was programmed in Inquisit (*Inquisit 5 [Computer software]*, 2016) to enable a better pairing between the audio and visual stimuli.

Data Analysis

As in experiment 1, we employed a 2 (emotional context: negative or neutral) × 3 (gaze: direct, averted or mixed) within-subjects design. We conducted the same analyses described in experiment 1, following our preregistered analyses plan.

Results and Discussion

Empathy ratings

As in experiment 1, ANOVA results revealed a main effect of gaze direction on empathy perceptions ($F(2,322) = 20.62$, $p < .001$, $\eta^2_{gen} = .027$) and a main effect of emotional context ($F(1,161) = 99.29$, $p < .001$, $\eta^2_{gen} = .063$). Differently than experiment 1, the interaction between gaze and emotional context was significant ($F(2,322) = 4.54$, $p = .011$, $\eta^2_{gen} = .006$). Post-hoc *t*-tests showed that mixed gaze led to higher perceptions of empathy than averted gaze in both emotional contexts (neutral: $t(161) = 4.21$, $p < .001$, $d = .351$; $M_{mixed} = 2.72$, SD = 1.31, $M_{averted} = 2.23$, SD = 1.34; negative: $t(161) = 4.92$, $p < .001$, $d = .424$; $M_{mixed} = 3.70$, SD = 1.49, $M_{averted} = 3.03$, SD = 1.55), whereas perceptions of mixed gaze only led to higher empathy ratings than direct gaze in the negative emotional context (neutral: $t(161) = -.098$, $p > .99$, $d = .005$; $M_{direct} = 2.73$, SD = 1.35; negative: $t(161) = 4.07$, $p < .001$, $d = .\ 351$; $M_{direct} = 3.17$, SD = 1.49). In contrast, target persons displaying direct gaze were perceived to be more empathic than those displaying averted gaze only in the neutral condition (neutral: $t(161) = 3.99$, $p < .001$, $d = .354$; Negative: $t(161) = 1.06$, $p > .99$, $d = .079$). Results can be seen in Table 1 and Figure 4 (left panel).

Table 1

Means and standard deviations of empathy, perspective-taking ratings and giving behavior in the Dictator game.

Gaze	Emotional context	Empathy	Perspective-taking	Giving
Mixed	Neutral	2.72 (1.31)	2.67 (1.32)	2.07 (1.91)
Averted	Neutral	2.23 (1.34)	2.25 (1.39)	1.78 (1.99)
Direct	Neutral	2.73 (1.35)	2.83 (1.43)	2.22 (2.18)
Mixed	Negative	3.7 (1.49)	3.72 (1.45)	2.97 (2.48)
Averted	Negative	3.03 (1.55)	2.96 (1.51)	2.47 (2.38)
Direct	Negative	3.17 (1.49)	3.02 (1.49)	2.51 (1.24)

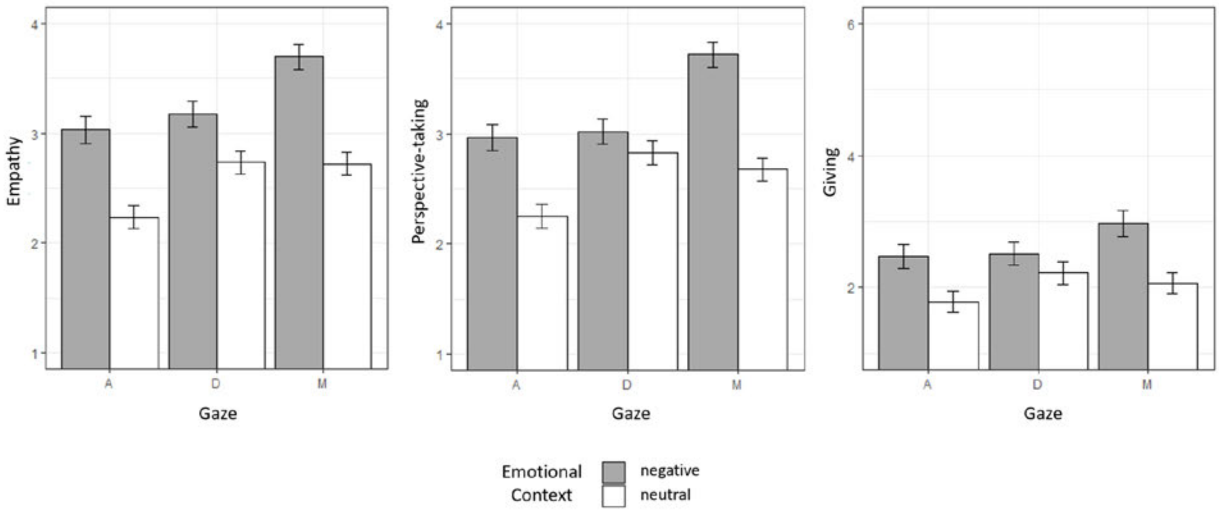
Note. Standard deviations are reported between parenthesis.

Perspective-taking ratings

Similarly to the empathy ratings, ANOVA results showed a main effect of gaze direction ($F(2,322) = 19.15$, $p < .001$, $\eta^2_{gen} = .028$) and a main effect of emotional context on perspective-taking ratings of the target person ($F(1,161) = 71.52$, $p < .001$, $\eta^2_{gen} = .049$). The interaction between gaze direction and emotional context was also significant ($F(2,322) = 12.16$, $p < .001$, $\eta^2_{gen} = .015$). Post-hoc *t*-tests showed the same pattern of empathy ratings, such that mixed gaze led to higher perspective-taking ratings than averted gaze in both emotional context conditions (neutral: $t(161) = 3.73$, $p = .001$, $d = .289$; $M_{mixed} = 2.67$, SD = 1.32, $M_{averted} = 2.25$, SD = 1.39; Negative: $t(161) = 5.55$, $p < .001$, $d = .485$; $M_{mixed} = 3.72$, SD = 1.45, $M_{averted} = 2.96$, SD = 1.51), whereas the comparison between mixed and direct gaze was only significantly different in the negative emotional context (neutral: $t(161) = 1.29$, $p > .99$, $d = .116$; $M_{direct} = 2.83$, SD = 1.43; negative: $t(161) = 5.12$, $p < .001$, $d = .472$; $M_{direct} = 3.02$, SD = 1.49) and the comparison between direct and averted gaze only differed in the neutral emotional context (Neutral: $t(161) = 4.38$, $p < .001$, $d = .392$; negative: $t(161) = .41$, $p > .99$, $d = .017$).

Figure 4

Results per gaze condition and emotional context (Experiment 2)



Note. Participants' ratings of the target person regarding their capacity to empathize with the speaker (left panel) and take their perspective (middle panel) are shown on a 6-point scale, where higher ratings indicate higher capacity for empathy and perspective taking. The average number of chips (from 0 to 10) transferred to the target person in the Dictator Game are shown in the right panel. The x-axis shows the gaze conditions (A: averted gaze; D: direct gaze; M: mixed gaze).

In addition, we also found that emotional context had no effect when the target person displayed direct gaze ($t(161) = 1.49$, $p = .41$, $d = .135$), whereas negative emotional context led to higher perspective-taking ratings in the mixed gaze ($t(161) = 8.6$, $p < .001$, $d = .756$)

and averted condition ($t(161) = 5.5$, $p < .001$, $d = .495$) compared to the neutral emotional context.

Giving in the DG

Results followed a similar pattern of empathy and perspective-taking ratings. On average, participants transferred 2.3 (SD = 2.21) chips to the target person. ANOVA results showed a main effect of gaze direction ($F(2,322) = 7.5$, $p = .001$, $\eta^2_{gen} = .005$) and emotional context on the amount transferred to the target person in the DG ($F(1,161) = 35.16$, $p < .001$, $\eta^2_{gen} = .02$), as well as a significant interaction between gaze direction and emotional context ($F(2,322) = 4.13$, $p = .017$, $\eta^2_{gen} = .003$). Post-hoc *t*-tests showed the same pattern of perspective-taking ratings. In particular, giving was higher for mixed compared to averted gaze for both neutral and negative context (neutral: $t(161) = 2.54$, $p = .072$, $d = .135$; $M_{mixed} = 2.07$, SD = 1.91, $M_{averted} = 1.78$, SD = 1.99; negative: $t(161) = 2.80$, $p = .036$, $d = .195$; $M_{mixed} = 2.97$, SD = 2.48, $M_{averted} = 2.47$, SD = 2.38). In contrast, the difference between mixed and direct gaze was statistically different in the negative emotional context, but not significantly different for neutral stories (neutral: $t(161) = 1.25$, $p > .99$, $d = .075$; $M_{direct} = 2.22$, SD = 2.18; negative: $t(161) = 2.92$, $p = .024$, $d = .186$; $M_{direct} = 2.51$, SD = 1.24). Furthermore, direct and averted gaze were significantly different for neutral but not for negative emotional context (neutral: $t(161) = 2.93$, $p = .024$, $d = .199$; negative: $t(161) = .25$, $p > .99$, $d = .015$). Moreover, emotional context did not influence giving behavior when the target person displayed direct gaze ($t(161) = 1.67$, $p = .29$, $d = .148$). In contrast, a higher amount was transferred to target persons in the DG in the negative emotional context when they displayed mixed gaze ($t(161) = 5.65$, $p < .001$, $d = .409$) and averted gaze ($t(161) = 4.4$, $p < .001$, $d = .315$).

Multiple mediation analysis

As in experiment 1, perceptions of empathy and perspective-taking were significantly correlated even after controlling for gaze direction and emotional context ($r = .827$, $p < .001$).

Nevertheless, there was no indication of multicollinearity issues, as variance inflation factors were low (below 3.23) and tolerance levels were high (above 0.31). Thus, as preregistered, we conducted a multiple mediation analysis with gaze direction and emotional context as independent variables, giving in the DG as the dependent variable, empathy as the first mediator and perspective-taking as the second mediator. Because the significant difference in gaze behavior was between averted and mixed gaze, we only considered these two conditions in this analysis.

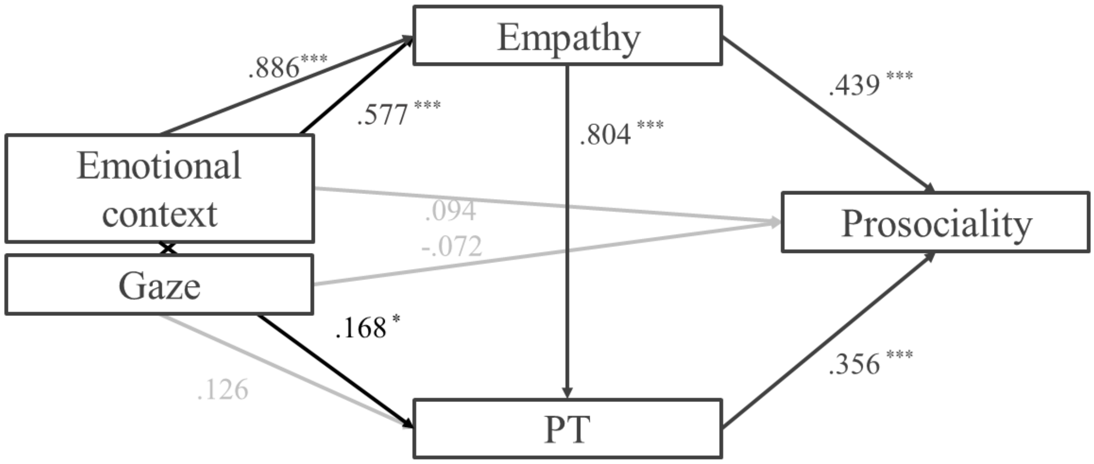
Results can be seen in Figure 4. Similar to experiment 1, we found that the indirect effect of gaze direction and emotional context on giving through empathy was significant (b = .64, SE = .15, $p < .001$), as well as the indirect effect of gaze direction and emotional context on giving through perspective-taking (b = .39, SE = .1, $p < .001$). The direct effect of gaze direction and emotional context was not significant (b = .021, SE = .23, $p = .93$; total effect: b = 1.05, SE = .23, $p < .001$).

These results suggest that the effects of gaze direction and emotional context on prosocial behavior are mediated by participants' perceptions of the target person's capacity for empathy and perspective-taking. Results are similar if the order of mediators is inverted. Excluding participants who did not believe they were playing with other real participants (N = 4) did not change results.

Importantly, based on the results of experiment 1, we did not anticipate the significant interaction between gaze direction and emotional context on giving behavior. Because ANOVA results showed a significant interaction between gaze and emotional context on giving behavior, we conducted an exploratory (not preregistered) multiple mediation analysis including the interaction term between these two independent variables. For this analysis, all gaze conditions were included. Results are similar, as we found a significant indirect effect of gaze and emotional context on giving through empathy (b = .49, SE = .23, $p = .035$) and through perspective-taking (b = .26, SE = .12, $p = .03$), while the direct effect of gaze and emotional context on giving behavior was not significant (b = -.02, SE = .68, p = .28; total effect: b = .73, SE = 1.07, p = .28).

Figure 4

Multiple mediation model displaying the direct and indirect effects of gaze and emotional context on giving behavior.



Note. Mixed gaze as well as negative emotional context significantly increased participants' perceptions of the target person capacity for empathy, leading to higher ratings of perspective taking and higher giving the in the Dictator Game.

General Discussion

Humans rely on a range of verbal and non-verbal cues to decide whether or not being generous towards others is worth the costs associated with prosociality. Here we investigated whether non-verbal cues in social interactions influence the behavior of observers. Although the effect of gaze on the prosocial behavior of interacting dyads has already been established (Behrens et al., 2020; Kret et al., 2015), less is known about how these effects are shaped by the emotional context of the encounter and the perception of the other person's social cognition skills. Across two experiments, we replicate previous findings showing that gaze behavior as well as the emotional context of the conversation have important implications to how individuals are evaluated by observers (Breil & Böckler, 2021). In addition, we expanded previous findings by demonstrating that these non-verbal cues affect not only observers' perceptions of the listener in a conversation, but their prosocial behavior in a subsequent interaction with the listener. Specifically, listeners who displayed mixed gaze (as compared to averted gaze) and who heard a negative story were rated higher in terms of empathy and perspective-taking and were met with more generous behavior. Interestingly, our findings shed light into the process by which non-verbal cues might affect observers' prosocial behavior. In both studies we find that observers' perceptions of the listeners' social cognition mediate the relationship between gaze direction, emotional context and prosocial behavior.

Because we find a significant influence of gaze direction and emotional context on the prosocial behavior of observers in both strategic (i.e., the TG) and non-strategic settings (i.e., the DG), we rule out that participants behaved nicely towards the listeners purely due to strategic concerns to maximize their payoffs. Instead, we believe our results are in line with indirect reciprocity. That is, participants might have been more generous to listeners who they perceived to have acted nicely towards the speaker.

Indirect reciprocity is essential for the sustenance of cooperation in large societies (Rand & Nowak, 2013). By observing others interact, individuals can learn important information about other's reputations, which is then selectively used in subsequent interactions. While previous studies have focused on prosocial behavior towards another as a signal of reputation (Almenberg et al., 2011; Watanabe et al., 2014; Wedekind & Milinski, 2000), our results suggest that non-verbal behaviors in social interactions can also serve as an indication of one's reputation for tending to others' needs.

Interestingly, previous studies show that being seen acting nicely towards others not necessarily translates in a good reputation. The motives underlying one's behavior are crucial for reputation building (Berman & Silver, 2022). While helping others due to self-interest is negatively perceived, individuals who show signs of emotions or empathy while doing good deeds are usually seen as more altruistic (Erlandsson et al., 2020). This is because emotions are perceived as a genuine signal of one's concern for others (Barasch et al., 2014). Similarly, we argue that non-verbal cues such as where individuals look during conversations and whether they are confided a negative story might provide insights into one's social affect and social cognition, and as consequence, be perceived as genuine and reliable signs of one's prosocial inclinations. This in turn leads to reciprocity.

The influence of non-verbal cues on prosocial behavior has important implications regarding the perception and treatment of non-neurotypical individuals. Research has shown that gaze behavior of individuals in the autism spectrum disorder differs from neurotypical individuals (Senju & Johnson, 2009a). As a consequence, they might be misperceived as being less empathic (Johnson et al., 2009), which might have important consequences for how they are treated, ultimately impacting their well-being (Mitchell et al., 2021). Future studies could examine whether social cognition training as an intervention for non-neurotypical individuals as well as inclusive education for neurotypical individuals could minimize this misperception as well as its consequences. Furthermore, health workers, such as psychotherapists or clinicians, could receive tailored training to help them adapt their gaze behavior to the situational context in order to create a safe atmosphere and promote trust in their patients. Similarly, both physical consultants and artificial agents in the economic sector could be trained to pay attention to the gaze behavior of their customers as well as to avoid or establish eye contact, depending on the emotional shading of the situation. In addition, it might be interesting to investigate how the perception of non-verbal cues in social interaction as well as its downstream consequences in prosocial behavior differ in non-neurotypical populations or different cultures. Finally, given the richness of social interactions, future studies could investigate whether our results expand beyond the non-verbal cues examined here. As other non-verbal cues such as accompanying gestures have been associated with increased empathy (Chu et al., 2014), it could be that their observation by third-parties also contribute to social perception and prosocial behavior.

We note that despite our efforts to keep the experiments as similar as possible to each other, we observed one difference in the results: while emotional context modulated the effects of gaze in prosocial behavior in experiment 2, it did not in experiment 1. In line with previous research (Breil & Böckler, 2021), we found in experiment 2 that averted gaze was more accepted and less penalized while the listener allegedly listened to negative stories rather than neutral ones. Because the modulation of gaze behavior by emotional context affected both ratings and prosocial behavior in experiment 2, we believe the differences observed between studies cannot be attributed to the different measure of prosocial behavior. Differences are more likely due to measurement error or to the fact that the software used in experiment 1 rendered the encounter between listeners and speakers slightly less believable by participants. Future studies should further investigate the interaction between gaze behavior and context as well as possible boundary conditions to this relationship.

MANUSCRIPT 6

THE LENS SHAPES THE VIEW: ON TASK DEPENDENCY IN TOM RESEARCH

Christina Breil[1], Anne Böckler[1,2]

[1]Julius-Maximilians-University of Wuerzburg, Germany

[2]Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

This article has been published in *Current Behavioral Neuroscience Reports* (2020) with DOI 10.1007/s40473-020-00205-6

Abstract

Purpose of review*:* This article provides an overview of current findings on Theory of Mind (ToM) in human children and adults and highlights the relationship between task specifications and their outcome in socio-cognitive research.

Recent findings: ToM, the capacity to reason about and infer others' mental states, develops progressively throughout childhood - the exact time course is still a matter of debate. Neuroimaging studies indicate the involvement of a widespread neuronal network during mentalizing, suggesting that ToM is a multifaceted process. Accordingly, the tasks and trainings that currently exist to investigate and enhance ToM are heterogeneous, and the outcomes largely depend on the paradigm that was used.

Summary: We argue for the implementation of multiple-task batteries in the assessment of socio-cognitive abilities. Decisions for a particular paradigm need to be carefully considered and justified. We want to emphasize the importance of targeted research on the relationship between task specifications and outcomes.

Key words: Theory of Mind; mentalizing; perspective-taking; social cognition; social interaction; task dependency

Introduction

Humans are extraordinarily social beings: we interact with various and varying groups, we bicker and play with each other, and we cooperate, trade and deceive. Understanding and predicting the behavior of others is of crucial importance in our everyday lives and the ability to do so is based on Theory of Mind (ToM; also termed mentalizing, cognitive perspective-taking). ToM, the ability to reason about or infer others' mental states, has been a core topic of social sciences for more than 40 years (Premack & Woodruff, 1978) and ever since has been investigated with a broad range of paradigms that make use of diverse materials. Participants in mentalizing research have read or memorized stories (Happé, 1994; Kindermann et al., 1998), played games (Gallagher et al., 2002), and watched comic strips or film sequences (Brunet et al., 2000; Kanske et al., 2015b), all aiming to elicit thoughts about other people's minds. Table1 and Figure1 present examples for different ToM tasks. ToM has been investigated across various age groups (Johansson Nolaker et al., 2018; Wellman et al., 2001), in humans and in animals (C. Heyes, 2014; C. M. Heyes, 1998), in typically developing individuals (Johansson Nolaker et al., 2018), and in psychopathologies (Langdon et al., 2017). As data from different paradigms accumulated, it became clear that ToM is not a monolithic ability, but rather a multifaceted construct with distinct interrelated sub-processes. As a result, the existing paradigms for ToM assessment are heterogeneous, focusing on different aspects of mentalizing, and none of them can capture the concept in its entirety (Schurz et al., 2014). In this review, we aim to provide a brief overview of the current state and recent trends in human ToM research. Most importantly, we want to illustrate the impact a specific paradigm can have on the experimental outcome in this framework.

Table 1.

Examples for ToM tasks. This table provides a, by no means exhaustive, overview of classically employed ToM paradigms including short descriptions of the relevant ToM aspect, task specifications and main fields of application.

Paradigm	**Authors (selection)**	**ToM aspect**	**Task description**	**Dependent variable**	**Main application field**
Classic change of location false belief (FB) task	Wimmer & Perner (1983)(Wimmer & Perner, 1983); Baron-Cohen, Leslie & Frith (1985)(Baron-Cohen et al., 1985)	First-order false belief (FB) attribution	Watch agent A place an object in location 1. Watch agent B place the same object in location 2 while agent A is absent (FB) or present (TB)	Prediction of the searching behavior of agent A (correct or incorrect location)	Developmental research; neuroscientific research; clinical research
Violation Of Expectation	Onishi & Baillargeon (2005)(Onishi & Baillargeon, 2005); Kovács, Téglás & Endress (2010)(Kovacs et al., 2010)	Implicit first-order FB attribution	Habituation phase: Watch an object being retrieved repeatedly from the same box; Test phase: Watch a typical FB task scenario	Looking times in response to expected versus unexpected events	Developmental research; animal research
Anticipatory Looking	Clements & Perner (1994)(Clements & Perner, 1994);	Implicit first-order FB attribution	Watch a typical FB task scenario	Location of anticipatory looking response	Developmental research; animal research
Interactive paradigms	Buttelmann, Carpenter & Tomasello (2009)(Buttelmann et al., 2009)	Implicit first-order FB attribution	Watch a typical FB task scenario, then watch the agent trying to open one of the boxes	Helping behavior for opening the boxes	Developmental research; animal research
Diverse	Wellman & Liu	First-order desire	State own preference about two	Prediction of the choice	Developmental

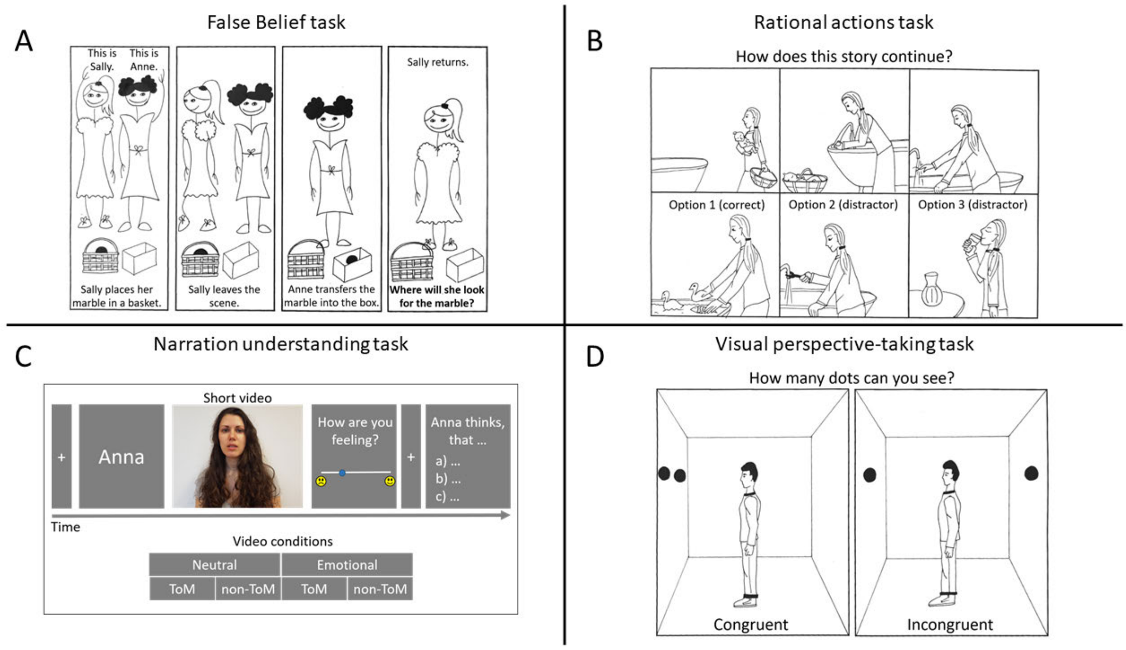
Desires	(2004)(Wellman & Liu, 2004)	attribution	objects, then learn that another agent has the opposite preference	of the agent	research
Diverse Beliefs	Wellman & Liu (2004)(Wellman & Liu, 2004)	First-order belief attribution (not knowing the truth)	State own belief about the location of an object while not knowing the true location. Then learn that another agent has a different belief	Prediction of the searching behavior of the agent	Developmental research
Knowledge Access	Wellman & Liu (2004)(Wellman & Liu, 2004)	First-order knowledge attribution	Learn about the content of an unlabeled box. Then learn about the ignorance of the agent about the contents of the box	Judgement of the knowledge of the agent	Developmental research
Contents FB	Wellman & Liu (2004)(Wellman & Liu, 2004)	First-order belief attribution (knowing the truth)	Learn about the content of an incorrectly labeled box. Then learn about the ignorance of the agent about the true contents of the box	Judgement of the knowledge of the agent	Developmental research
Real-Apparent Emotion	Wellman & Liu (2004)(Wellman & Liu, 2004)	Felt vs. displayed emotion	Listen to the story of an agent hiding his real emotions	Recognition of the real versus the displayed emotion on pictures	Developmental research
Second-order FB task	Perner & Wimmer (1985)(Perner & Wimmer, 1985)	Second-order belief attribution	Listen to the story of two agents, both forming a TB while thinking that the other agent holds a FB	Prediction of the behavior of the agents	Developmental research; inter-individual differences research

Second-order FB task with deception	Sullivan, Zaitchik & Tager-Flusberg (1994)(Sullivan et al., 1994)	Second-order belief attribution	Listen to the story of agent A deliberately misinforming agent B. Then learn about agent B finding out the truth without agent A knowing it	Judgement of the belief of agent A regarding the mental state of agent B	Developmental research; inter-individual differences research
Rational actions	Brunet, Sarfati, Hardy-Baylé & Decety (2000)(Brunet et al., 2000)	Attribution of intentions	View comic stories with open ending	Choice of a logical ending of the story among several options	Developmental research; neuroscientific research; clinical research
„Yoni“ task	Shamay-Tsoory & Aharon-Peretz (2007)(Shamay-Tsoory & Aharon-Peretz, 2007)	Mental state judgement	Watch a comic face surrounded by four objects of the same category, e.g. fruits	Choice of one of the objects based on the expression or gaze direction of the face	Developmental research; neuroscientific research; clinical research
Reading-The-Mind-In-The-Eyes test	Baron-Cohen et al. (2001)(Baron-Cohen et al., 2001)	Emotion/mental state recognition	Look at pictures of human eye regions	Choice of one out of four words to describe best what the person is feeling	Neuroscientific research; clinical research
Strategic games	Kircher et al. (2009)(Kircher et al., 2009); Sripada et al. (2009)(Sripada et al., 2009)	Advanced ToM	Play the ultimatum game or the prisoners dilemma game with a human counterpart	Choices during the game	Neuroscientific research
Social animations	Castelli, Happé, Frith & Frith (2000)(Castelli et al., 2000); Blakemore et	Attribution of intentions	Watch video animations of interacting geometrical shapes	Explanations of behavior or answer to questions (open format)	Neuroscientific research; clinical research

	al. (2001)(Blakemore et al., 2001)				
Interaction observation	Baksh, Abrahams, Auyeung & McPherson (2018; "Edinburgh Social Cognition Test (ESCoT) ")(Baksh et al., 2018); Dziobek et al. (2006; "Movie for the Assessment of Social Cognition (MASC)")(Dziobek et al., 2006)	Higher-order ToM, social norm understandin g	Watch dynamic interactions with or without social norm violation	Explanation and interpretation of the interactions (open format)	Inter-individual difference research; neuroscientific research
Narration understanding	Kanske, Böckler, Trautwein & Singer (2015; „EmpaToM")(Kanske et al., 2015b)	Higher-order ToM, empathy	Watch short videos of autobiographical narratives	Answers to multiple-choice questions requiring interpretation of the stories	Inter-individual differences research; neuroscientific research; clinical research
Visual perspective-taking	Keysar, Dale, Barr, Balin & Brauner (2000)(Keysar et al., 2000); Samson, Apperly, Braithwaite, Andrews & Bodley Scott (2010)(Samson et al., 2010)	Visual perspective-taking	Take the point of view of another person or avatar in order to follow their instructions or to judge their field of view	Number of gaze fixations or RT of judgement of objects visible to oneself versus the other person or avatar	Inter-individual differences research; neuroscientific research

Figure 1

Overview of ToM task categories and typical examples



Note. Panel A: Depiction of the Sally-Anne task as an example for a False Belief task (based on Baron-Cohen et al. (1985)*(Baron-Cohen et al., 1985)*). Participants need to understand that Sally holds a false belief (that differs from their own) in order to solve the task. *Panel B*: Example for a rational actions task (based on Brunet et al. (2000)*(Brunet et al., 2000)*). Selection of the correct picture requires an understanding of the depicted agent's goal. *Panel C*: The EmpaToM as an example for a more naturalistic and dynamic narration understanding task. Short video clips depict fictional characters telling short autobiographic stories. The content of these narrations can be neutral or emotional, and the stories can require mentalizing or not. Participants indicate how they feel (as a measure of empathic responding) and answer multiple-choice questions requiring inferences about the mental states of the narrator (ToM condition) or factual reasoning (control condition) (based on Kanske et al. (2015)*(Kanske et al., 2015b)*). *Panel D*: The Samson task as an example for a visual perspective-taking task. Participants judge the number of dots seen by the avatar or from their own perspective. 1)

Congruent condition (avatar and participant see the same number of dots). 2) Incongruent condition (avatar and participant see different numbers of dots) (based on Samson et al., 2010). Slower responses in the incongruent condition are taken as an indication for the tendency to represent not just one's own, but also the avatar's perspective. All parts of this figure are original.

Developing ToM

Understanding other people's mental states is a socio-cognitive competence that develops throughout childhood. Many researchers attribute this process to the sequential emergence of multiple interrelated concepts rather than a single event (Cadinu & Kiesner, 2000; Wellman et al., 2001). Nevertheless, ToM advancements can be roughly divided into three stages: early ToM, which emerges in the first months of life, basic ToM, which is typically developing around the age of four years, and advanced ToM, which does not evolve until six to eight years (Hutchins et al., 2012) and keeps developing throughout adolescence (Osterhaus et al., 2016). Findings from neuroimaging studies suggest a common neuronal basis across the three types of ToM in four-to-eight-year-old children, with particularly strong similarities between basic and advanced ToM (Xiao et al., 2019).

Early ToM

One of the most central debates in current ToM research concerns the mentalizing skills of young infants. The development of new paradigms with more implicit measures, such as spontaneous gaze behavior, paved the way for the investigation of ToM performance in children below the age of two years. Some studies suggested that infants as young as seven to 15 months can master false belief (FB) tasks when implicit paradigms are used (Baillargeon et al., 2010; Kovacs et al., 2010). More recently, however, the generalizability of this notion has been queried. For example, a meta-analysis revealed that infants' correct performance in implicit FB tasks is highly influenced by the choice of paradigm (Barone et al., 2019). Children were more likely to pass the test when a Violation of Expectation (VOE) paradigm was implemented in the study, compared to anticipatory-looking (AL) or more interactive paradigms. In the VOE paradigm, an expectation, for instance about an agents' behavior, is generated in an initial habituation phase after which the child is presented with either an expected or an unexpected event. The gaze behavior of the infant serves as indication for their inference about the agent's mental state. This is both the benefit and the vulnerability of the paradigm. On the one hand, without any language requirements, even the youngest infants can participate in this task. On the other hand, without explicit responses, longer looking times in the test phase leave much room for interpretation; while they are typically taken as an indication of surprise about an event that is unexpected given the agent's mental state, longer looking times could also reflect a more basic response to a novel stimulus (Barone et al., 2019; C. Heyes, 2014). Thus, deliberate construction of control conditions and habituation phases are necessary to prevent this potential confound – a requirement that many studies fail to satisfy (C. Heyes, 2014; Kulke et al., 2018, 2019). Besides the choice of experimental paradigm, a broad range of task specifics can account for variance in the ToM performance of infants. These include the type of agent and the salience of its mental state as well as the movements of involved objects and whether or not deception was included in the task (Barone et al., 2019).

A recent study revealed the significance of another characteristic of implicit ToM tasks. Fizke et al. (2017) tracked the helping behavior of two-to-three-year old children in two versions of a FB task: one version included aspectuality whereas the other version of the task did not. Aspectuality denotes incompatible beliefs about an object or a person under two different aspects, for example knowing the person Clark Kent as himself versus knowing him as

Superman without being aware of his private identity. Each of the two task versions used by Fizke et al. consisted of a true and a false belief condition. The toddlers reacted differently to the agent's true versus false belief only when aspectuality was not involved in the task. This pattern was taken as an indication of conceptual deficits in infants and is in line with the finding that, below the age of two, they are capable of tracking mental states and can master implicit FB tasks as long as an understanding of aspectuality or of other propositional attitudes is not necessary to pass the test (Fizke et al., 2017; Oktay-Gür et al., 2018; Rakoczy, 2017; Rakoczy et al., 2015).

Taken together, while spontaneous perspective-taking in young infants appears to be a real phenomenon, it is highly dependent on formal and content-related aspects of the paradigm.

Basic ToM

As children grow older, direct questions can be used to examine their ToM skills. Classical investigations employing such elicited-response tasks showed that children from about four years of age are able to attribute mental states to others even when those states differ from their own (Wellman et al., 2001). Around this age, children acquire competence for a large variety of ToM tasks, and the high correlation between performance in these explicit first-order ToM tasks indicates the emergence of a conceptual capacity. Similarly, and in contrast to implicit paradigms, specifics of explicit FB tasks, such as characteristics of the protagonist or the type of question, appear to have no effect on performance (Wellman et al., 2001). This pattern speaks for a more tangible belief conception in children of four years and above, which is largely independent of FB task variations.

Whereas the reported within-task variance appears to be negligible, the content of the other‘s mind has an impact on explicit ToM performance in pre-school children. Wellman and Liu

(Wellman & Liu, 2004) developed a scaled set of first-order ToM-tasks and showed that understanding of different mental states in children ages four to six develops in a regular order with progressively broadening comprehension of subjectivity. Specifically, an understanding of desire and intention appears to emerge before an understanding of belief, while an understanding of hidden emotions arises much later. Findings from a recently developed auditory equivalent of the scale showed that children pass the tasks in almost the same order when auditory instead of visual material was presented, which indicates that the assessment of ToM development is modality independent (Hasni et al., 2017). An auditory version of the scale could be especially useful for the assessment of children who show a delay in ToM development and face visual challenges, such as in children with congenital blindness (Peterson et al., 2000).

Burnel et al. (Burnel et al., 2018) continued on this path and designed low verbal versions of Wellman and Liu's tasks with largely similar outcomes. Taken together, these findings exemplify the sequential acquisition of specific ToM skills during childhood and emphasize the importance of a broad assessment of ToM performance during the pre-school years that goes beyond false belief understanding and includes scaled task batteries.

Besides the progressive understanding of mental states, linguistic abilities have a strong influence on ToM performance. The apparent differences in the age of ToM acquisition between studies can often be explained by differences in linguistic task demands (Burnel et al., 2018; Kamawar & Olson, 2011; Rakoczy et al., 2015). Together with the notion of a close correlation between ToM and language development (Atkinson et al., 2017; Miller, 2009), this finding demonstrates the impact of linguistic requirements in ToM assessment, especially when working with children.

Higher-order ToM and advanced ToM

Along with cognitive development, children acquire the competence to pass more complex mentalizing tasks, so called second-order ToM tasks. While first-order ToM refers to what people think about real events, second-order ToM goes one step further and encompasses what people think about other people's thoughts. As a result, these tasks are inherently more complex and children are generally older when they first accomplish this level of mental state representation. Representations of second-order false beliefs are typically tested with the story vignettes approach by Wimmer and Perner (Perner & Wimmer, 1985). Initial findings suggested that children pass second-order FB tasks under optimal conditions at the age of six or seven years. However, by substantially reducing task complexity and linguistic demands, even five-year-old children showed high success rates. Further facilitative effects have been reported when adding an extra question to prompt the mental state of the agent, such as "Does John know that Mary knows where ice-cream man is now?" (Miller, 2009; Sullivan et al., 1994).

Higher-order ToM includes even more levels than second-order ToM, whereas advanced ToM involves complex understandings of features such as irony, metaphors, or double deceptions. These more complex forms of ToM are acquired later than second-order FB reasoning, between eight and 13 years(Devine & Hughes, 2013) and improve throughout adulthood (Dumontheil et al., 2010). Recently, some of the most widely used paradigms to investigate these forms of social reasoning, in particular the Strange Stories Task (Happé, 1994), have been criticized for low internal consistency (E. O. Hayward & Homer, 2017) and a multifactorial structure of these paradigms has been suggested (Osterhaus et al., 2016). Specifically, (advanced) ToM seems to be an assembly of distinct socio-cognitive competences, including trait judgements, reasoning about rational behavior, and reasoning about ambiguity (Osterhaus et al., 2016; Schurz et al., 2014). Accordingly, capturing the development of advanced ToM throughout adolescence may require a carefully selected battery of tasks that allow targeting the specific underlying socio-cognitive processes.

Mature ToM

Two core questions dominate the investigation of fully developed socio-cognitive capacities. First, fanned by the rapid technical and methodological advances in imaging research, numerous studies addressed the neuronal underpinnings of ToM. Secondly, inter-individual differences in ToM performance and their relation to other constructs, such as executive functions, are informative about the nature of ToM. While paradigms typically used in neuroimaging research are relatively easy and often elicit performance that is at ceiling, research on inter-individual differences requires tasks with a higher level of difficulty.

Neuronal basis

The neuronal activation pattern that accompanies performance of ToM tasks has inspired imaging research for more than two decades. A wide range of experimental paradigms has been deployed and, consequently, findings have been heterogeneous. It is uncontested, however, that a distributed brain network is engaged during mentalizing (Carrington & Bailey, 2009; Frith & Frith, 2006). Two core regions of this network are the temporo-parietal junction bilaterally, which is most specifically engaged in reasoning about other person's mental states (Carrington & Bailey, 2009; Saxe & Kanwisher, 2003; Schurz et al., 2017), and the medial prefrontal cortex (Carrington & Bailey, 2009), which has been suggested to be more generally involved in processing socially and emotionally relevant information (Schurz et al., 2014). Other regions frequently associated with the mentalizing network include the posterior cingulate cortex and parts of the precuneus, the orbitofrontal cortex, the anterior temporal lobes, and the amygdala. Recent endeavors specifically investigated neuronal activation patterns during mentalizing in relation to the task that was employed and found that activation varies with study methodology (Carrington & Bailey, 2009; Frith & Frith, 2006; Schurz et al., 2017). A direct comparison of ToM tasks within one participant sample revealed distinct neuronal activation patterns for different ToM tasks (Spunt & Adolphs, 2014) and specific features of the task, such as the mental state it taps into or whether belief reasoning refers to similar or dissimilar others, differentially engage specific regions of the ToM network (Carrington & Bailey, 2009). As such, neuroimaging research supports the conceptualization of ToM as a multifaceted capacity with varying specifications depending on the context. Accordingly, future research should advance systematic comparisons of neuronal activation and their relation to different paradigms and task aspects (Schaafsma et al., 2015). This endeavor could provide valuable insights about the particular sub-processes that contribute to successful mentalizing.

ToM and Executive Functions

Like with so many other challenges in life, some people are better at ToM than others, and one important role in this context is played by executive functions (EF) (Aboulafia-Brakha et al., 2011; Wade et al., 2018). EF is an umbrella term for cognitive processes that foster goal-directed behavior and problem solving, such as inhibition, updating of working memory, and cognitive flexibility (Miyake et al., 2000). The strong relationship between EF and ToM and the fact that both constructs comprise a large number of processes beg the question whether ToM tasks specifically measure mentalizing or whether – and to what extent – performance in these tasks relies on other, more general, capacities. For instance, the inhibition of prepotent responses, that is critical in EF tasks, and the inhibition of one’s own mental states when inferring others’ mental states in ToM tasks might be very similar inhibition processes. Indeed, neuroscientific evidence suggests that areas associated with EF are involved in mentalizing (Saxe et al., 2006). A strong relationship has been demonstrated in first-order FB tasks, whereas the evidence for effects in second-order FB reasoning is less consistent (Miller, 2009).

Critically, the association of the two constructs can bias findings in ToM research, particularly in groups with limited or impaired EF, for example children, older adults or patients with schizophrenia (Johansson Nolaker et al., 2018; Langdon et al., 2017; Wade et al., 2018). A well-designed task as well as the use of adequate comparison conditions is therefore especially important in these samples. In the case of schizophrenia, a fruitful approach to tap into ToM capacities irrespective of EF is the employment of instructions that only indirectly refer to ToM, for example sorting cartoon pictures (concerning the mental states of the displayed agents) in a logical order or explaining a joke (Langdon et al., 2017). Older adults, on the other hand, could benefit from verbal tasks because vocabulary increases with age (Verhaeghen, 2003). Other important methodological parameters in this context include task complexity and time constraints as well as stimulus material and the modality of presentation (Laillier et al., 2019).

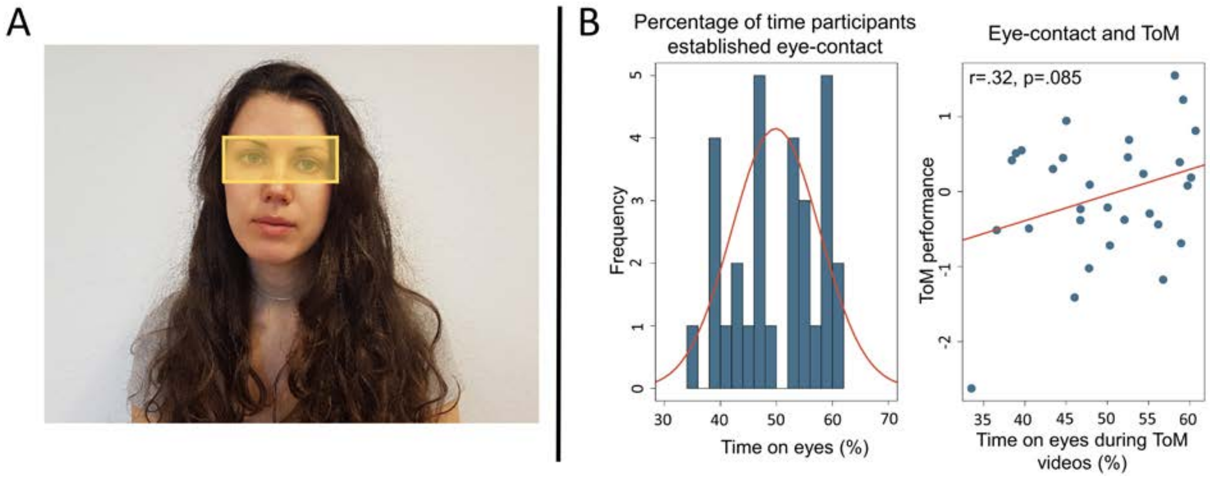
Recent Advances

A central characteristic shared by most FB and other ToM tasks is the binary response format. The resulting pass-or-fail interpretation, together with the fact that performance in those tasks is usually at ceiling in adolescents and adults, makes it difficult to capture variance in mental state representation. Therefore, an important recent trend has been the extension of classical paradigms with continuous measures that allow for the investigation of inter-individual variability. For example, Bradford, Gomez and Jentzsch(Bradford et al., 2019) combined measures of correct performance, reaction time (RT) and electroencephalography (EEG) to investigate the role of perspective-shifting in a ToM task. Other recent RT-based studies demonstrate a connection between visual perspective-taking and cognitive perspective-taking(Bio et al., 2018; Conway & Bird, 2018). Compared to exclusively relying on correct versus incorrect answers, the incorporation of RT measurement better allows for revealing inter-individual variability.

Another promising approach to capturing inter-individual variability in advanced ToM was introduced in the Edinburgh Social Cognition Test (ESCoT) (Baksh et al., 2018). The test employs cartoon-style dynamic interactions together with open questions that are rated based on the quality of the answer. With the dynamic stimulus material, the ESCoT also addresses another obvious yet often overlooked shortcoming of classic social cognition paradigms: their limited ecological validity. Some aspects of ToM are inherently interactive and therefore need to be studied in more complex, dynamic, and naturalistic settings. Other examples of new paradigms that incorporate this idea are the Strange Stories Film Task (Murray et al., 2017), that was based on the original stories from Happé (Happé, 1994), and the EmpaToM (Kanske et al., 2015b), that allows for a simultaneous manipulation and assessment of empathy and ToM with sufficient inter-individual variance in adults. A sample trial sequence of this video-based task is depicted in panel C of Figure1. In a recent pilot experiment, we combined the EmpaToM with eye-tracking to investigate the relationship of basic gaze processes with empathic responding and ToM in a naturalistic social setting. Results are presented in panel B of Figure2. First, we found substantial variance in the individual tendency to establish eye contact with the narrator during the video. Participants spent between 34% and 61% of the time looking at the eye region. In addition, participants who showed a higher empathy-tendency spent less time overall looking at the eyes of the narrator during videos with negative valence ($r=-.44$, $p=.015$). This pattern is in line with the notion of a self-regulative role of gaze behavior in emotionally charged situations (Kendon, 1967b). Hence, empathic participants may have down-regulated their own emotions by looking away from the eye region during emotionally negative videos. Interestingly, the more time participants spent looking at the eyes of the narrator (relative to other areas) during videos with mental state interference was marginally positively related to performance in the subsequent ToM question (r=-.32, p=.085). This finding suggests that eye contact during a conversation might enhance the efficiency of mentalizing processes (Senju & Johnson, 2009b). Given that present results are based on only 30 participants and that effects are relatively small, further studies are certainly necessary before strong conclusions can be drawn. However, we think our pilot study suggests that probing the relation between basic perceptual and behavioral processes on the one hand and performance in ToM tasks on the other hand can be promising.

Figure 2

The EmpaToM



Note. Panel A: Example for the region of interest (eye region) for one of the narrators in the EmpaToM. *Panel B:* Pilot findings of gaze behavior in relation to ToM. The histogram displays how much time participants spent looking at the eye region during the videos (in percent). The scatter plot shows the correlation between the relative duration participants looked at the eye region during ToM videos and performance in ToM questions (composite score integrating speed and accuracy). All parts of this figure are original.

Rapid technical advances pave the way for even more naturalistic paradigms in adapting a second-person account. Live video-feed, mobile eye-tracking or motion capture are promising ways to study social cognition in a more interactive and ecologically valid fashion (see Lehmann, Maliske, Böckler & Kanske (Lehmann et al., 2019b) for a review). As virtual reality (VR) technology becomes more available, it is increasingly integrated in social cognition paradigms as well (Pan & Hamilton, 2018). For example, in a recently developed VR task for the investigation of ToM in schizophrenia, participants run errands in a virtual shopping center (Canty et al., 2017). The scenario involves social interactions which are complemented with multiple-choice questions requiring an interpretation of the encounter. The great opportunity of VR is the potential to bridge the gap between ecological validity and experimental control. Changes of specific variables, for example the gender of the interaction partner, can be easily implemented while keeping all other parameters constant. Moreover, VR facilitates reproducibility because, once created, scenarios can be shared across laboratories. In view of the replicability crisis, this is an opportunity of special importance.

Enhancement of developing and mature ToM

Even though ToM development follows a relatively consistent pattern across children, it can be promoted during childhood. In the first years of life, mental-state talk of the caregiver is related to children's later understanding of the mind (Meins, 2012; Ruffman et al., 2018; Taumoepeau & Ruffman, 2008). Storybook interactions with a special focus on the mental states of the character are an easy way for parents to support false belief understanding in this age group (Tompkins, 2015). Later, during the first years of school, conversations about the mind and group discussions about mental states, which can be delivered by the teacher (Bianco & Lecce, 2016), can successfully enhance ToM skills (Bianco et al., 2016; Lecce, Bianco, Devine, et al., 2014; Ornaghi et al., 2014). While meta-analyses show that shorter periods of training with longer session durations seem to be more efficient, the discovery of the most effective training practices requires further research (Hofmann et al., 2016).

Interestingly, some studies incorporated additional outcome measures – with mixed results. For instance, training of first-order ToM can transfer onto more advanced forms of ToM (Lecce, Bianco, Demicheli, et al., 2014) and a training that was mainly constructed to enhance children’s emotion understanding through conversational interventions on emotions also showed a positive effect on other social cognition aspects, such as ToM (Ornaghi et al., 2014). On the other hand, a storybook interaction approach intended to promote emotion understanding, social competence, and false belief understanding in pre-school children, only had an effect on the latter (Tompkins, 2015). Training of an isolated feature, for example false belief understanding, cannot do justice to a multifaceted construct such as ToM. It is therefore not surprising that the increase in specific ToM skills in autistic children and adults after trainings with standardized tests often fail to transfer onto more generalized ToM measures or social competence in real life (Begeer et al., 2011; Golan & Baron-Cohen, 2006; Ozonoff & Miller, 1995).

Recent research suggests that ToM performance can also be enhanced in healthy adults. A mental training protocol that targeted a rather wide range of socio-cognitive skills, such as flexible perspective-taking on self and others and observing one’s own thoughts, led to increased performance in an advanced and high-level ToM task (EmpaToM, Böckler et al., 2017; Trautwein et al., 2020). The observed behavioral improvement was accompanied by changes in grey-matter volume in neuronal regions that are consistently associated with ToM (Valk et al., 2017).

The promotion of socio-cognitive capacities is of special interest in ageing populations, as ToM has been found to decrease with age (Henry et al., 2013). Fortunately, older adults benefit no less from ToM training than younger adults when a conversational approach is used (Rosi et al., 2016). Diversified ToM trainings that include practicing visual perspective-taking, first- and higher-order ToM, and mentalizing in various real-life contexts seem suitable to enhance performance in different ToM measures in older adults (Cavallini et al., 2015; Lecce et al., 2015).

Taken together, ToM performance can be promoted throughout life, but the effects of social cognition trainings seem to critically depend on their content (Begeer et al., 2011; Cavallini et al., 2015; Golan & Baron-Cohen, 2006; Ozonoff & Miller, 1995). An improvement of ToM in its entirety requires training of the whole spectrum of the concept. In this context, more true-to-life procedures are a promising avenue; six months after a five-week VR-based social cognition training, autistic individuals reported increased social skills, such as maintaining a conversation and establishing relationships, in their everyday life (Kandalaft et al., 2013).

Conclusions

In this article, we illustrate how the choice of paradigm and its characteristics shape the outcome of ToM assessment throughout all age groups. In young infants, spontaneous mentalizing skills as investigated with implicit designs largely depend on formal and content-related aspects of the task. In addition, linguistic requirements and the strong relationship between ToM and EF are critical when assessing ToM in childhood. A multiple-task battery allows a broad investigation, which enables a more comprehensive assessment of ToM capacities and helps to determine the current stage of ToM development in children (Wellman & Liu, 2004). In adults, behavioral observations and neuronal activation patterns exemplify the task-dependent and multifaceted nature of ToM. Similarly, while ToM performance can be promoted by training programs in both children and adults, the generalizability of training effects depends on the scope of the training, supporting the view that “you get what you give”.

Based on the findings reviewed in this article, we want to promote a multifaceted approach in the assessment of socio-cognitive competences. The application of multiple-task batteries instead of a monolithic treatment of ToM is of central importance in this context. In line with this point, we want to emphasize the significance of making deliberate and well-informed decisions about the paradigms, specific variations, and control conditions that are incorporated in research.

To achieve these objectives, further research needs to probe the precise relationship between task settings and their behavioral and neuronal outcomes in more detail. Existing meta-analyses on this issue provide a good basis (Barone et al., 2019; Henry et al., 2013; Schurz et al., 2014, 2017; Wellman et al., 2001). Systematic comparisons of different paradigms and their variations within the same population are vital for future research. Based on the notion that cultural variations exist in mentalizing (C. Heyes, 2003; C. M. Heyes & Frith, 2014), we believe that cross-cultural comparisons could be a fruitful addendum to this new line of research. A better understanding of the nature and the evolution of ToM could contribute to a well-grounded approach of future mentalizing assessment.

The incorporation of continuous measures and naturalistic stimuli are promising ways towards a more profound and comprehensive assessment of socio-cognitive capacities. This approach could be extended with a combination of diverse behavioral and physiological measures to capture the vast range of processes that contribute to and are involved during mentalizing. As an example, our above mentioned pilot findings suggest a relationship between basic attentional processes and advanced ToM capacity in adults: participants who spent more time looking at the eyes of narrators were somewhat better in understanding their mental states. Investigating the relationship between basic processes and ToM can pave the way for new approaches to promote mentalizing skills. Research revealed that both developing and mature ToM can be enhanced by relatively short training programs (Lecce, Bianco, Demicheli, et al., 2014; Lecce et al., 2015). Enhanced generalizability of these effects could be gained by training schedules that take the multifaceted nature of ToM into account. Furthermore, a better understanding of the exact mechanisms that drive training success is needed to further enhance the efficiency of these programs (Hofmann et al., 2016). Of crucial importance in this context is a thorough investigation of the transfer effects of ToM trainings. These effects can shed light on the impact that mentalizing skills have outside of the laboratory, in terms of their contribution to enabling successful social interactions, as well as ensuring physical and mental health in everyday life.

Acknowledgments

This work was supported by the Emmy Noether Programme of the German Research Foundation (grant number BO 4962/1-1).

MANUSCRIPT 7

A REVISED INSTRUMENT FOR THE ASSESSMENT OF EMPATHY AND THEORY OF MIND IN ADOLESCENTS: INTRODUCING THE EMPATOM-Y

Christina Breil[1], Philipp Kanske [2,3], Roxana Pittig[4], Anne Böckler[1,2]

[1]Leibniz-University Hannover, Germany

[2]Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[3]Technische Universität Dresden, Germany

[4]Julius-Maximilians-University of Wuerzburg, Germany

This article has been published in *Behavior Research Methods* (2021) with DOI 10.3758/s13428-021-01589-3

Abstract

Empathy and Theory of Mind (ToM) are two core components of social understanding. The EmpaToM is a validated social video task that allows for independent manipulation and assessment of the two capacities. First applications revealed that empathy and ToM are dissociable constructs on a neuronal as well as on a behavioral level. As the EmpaToM has been designed for the assessment of social understanding in adults, it has a high degree of complexity and comprises topics that are inadequate for minors. For this reason, we designed a new version of the EmpaToM that is especially suited to measure empathy and ToM in youths. In experiment 1, we successfully validated the EmpaToM-Y on the original EmpaToM in an adult sample (N = 61), revealing a similar pattern of results across tasks and strong correlations of all constructs. As intended, the performance measure for ToM and the control condition of the EmpaToM-Y showed reduced difficulty. In experiment 2, we tested the feasibility of the EmpaToM-Y in a group of teenagers (N = 36). Results indicate a reliable empathy induction and higher demands of ToM questions for adolescents. We provide a promising task for future research targeting inter-individual variability of socio-cognitive and socio-affective capacities as well as their precursors and outcomes in healthy minors and clinical populations.

*Keywords***:** social cognition, social understanding, theory of mind, mentalizing, empathy, adolescence, development

Introduction

Today's youth is more closely connected, better educated and more diverse than any generation before. These chances bring about novel challenges. In a globalized world, problems arise at a bigger scale and constructive cooperation is more important than ever. Two key social capacities are necessary for this endeavor. First, feeling for somebody or sharing someone's affect, which is commonly referred to as "empathy" (de Vignemont & Singer, 2006; Oliver et al., 2018; Singer & Lamm, 2009), and second, the capability to represent other's intentions and beliefs, commonly referred to as "theory of mind" (ToM) or "mentalizing" (Frith & Frith, 2006). Even though these two concepts have many features in common, they can be clearly dissociated on a behavioral and on a neuronal level (Kanske et al., 2015b; Schurz et al., 2020).

As early as on the first day of their lives, human infants spontaneously respond to hearing other infants' cries (Martin & Clark, 1982). In parallel, the neural networks associated with empathy are subject to profound maturation processes until adulthood (Decety & Michalska, 2010). On a behavioral level, findings regarding age trends in empathy-related responding during adolescence are inconsistent: While Decety and Michalska (2010) found reduced intensity of pain perceptions in others with increasing age, other studies report an age-related increase in empathic responding, and some did not find any differences (Eisenberg et al., 2009).

There are profound inter-individual differences in adolescent empathy that remain stable across several decades (Allemand et al., 2015). These differences in empathy reflect on various other life domains: Impairments in empathic responding have been associated with aggression and criminal behavior across all age groups (Blair, 2018; van Hazebroek et al., 2017; van Zonneveld et al., 2017; Winter et al., 2017). In adolescents, empathy is negatively related to delinquency, bullying and externalizing problems – but positively related to numerous socially desirable characteristics, such as pro-social goals, social competence and supportive relationships (Eisenberg et al., 2009). Critically, the level of initial empathy as well as the degree and direction of development during adolescence predict inter-individual differences in social competence two decades later (Allemand et al., 2015) and an accumulation of adverse relationships in youths is considered an unspecific risk factor for psychopathologic development from adolescence to early adulthood (Adam et al., 2011). As such, adolescent empathy is not only a protective factor at the time being, but also an important resource for social functioning and mental health as an adult. Yet, the literature on empathy development from ages 12 – 18 is limited and findings have been inconsistent (Eisenberg et al., 2009), indicating that further research in this area is urgently needed.

In a similar vein, research in children suggests a reliable development of ToM capacities during the first years of life (Hughes & Leekam, 2004; Wellman et al., 2001) with first attempts of spontaneous perspective-taking at the age of 7 – 15 months (Baillargeon et al., 2010; Kovacs et al., 2010) and a progressive understanding of more complex forms of mentalizing throughout adolescence (Devine & Hughes, 2013) and adulthood (Dumontheil et al., 2010). Difficulties in ToM performance have been linked to a variety of psychological disorders, such as depression, social anxiety disorder, autism spectrum disorder (ASD) and schizophrenia (Berecz et al., 2016; Bora et al., 2009; Leppanen et al., 2018; Washburn et al., 2016).

So far, most research has focused on the early childhood and pre-school years (Baillargeon et al., 2010; Cadinu & Kiesner, 2000; Wellman et al., 2001), and on clinical populations with social deficits, for instance individuals with ASD (Altschuler et al., 2018; Baron-Cohen et al., 1985; Deschrijver et al., 2016) or schizophrenia (Bora et al., 2009; Frith & Corcoran, 1996). More recently, the neural underpinnings of mentalizing have received considerable attention and new paradigms have been developed to investigate inter-individual variability in healthy adults (Baksh et al., 2018; Murray et al., 2017; Schurz et al., 2014, 2020). In striking contrast, very little attention has been devoted to ToM development, its precursors and its outcomes in healthy teenagers. Pioneer fMRI studies show that activity during mentalizing processes in frontal regions decreases from adolescence to adulthood (Blakemore et al., 2007; Sebastian et al., 2012; Wang et al., 2006), which could be indicative of synaptic reorganization processes in the prefrontal cortex (Blakemore, 2008). One recent study demonstrated that only from the age of 10 – 12 years onwards do children begin to understand that two people can represent the exact same information differently. This type of reasoning has been shown to be protective of serious behavior problems and social conflict in high school (Weimer et al., 2017). Strong interactions between peer acceptance and social understanding have been demonstrated in children (Banerjee et al., 2011; Hughes & Leekam, 2004) and pre-adolescents (Bosacki & Wilde Astington, 2001). A thorough investigation of social understanding in teenagers is therefore highly necessary.

Critically, the endeavor of assessing the development of empathy and ToM in youths demands measures that allow for an assessment of the full range of skills that are required to prosper in the adolescent social system. The false belief task (Wimmer & Perner, 1983) is widely regarded as the litmus test for ToM, but is already mastered by normally developing children from the age of four years on (Wellman et al., 2001), and even the more complex variations of this or related paradigms are usually at ceiling in healthy adults (but see Keysar et al. (2003)). These issues lower the chances of capturing variance and improvement in mental state representation in healthy adult and adolescent samples. In this light, the development of complex ToM measures, such as the Edinburgh Social Cognition test (Baksh et al., 2018) and the EmpaToM (Kanske et al., 2015b), has been an important recent trend.

The EmpaToM is a promising tool, as it allows for simultaneous manipulation and assessment of empathy and ToM. By using naturalistic dynamic stimuli, the EmpaToM is akin to real-life situations and interactions, and its compatibility with physiological measures and imaging techniques allows for a full-range investigation of social cognition in healthy samples and clinical populations (Preckel et al., 2016). The task consists of short video sequences that depict an unknown person narrating an autobiographical episode. The episode is either of negative emotional valence, thereby eliciting an empathic response, or neutral as a control condition. Participants empathic tendency is derived from affect ratings after each video. ToM performance is measured by means of content-related questions on the previously seen video that either require mental perspective taking of the narrator (ToM), or not (nonToM). Hence, affect sharing and ToM are orthogonally manipulated in this task, comprising (i) negative and neutral videos for an assessment of subjective affect sharing of the participants and (ii) videos with or without a mentalizing component allowing for subsequent ToM questions and control questions on each story. The EmpaToM has been thoroughly validated, revealing specific brain-behavior relations for both capacities. Importantly, the task is sensitive for changes in social cognition across the adult lifespan (Reiter et al., 2017) and for plasticity induced by mental trainings (Trautwein et al., 2020). However, the EmpaToM in its present form is inappropriate for an assessment of empathy and ToM in adolescents for three reasons in particular. For one, this task encompasses several episodes that are inadequate for minors on an affective level. These episodes include war experiences, sexual and physical abuse, family tragedies and deadly accidents, and could lead to intolerable emotional distress in teenagers. Second, the EmpaToM has a high level of difficulty resulting from complex issues that are alluded to, but not explicitly named in the videos, and from questions that require common knowledge that may only be acquired with age. Finally, the EmpaToM mainly entails "adult topics" that could be difficult to imagine for teenagers. While empathizing with other persons even when they are in situations one cannot easily relate to is a core competence of social understanding and should hence not lower ecological validity, the exclusive implementation of such unfamiliar topics could lead to low motivation or even negligence at task execution.

In summary, even though social cognition likely continues to develop beyond the well-studied hallmarks during childhood, a thorough understanding of the representation of other people's minds in adolescents is still lacking. This gap is especially problematic because social competence is vital for a healthy and adaptive coming of age with intact peer relationships. Critically, the neglect of adolescent social cognition in research goes hand in hand with a shortage of appropriate tools for a comprehensive assessment of social understanding in healthy individuals of this age group.

We aimed to fill this gap by providing a new instrument for the assessment of empathic affect sharing and ToM in teenage samples. Our goal was to design a measure that allows for a full-range investigation of adolescent social understanding with inter-individual variability in a naturalistic setting. To this end, we created a version of the EmpaToM that is especially tailored to the abilities and needs of a younger age group, namely (i) eliciting sufficient inter-individual variance while being generally solvable by teens and (ii) age-appropriate content of the stories with (iii) younger narrators talking about issues that teenagers can more easily relate to. For our new instrument, the EmpaToM-Y, we kept the general design of the EmpaToM and developed new videos and questions that are less complex and more appropriate for adolescents.

Because the original EmpaToM has been extensively validated, we first behaviorally tested the EmpaToM-Y on the existing measure in an adult sample group (N = 61, experiment 1). We decided for this age group because the original EmpaToM is inappropriate for adolescents. We therefore conducted a second experiment (N = 36, experiment 2) in which we assessed the feasibility of our new instrument in a sample of adolescents. For further external validation in experiment 2, we added a standardized measure of self-reported empathy and ToM.

Experiment 1

Method

We report how the sample size was determined, all manipulations and measures that are collected and all data exclusions (Simmons et al., 2012). In experiment 1, we apply the

EmpaToM-Y together with the existing EmpaToM in an adult sample to behaviorally validate the new measures.

Participants

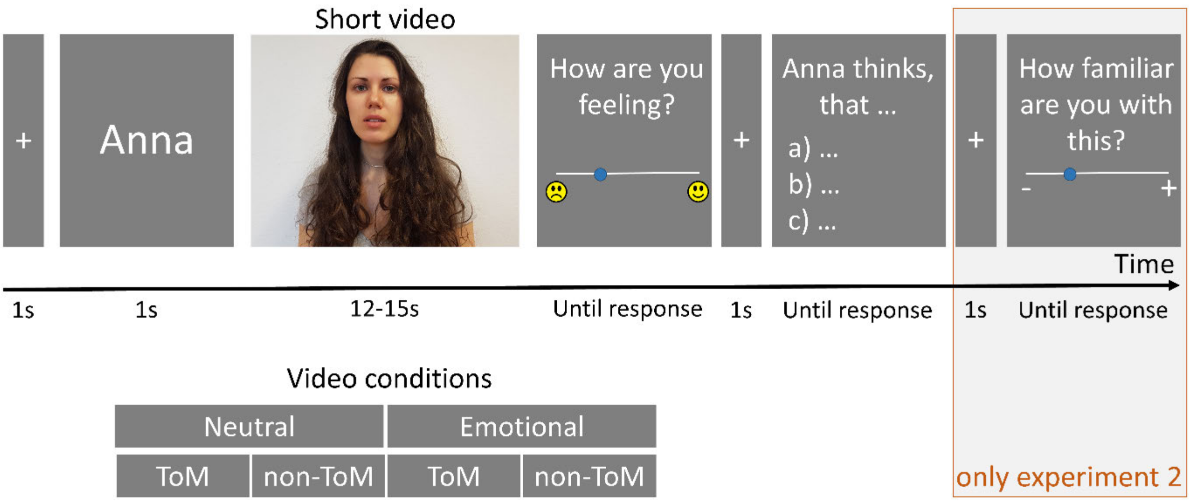
Ninety-nine participants took part in experiment 1 in return for course credit or 10€ and completed an informed consent form. All participants were recruited via the participant database of the University of Wuerzburg, were fluent in German and reported normal or corrected to normal vision. We had to exclude the data of 18 participants because they reported to be acquainted with one of the persons that was displayed in the videos, or because of language barriers. Due to technical difficulties with one of the testing computers, the data of 17 further participants was corrupt and could not be entered into the analysis. Of the remaining 64 participants, three data sets were removed due to implausibly high error rates in the ToM and nonToM questions (above 33%), leaving 61 participants (mean age = 28.7, SD = 8.88, range: 20-56; 47 females; 57 right-handed) for the final analysis. The present study is compliant with the ethical standards of the 1964 Declaration of Helsinki regarding the treatment of human participants in research and was approved by the local ethics committee.

Task

The EmpaToM-Y is a German video-based task that simultaneously manipulates empathic affect sharing and ToM. Each trial started with a fixation cross (1s) after which the name of the person who is speaking in the following video was displayed (1s; Figure 1). Each video lasted about 15s and presented an unknown character allegedly recounting an autobiographical episode. The videos differed in terms of valence (neutral or negative) and ToM-affordance (ToM or nonToM). After each video, participants were required to rate their own emotional state on a rating scale ranging from negative to positive (affect rating). We derived a measure for the tendency to share others‘ affect (affect sharing tendency) by comparing the participants' rating after negative versus neutral videos. The affect rating was followed by a multiple-choice question regarding the video content. Each question had three response options (one correct answer) that appeared in randomized order. The questions either entailed mental perspective-taking (ToM) or factual reasoning (nonToM). The EmpaToM-Y consisted of 40 trials (ten for each combination video valence and ToM requirement), with four videos per narrator (one per condition). Forty trials of the original EmpaToM task were presented intermixed with the 40 trials of the EmpaToM-Y in randomized order with a short break every 20 trials.

Figure 1

Trial sequence of experiment 1



Note. After a fixation cross and the name of the person in the video are displayed for 1s each, a short video (12-15s) is played. The video is followed by a rating scale measuring empathic affect and a multiple-choice question for ToM assessment or factual reasoning, both displayed until a response is made. In experiment 2, this was followed by a second rating question to assess familiarity with the situation in the video.

Stimuli

Twenty-four novel videos, six for each condition, were created specifically for the EmpaToM-Y. Each episode was designed to resemble an extract of a presumably longer dyadic conversation and was either neutral or emotionally negative, thus entailing experiences of disappointment, loss or regret. We took special care to avoid any age-inappropriate content, such as war experience, heavy violence or family drama, and to include more age-related topics like school life or peer group experiences. An example story and the corresponding question for each condition can be found in Appendix A. Six young amateur actors, three females and three males, were recruited for the shooting of the videos and were compensated with payment (10€/hour). One actress was Afro-German, the other five were Caucasian. The camera, light and audio settings were held constant throughout all of the videos. The film footage was cut to a length of 12-15s per video and converted to MP4 using Windows Movie Maker (version 2012; Microsoft, Redmond, Washington, USA). Sixteen videos (four per condition) of the original EmpaToM were suitable for teenage participants and were hence included in the EmpaToM-Y. These videos displayed two male and two female Caucasian adults. The corresponding questions were reduced in complexity. Two videos with modified questions from the original EmpaToM served as training trials. Crucially, none of the videos that was used for the EmpaToM-Y appeared in the EmpaToM version applied here.

For each trial of the EmpaToM-Y, we created a multiple-choice question with one correct response option and two distractor options. ToM questions referred to mental state aspects of the narrator, such as thoughts, goals or intentions, that were not explicitly mentioned in the video. Hence, identifying the correct answer to ToM-questions required taking the mental perspective of the previously seen person. Control questions entailed no ToM processes but similarly complex factual reasoning. We devoted considerable effort to ensure a constant level of linguistic demands across the total trials of all four conditions and matched the conditions regarding syntactic complexity and number of words (Table 1). Similarly, the length of the answers was equal across conditions (all Fs ≤ 1). For the trials of the EmpaToM, the original questions were used.

Table 1

Results of analyses on grammatical complexity of the EmpaToM-Y

Dependent variable	Test statistic	*p*-value	Effect size ($\eta^{2)}$
Number of words	$F(1, 14) = 0.00$	.948	< .01
Frequency of future tense	$F(1, 14) = 0.08$	.787	.01
Frequency of past tense	$F(1, 14) = 1.91$	.189	.12
Number of conditional sentences	$F(1, 14) = 1.00$	.334	.07
Frequency of subordinate clauses	$F(1, 14) = 0.38$	.546	.03

Note. For all questions, the number of words, frequencies of future and past tense, number of conditional sentences and the frequency of subordinate clauses were submitted to separate one-way ANOVAs with the within-subject factor condition (neutral-nonToM, neutral-ToM, negative-nonToM, negative-ToM).

Procedure

All participants provided written informed consent to the experimental procedures. For the experiment, participants sat 80cm away from a 60-cm monitor and were provided with a pair of over-ear headphones. The experiment started with a standardized instructions screen, followed by two training trials. For the affect rating, participants were specifically instructed to spontaneously indicate their own emotional state with respect to the video, but to carefully choose their answer of the multiple-choice question. The training block and each of the four test blocks could be started self-paced by pressing the space bar. After the training trials, the participants were given the chance to pose questions to the experimenter, and they had the opportunity to take a break between the blocks. Altogether, it took about one hour to complete the experiment.

Analyses

Mean absolute affect ratings, mean error rates and mean RTs were submitted to three separate 2 (ToM requirement: ToM, nonToM) × 2 (video valence: negative, neutral) × 2 (task: EmpaToM-Y, EmpaToM) repeated measures ANOVAs in order to assess (i) whether the EmpaToM-Y was in fact easier, hence, elicited lower error rates and faster responses than the original EmpaToM and (ii) effects of the valence manipulation were comparable across tasks. Post-hoc *t*-tests with Bonferroni correction were performed to resolve ANOVA interaction effects. Additionally, we investigated the effects of video valence and ToM requirements for each of the two instruments individually. The results of these separate ANOVAs can be found in Tables B1-B6 in Appendix B. We report generalized η^2 as effect size.

For each participant, we calculated individual empathic affect sharing by subtracting the mean affect rating after negative videos from the mean affect rating after neutral videos. Larger values hence indicate a stronger tendency to be influenced by the emotionality of the video and represent greater empathic affect sharing. While difference scores have been criticized for low test-retest reliabilities (Paap & Sawi, 2016) they provide an option to control for each participant’s baseline and led to reasonable outcomes in previous studies with a similar design (e.g. Bernhardt, Klimecki, et al., 2014). This difference score was calculated separately for trials of the EmpaToM and the EmpaToM-Y, resulting in two values for each participant. We calculated the Pearson correlation between the affect sharing measures of both tasks. Furthermore, we calculated the following Pearson correlations between the EmpaToM and the EmpaToM-Y: (i) mean error rates of ToM questions, (ii) mean error rates of nonToM questions, (iii) mean response times (RTs) for ToM questions and (iv) mean RTs for nonToM questions. RT was defined as the time from question onset until key press.

Following previous studies (Kanske et al., 2015b, 2016; Trautwein et al., 2020), we additionally generated composite measures of ToM and nonToM performance by z-transforming the error rates and mean RTs and taking the average of both. Again, we did this separately for the EmpaToM and the EmpaToM-Y to calculate the Pearson correlation between them as well as the partial correlation for the ToM composite values controlling for the nonToM composite values.

Finally, we calculated the internal consistency (Cronbach α) and the item total correlation of each instrument as well as item-specific difficulty and reliability values.

Results

The datasets generated and analyzed during the current study are available in the Open Science framework repository (DOI: 10.17605/OSF.IO/8Y95B). All stories and questions as well as an example video of each condition can be found at the same location. The full video set of the EmpaToM-Y is available on request (DOI: 10.17605/OSF.IO/3RYSN). Both experiments were not pre-registered.

Mean affect ratings, error rates and RTs for each condition of the EmpaToM and the EmpaToM-Y are visualized in Figure 2 and summarized in Table C1 in the Appendix C.

Combined ANOVA

Affect ratings. In a conjunct analysis of the EmpaToM-Y and the EmpaToM, participants reported significantly more negative affect after videos with negative valence than after neutral videos, reflected in a main effect of video valence ($F(1, 60) = 351.77$, $p < .001$, $\eta^2 = .85$). This pattern is in line with earlier findings and suggests the effectiveness of the empathy induction. Participants also reported more negative affect after nonToM-videos, leading to a main effect of ToM requirement ($F(1, 60) = 58.60$, $p < .001$, $\eta^2 = .49$). The between-subjects factor task (EmpaToM-Y, EmpaToM) was significant ($F(1, 60) = 101.27$, $p < .001$, $\eta^2 = .63$), indicating overall more negative affect in the EmpaToM. This finding likely reflects our decision to remove videos reporting serious negative instances such as abuse and war experiences from the EmpaToM-Y.

There was a significant interaction effect of ToM requirement × video valence (F(1, 60) = 18.94, $p < .001$, $\eta^2 = .24$), indicating that the difference between ratings after ToM versus after nonToM videos decreased from neutral to negative, but remained significant (neutral: t(121) = 7.25, $p < .001$; negative: t(121) = 3.295, $p < .001$). Furthermore, a significant video valence × task interaction effect (F(1, 60) = 17.85, $p < .001$, $\eta^2 = .23$) indicates that the difference in affect ratings between the EmpaToM and the EmpaToM-Y was larger after videos with negative valence, but significant in both conditions (neutral: t(121) = 3.90, $p < .001$; negative: t(121) = 10.46, $p < .001$). No other interactions reached significance (ToM × task: $p = .223$; ToM × video valence × task: $p = .087$).

Performance. Participants produced significantly more errors in the original EmpaToM, reflected in a main effect of task (F(1, 60) = 276.66, $p < .001$, $\eta^2 = .82$). Hence, as intended, the EmpaToM-Y had reduced levels of difficulty. We also found more errors for neutral videos compared to negative videos, reflected in a main effect of video valence (F(1, 60) = 15.80, $p < .001$, $\eta^2 = .21$). The main effect of ToM requirement was not significant (p = .134).

We found a significant interaction of ToM requirement × video valence (F(1, 60) = 10.23, $p = .002$, $\eta^2 = .15$). This interaction was due to higher error rates for ToM questions, but not for nonToM questions, after neutral compared to after negative videos (ToM: t(121) = 4.553, $p < .001$; nonToM: $p = .448$). In addition, there was a significant interaction of video valence × task (F(1, 60) = 5.47, $p = .023$, $\eta^2 = .08$), resulting from more errors in in the EmpaToM, but not the EmpaToM-Y after neutral than after negative videos (EmpaToM: t(121) = 3.485, $p = .004$; EmpaToM-Y: $p = .527$). Critically, the ToM requirement × task interaction was not significant ($p = .214$), indicating that the ToM manipulation had similar effects on performance in both tasks.

There was a significant three-way interaction (F(1, 60) = 32.21, $p < .001$, $\eta^2 = .35$), resulting from an advantage for neutral videos at nonToM questions in the EmpaToM-Y (t(60) = 3.29, $p = .006$), but at ToM questions in the EmpaToM (t(60) = 6.567, $p < .001$).

Effects in error rates were paralleled by significantly faster responses for nonToM questions, after negative videos and for questions of the EmpaToM-Y, reflected in significant main effects ToM (F(1, 60) = 19.21, $p < .001$, $\eta^2 = .24$), video valence (F(1, 60) = 4.20, $p = .045$, $\eta^2 = .07$) and task (F(1, 60) = 303.65, $p < .001$, $\eta^2 = .84$), respectively.

The interaction effect of ToM requirement × video valence (F(1, 60) = 11.01, $p < .001$, $\eta^2 = .15$) was significant, indicating faster responses to nonToM questions after neutral videos (t(123) = -2.92, $p = .016$), but faster responses to ToM questions after negative videos (t(121) = 2.56, $p = .007$). A significant video valence × task interaction (F(1, 60) = 12.57, $p = .002$, $\eta^2 = .17$) suggested faster responses after negative videos in the EmpaToM (t(121) = 2.75, $p = .020$), but no difference in the EmpaToM-Y ($p = .158$). The two-way interaction of ToM requirement × task was not significant ($p = .839$), indicating similar effects of the ToM manipulation across tasks. There was a significant three-way interaction (F(1, 60) = 11.41, $p < .001$, $\eta^2 = .16$), indicating that the interaction effect of ToM requirement × video valence was significant only for the EmpaToM with faster responses to ToM questions after negative videos (t(60) = 4.34, p < . 001).

Taken together, these results indicate effective empathy inductions in both tasks. Also, we successfully reduced task difficulty in the EmpaToM-Y, reflected in both reduced errors and RTs at ToM and nonToM questions. Finally, no main effects of ToM requirements on error rates suggest that overall levels of difficulty were comparable for ToM and nonToM questions in both tasks.

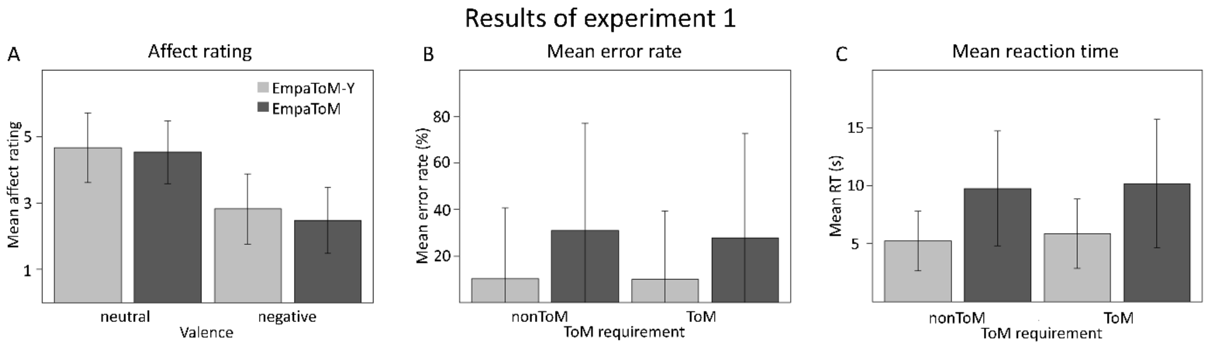
Correlations

The correlations of affect sharing tendency and ToM performance are presented in Figure 3. The mean affect sharing tendency was 2.18±.80 (4.73-2.55) for the EmpaToM-Y and 2.96±0.87 (neutral-negative: 4.54-2.48) for the EmpaToM. The Pearson correlation between the two sets was $r = .901$ ($p < .001$).

The mean error rate for ToM questions was 10.65±30.85% for the EmpaToM-Y and 28.39±45.11% for the EmpaToM, with a Pearson correlation of $r = .617$ ($p < .001$). The mean error rate for nonToM questions was 11.05±31.36% for the EmpaToM-Y and 31.7±46.54% for the EmpaToM, and the Pearson correlation was $r = .637$ ($p < .001$). When we controlled nonToM on the relationship between ToM responses in the EmpaToM-Y and the EmpaToM, we found a significant partial correlation of $r = .489$ ($p < .001$).

Figure 2

Absolute affect ratings, error rates and RTs per condition in the EmpaToM and the EmpaToM-Y



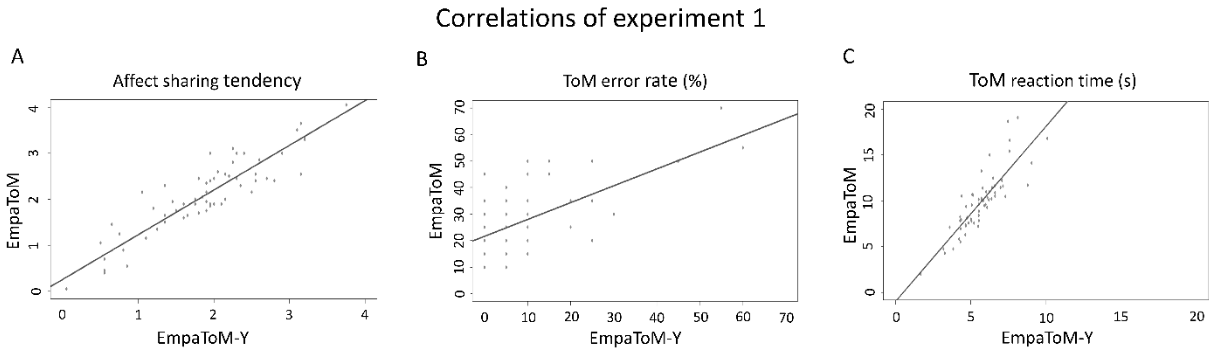
Note. ToM = Theory of Mind. RT = response time. Error bars represent standard errors. *Panel A*: Mean affect ratings on a 7-point scale. *Panel B*: Mean error rates at questions in %. *Panel C*: Mean response times to questions in seconds.

For the EmpaToM-Y, the mean RT for ToM questions was 5.71±1.57s, whereas it was 10.01±5.56s for the EmpaToM. The Pearson moment correlation between the tasks was r =

.849 (p < .001). The mean RT for nonToM questions was 5.19±2.37s for the EmpaToM-Y and 9.61±5.56s for the EmpaToM, with a Pearson moment correlation of r = .628 (p < .001). The partial correlation was significant with r = 783 (p < .001).

Figure 3

Correlations of affect sharing tendencies as well as errors and RTs in ToM questions between the EmpaToM and the EmpaToM-Y



Note. ToM = Theory of Mind. *Panel A*: Correlation of affect sharing tendency (difference between ratings after neutral and negative videos) between the EmpaToM and the EmpaToM-Y. Higher values indicate a higher individual tendency for empathic affect sharing. *Panel B*: Correlation of individual percentages of error rates for ToM questions between the two tasks. *Panel C*: Correlation of mean response times for questions with ToM requirements between both measures.

The Pearson moment correlation of the composite scores between the EmpaToM-Y and the EmpaToM was r = .641 (p < .001) for ToM performance and r = .494 (p < .001) for nonToM questions, with a partial correlation of r = .642 (p < . 001).

Overall, we found significant, medium to strong correlations of all measures of empathic affect sharing and ToM between the EmpaToM-Y and the EmpaToM, suggesting that our novel task measures the same constructs as the thoroughly validated EmpaToM.

Item analyses

The mean error rate of the EmpaToM-Y was 11% both for ToM questions (2-33%) and for nonToM questions (5-39%). The internal consistency of ToM questions was α = .82 (standardized Cronbach α) with an average inter-item correlation of r = .19. The correlations between individual items and the total scale ranged from r = .21 to r = .81. NonToM questions had an internal consistency of α = .88 with an average inter-item correlation of r = .27. There was a range of correlations between single items and the total scale of r = .16 to r = .85.

The mean error rate of the EmpaToM was 29% (5-59%) for ToM questions and 32% (13-59%) for nonToM questions. ToM questions had an internal consistency of α = .57 and an average inter-item correlation of r = .06. The correlation of individual items with the total scale ranged between r = .07 and r = .63. NonToM questions had an internal consistency of α = .67 with an average inter-item correlation of r = .09. Single items had a correlation with the total nonToM scale between r = .15 and r = .58.

In sum, the results indicate strong internal consistency of both ToM and nonToM scales of our new measure.

Discussion

Showing strong correlations with an established measure of empathic affect sharing and ToM in adults, experiment 1 demonstrates the validity of our new task (Kanske et al., 2015b; Schober et al., 2018). Reduced task demands make the EmpaToM-Y a useful and promising tool for the investigation of social understanding in adolescent samples.

Two findings in particular suggest the validity of assessment of empathic affect sharing in the EmpaToM-Y: First, we found a high correlation between the respective measures of the two instruments. Subjective affect ratings in the EmpaToM are related to performance in other established paradigms for the assessment of empathy (the Socio-affective Video Task; Klimecki et al., 2013) and to neural activation in networks that are commonly associated with empathy (Kanske et al., 2015b). This finding is substantiated by the fact that the valence of the videos affected emotion ratings in both instruments. Participants in the present experiment indicated to feel significantly more negative after negative videos compared to after neutral videos. Unsurprisingly, this effect was more pronounced for the EmpaToM, given that this task contains traumatic episodes which are inherently more tragic and hence empathy-inducing than the toned down stories in the EmpaToM-Y. Note that one core goal of our endeavor to create a version of the EmpaToM that is suitable for adolescents was to exclude traumatic episodes.

We also demonstrate that our new tool is valid for the assessment of ToM by showing adequate correlations with the corresponding measure in the EmpaToM (Schober et al., 2018). This relation was evident both in error rates and RTs for questions that required cognitive perspective taking, and this pattern held even when the correlation between ToM performance in the two measures was controlled for nonToM performance. ToM performance in the EmpaToM has been shown to be related to performance in an established measure of high-level ToM (the Kinderman Imposing Memory task; Kinderman et al., 1998) and the task induced neural activation in regions that are reliably associated with ToM (Kanske et al., 2015b). We can thus conclude that the EmpaToM-Y validly measures ToM performance.

Importantly, the results show that the EmpaToM-Y has reduced task demands compared to the EmpaToM. The latter task was designed for adult samples and could be too demanding and tedious for adolescents. Our new task is considerably easier as evidenced by lower error rates and faster responses, yet it is still capable of revealing inter-individual differences in adults. No item has been answered either correctly or incorrectly by all participants of experiment 1, indicating that every trial of the EmpaToM-Y is suitable to detect inter-individual differences. This pattern is paralleled by convincing internal consistencies of both ToM and nonToM items (Hays & Revicki, 2005) and appropriate item-scale correlations (Piedmont, 2014).

In order to directly target the suitability of our new task for adolescents, experiment 2 applied the EmpaToM-Y to a sample of teenagers aged 14 to 18 years.

Experiment 2

In experiment 2, we tested the feasibility and appropriateness of the EmpaToM-Y in the intended age group. We employed the task in a group of adolescents and included an established questionnaire for the assessment of socio-cognitive and socio-affective understanding, the German version of the Interpersonal Reactivity Index (i.e. the Saarbrucken Personality Questionnaire, SPQ; Paulus, 2006), for further external validation.

Method

Participants

Forty-three adolescent participants were publicly recruited for experiment 2 and were compensated with payment (10€/hour). Prior to the testing day, all participants were asked to report about mental and neurological disorders as well as about medication. Five participants reported no clinical diagnosis at pre-screening but did so at the test appointment. Due to technical difficulties, the data of further two participants was missing, leaving 36 participants (14-18 years; mean age = 16.13, SD = 1.40; 24 females) for the final analysis. All of these participants were healthy and unmedicated and had normal or corrected to normal vision. The parents or legal guardians of minor participants provided written informed consent prior to the experiment. Participants of full age provided written informed consent themselves. The present study is compliant with the ethical standards of the 1964 Declaration of Helsinki regarding the treatment of human participants in research and was approved by the local ethics committee.

Measures

EmpaToM-Y. Only the EmpaToM-Y was employed in experiment 2. This task consisted of three training trials followed by two test blocks of 20 trials each. The trial sequence was similar to experiment 1, but we added a third question at the end of each trial (see Figure 1). Specifically, participants were asked to rate how familiar they were with the situation that was displayed in the previous video. This served as an indicator of how appropriate the videos are for adolescent samples.

Saarbrucken Personality Questionnaire (SPQ). The SPQ is the revised German version of the Interpersonal Reactivity Index (IRI; Davis, 1980), consisting of the four scales perspective taking (PT), fantasy (FS), empathic concern (EC), and personal distress (PD), with four items per scale. Example items of the IRI can be found in Appendix D. Three of these scales, namely FS, EC and PD, are related to empathy (Paulus, 2006). The remaining scale, PT, is described as the capacity to spontaneously take the psychological perspective of another person and is hence more closely related to ToM. We hypothesized correlations of PT with ToM performance in the EmpaToM-Y, and correlations of the scales FS, EC and PD of the SPQ with empathy tendency in our task. One advantage of the SPQ is the short time it takes to complete the questionnaire: the 16 items are answered in less than 10 minutes, giving us the opportunity to add an external validation measure without excessively prolonging the total duration of the experiment. The SPQ was administered prior to the EmpaToM-Y for half of the participants and after the task for the remaining half (randomized). Altogether, it took about one hour to complete the experiment.

Physiological data. In order to test for physiological responses to emotional videos we recorded electrodermal activity and pupillometry during the videos of the EmpaToM-Y.

Since we did not find any meaningful effects, the details about data collection and analysis as well as results are described in the Appendix E.

Analyses

As in experiment 1, we calculated separate 2 (video valence: neutral, negative) × 2 (ToM requirement: ToM, nonToM) repeated measures ANOVAs for the following dependent variables: (i) affect ratings, (ii) accuracy, (iii) RTs and (iv) composite scores. We created the latter for ToM and control questions by z-transforming mean correct RTs, defined as the time between question onset and the first key press, and mean amount of error rates, and then taking the average of both. Post-hoc t-tests were performed to resolve ANOVA interaction effects. Furthermore, we tested for correlations between affect ratings, ToM performance and familiarity ratings. We report generalized η^2 as effect size.

Results

ANOVAs

The mean affect ratings as well as mean accuracy rates, RTs and composite scores for each condition are visualized in Figure 4 and listed in Table C2 in the appendix C.

Affect ratings. The mean affect sharing tendency was 2.81 (SD = 1.21). Individual affect sharing tendency ranged from -0.25 to 4.80. Adolescents in experiment 2 reported to feel significantly worse after negative videos compared to after neutral videos, reflected in a main effect of video valence ($F(1, 35) = 196.09$, $p < .001$, $\eta^2 = .688$) and confirming a successful valence manipulation in our task. Ratings were also less positive after videos of the ToM conditions, leading to a main effect of ToM requirement ($F(1, 35) = 26.58$, $p < .001$, $\eta^2 = .037$). There was a significant interaction of video valence × ToM requirement ($F(1, 35) = 12.67$, $p = .001$, $\eta^2 = .016$) due to a larger affect sharing tendency after nonToM than after ToM questions (nonToM: $t(35) = -13.51$, $p < .001$; ToM: $t(35) = -13.02$, $p < .001$).

Performance. The mean error rates were 14.4% for ToM questions and 6.5% for nonToM questions. Individual error rates for ToM questions ranged between 0.0% and 35.2%, whereas errors ranged between 0% and 14.8% for nonToM questions. For a large majority of 26 participants, ToM questions were more difficult than nonToM questions. ToM and nonToM questions were equally difficult for 6 participants and for only 4 participants, ToM questions were easier than nonToM questions. Consistently, 10 participants answered all nonToM questions correctly whereas only 2 participants achieved this for ToM questions. One participant performed at ceiling in both conditions.

Hence, in contrast to our findings in adults in experiment 1, adolescents produced significantly more errors for ToM questions than for nonToM questions, reflected in a main effect of ToM requirement ($F(1,35) = 33.34$, $p < .001$, $\eta^2 = .135$). No main effect of video valence was found ($p = .381$). The video valence × ToM requirement interaction effect was significant due to a larger difference in error rates between ToM and nonToM questions after emotional videos ($F(1,35) = 4.92$, $p = .033$, $\eta^2 = .029$; emotional: $t(35) = 5.04$, $p < .001$; neutral: $t(35) = 2.36$, $p < .001$).

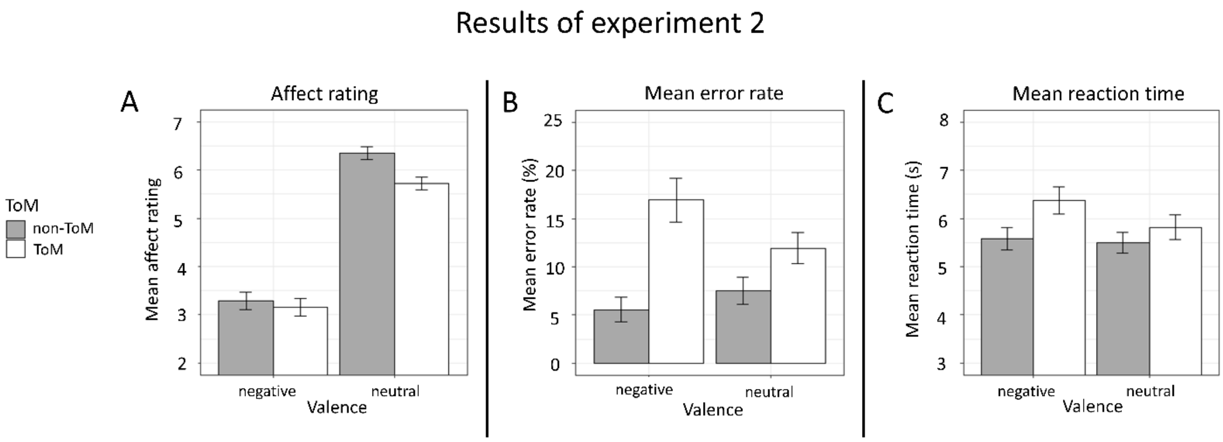
Effects in error rates were paralleled by significantly longer RTs to ToM questions than to nonToM questions, reflected in a main effect of ToM requirement ($F(1,35) = 17.40$, $p < .001$, $\eta^2 = .036$). The valence manipulation induced overall longer response latencies after negative videos ($F(1,35) = 13.66$, $p = .001$, $\eta^2 = .012$). We found a significant two-way interaction ($F(1,35) = 4.82$, $p = .035$, $\eta^2 = .006$), reflected in prolonged responses at ToM questions after emotional videos ($t(35) = -4.55$, $p < .001$), but not after neutral videos ($p = .121$).

Similar effects emerged in the analysis of the composite scores: Lower scores, indicating better performance, were found for nonToM questions, leading to a main effect of ToM requirement ($F(1,35) = 65.27$, $p < .001$, $\eta^2 = .162$). A significant main effect of video valence was due to lower scores for questions after neutral videos ($F(1,35) = 4.55$, $p = .04$, η^2

= .02). A significant two-way interaction ($F(1,35) = 10.49$, $p = .003$, $\eta^2 = .035$) indicated once more that the contrast in difficulty was larger after emotional videos (emotional: $t(35) = -7.21$, $p < .001$; neutral: $t(35) = -3.48$, $p < .001$).

Figure 4

Affect rating and performance results by condition of the EmpaToM-Y in the adolescent sample of experiment 2



Note. ToM = Theory of Mind *Panel A*: Mean affect ratings on a 9-point scale. *Panel B*: Mean error rates at questions in %. *Panel C*: Mean response times to questions in seconds.

Familiarity ratings. The overall familiarity rating was 4.77 with mean item ratings between 1.92 (SD = 1.34) and 7.03 (SD = 1.83) points on a nine-point scale. The SD of items ranged between 1.34 (M = 1.92) and 2.78 (M = 4.06). Mean familiarity ratings and SD of all items are listed in Table F4 in Appendix F.

Correlations

More positive valence was reported after videos that were rated as more familiar ($r = .136$, $p < .001$). High familiarity with a situation also was related to faster responses at ToM questions ($r = -.112$, $p = .005$). RTs and error rates for ToM questions were positively related, indicating that questions that were more likely to be answered correctly also were answered faster ($r = .217$, $p < .001$). No other correlations were significant after Bonferroni correction (all $p > .05$). Importantly, ratings of affect were unrelated to performance at ToM questions (accuracy: $p = .371$; RT: $p = .27$).

Item analysis

In the adolescent sample, the mean error rate of the EmpaToM-Y was 6.6% (0.0-27.8%) for nonToM questions and 14.4% for ToM questions (0.0-52.8%). Four of the nonToM items and two of the ToM items were always answered correctly but none was unsolvable for all participants. Mean error rates of individual items ranged between 0.0 and 52.8%. The internal consistency of ToM questions was $\alpha = .35$ (standardized Cronbach α) with an average inter-item correlation of $r = .03$. The correlations between individual items and the total scale ranged from $r = -.08$ to $r = .66$. NonToM questions had an internal consistency of $\alpha = .08$ with an average inter-item correlation of $r = .01$. There was a range of correlations between single items and the total scale of $r = .06$ to $r = .52$. Taken together, these results indicate that adolescent samples are heterogenous, producing more variance in ToM and nonToM questions than observed in adults (experiment 1).

SPQ

None of the hypothesized correlations between behavioral measures of the EmpaToM-Y and scales of the SPQ were significant (all $p > .05$). In particular, corrected for multiple testing, the scales FS, EC and PD of the SPQ were unrelated to affect sharing tendency of the EmpaToM-Y (FS: $p = .152$; EC: $p = .162$; PD: $p = .085$), and the scale PT was unrelated to ToM accuracy ($p = .611$) and RTs ($p = .825$).

Discussion

Experiment 2 demonstrates the adequacy of the EmpaToM-Y for the intended age group by showing general feasibility of the task and sound assessment of empathic affect sharing and ToM with inter-individual variance in adolescents.

The valence manipulation of the EmpaToM-Y induced measurable empathic responses: Participants in experiment 2 indicated to feel significantly more negative after videos with negative valence compared to after neutral videos.

The overall performance in experiment 2 suggests that the EmpaToM-Y is feasible for adolescents. This pattern fits our finding from experiment 1 that the new task is less difficult compared to the original EmpaToM. However, in contrast to results from adults, we found that ToM questions were generally more demanding than control questions for adolescent participants. This effect was evident on all performance measures and substantiates the notion that ToM capacity is still developing during adolescence (Blakemore, 2008; Blakemore et al., 2007; Sebastian et al., 2012; Symeonidou et al., 2020; Wang et al., 2006). While demand differences between ToM and control questions for this age group could constitute a confound in fMRI studies, they offer the opportunity to capture the ToM progression on a behavioral level and contribute to a holistic understanding of social cognition development across the lifespan. Interestingly, once developed, ToM appears to remain relatively stable and even seems to be protected from the overall cognitive decline in the elderly (Reiter et al., 2017).

Given the abovementioned finding that ToM capacity is still developing throughout adolescence, it is reasonable to expect a wide variability in individual ToM performance. As evident in better performance for nonToM questions, ToM was not yet fully emerged in the present sample of adolescents and, consequently, inter-individual variance was enhanced. Bearing this in mind, it is not surprising that we found relatively low values of internal consistency and item-scale correlations in experiment 2 while experiment 1 showed good internal consistency for ToM performance in adults. Furthermore, prior analyses with adult samples suggest that the items of the EmpaToM are representative for the respective item populations, producing consistent patterns of brain activation for empathy and ToM across item- and participant-wise analyses (Tholen et al., 2020). Given the great conceptual and empirical overlap between the two tasks (see experiment 1), it can be assumed that the same applies for the EmpaToM-Y.

Importantly, and in line with previous findings in adults (Kanske et al., 2016), ToM performance was unrelated to the tendency to share others' affective states, indicating a successful orthogonal manipulation of empathy and ToM in our new task. This feature allows to assess the development of both constructs independently from each other. The notion of a conceptual dissociation of empathy and ToM is becoming increasingly popular and has been empirically supported in various domains. First, research suggests independent developmental progress of the two capacities, with ToM preceding empathy in children (Brown et al., 2017), and empathy outliving ToM in older adults (Reiter et al., 2017). Second, a range of mental dysfunctions is known to selectively affect only one aspect of social cognition. The most profound example is a dissociation of social cognitive deficits in ASD and alexithymia (emotion description inability), with an impact of ASD on brain networks related to ToM but not empathy, while the opposite pattern is found for alexithymia (Bernhardt, Valk, et al., 2014; Santiesteban et al., 2021). And finally, evidence accumulates that, even in the typically developing brain, cognitive and affective networks in the social brain diverge (Kanske et al., 2015b, 2016; Singer, 2006) and can be selectively promoted by specific training modules (Trautwein et al., 2020; Valk et al., 2017). Interestingly, while empathy and ToM are two clearly dissociable tendencies that seem independent in terms of their neural underpinnings and inter-individual variance (for a review, see Stietz et al., 2019), some findings suggest that they interact on an intraindividual level. For instance, empathizing might be prioritized in highly emotional situations, which can hamper ToM performance in this instance (Kanske et al., 2016). For a better understanding of the orchestration of these social capacities within a given person and situation, the simultaneous assessment of these tendencies is critical. Also, for a more thorough understanding of the interplay of empathy and ToM development, a simultaneous assessment of both capacities in different age groups is necessary. The

EmpaToM-Y is a promising tool for this endeavor as it allows to pinpoint inter-individual variance in both components of social understanding in teenagers

Overall, the familiarity ratings show that the items represent circumstances that adolescents can relate to. There seems to be substantial inter-individual variance in the degree to which participants were familiar with the various situations presented in the videos. This pattern makes the EmpaToM-Y a well-suited task for the assessment of social understanding, because it allows probing these capacities not only in well-known situations, but also when encountering people living in and experiencing circumstances that differ from one's own. In fact, correlation analyses suggest that high familiarity with a situation might facilitate empathic affect and mental perspective taking. Future studies could use this additional variable to estimate the effect of between-group differences in experiences on social cognition.

We found none of the hypothesized correlations between measures of the EmpaToM-Y and scales of the SPQ. We do not believe, however, that this seriously undermines the validity of our new task. Social cognition is a complex and multifaceted construct and an absence of intercorrelations even between well-established measures is a pattern that has been found before (Dziobek et al., 2006; Osterhaus et al., 2016). Critically, while we assessed actual empathic affect sharing in the EmpaToM-Y, the SPQ is a measure of a person's conception of her- or himself. Self-reports have been shown to be unrelated to actual behavior in other domains of social cognition, such as altruism (Böckler et al., 2016), and a critical self-reflection of one's own social cognition capacities could be particularly difficult for adolescents. Furthermore, while the empathy manipulation in our task reflects a psychometric state, the SPQ is a measure of trait empathy (Ze et al., 2014). Finally, a missing relation between the measures could partly be explained by wide-ranging differences in formal aspects of the tasks which have been noted to be critical determinators of the outcome in ToM assessment (Breil & Böckler, 2020) and which should be investigated in future studies with larger samples. As mentioned above, we found strong correlations with an established measure of empathic affect sharing and ToM in experiment 1 (Kanske et al., 2015b). Taken together, we believe that, despite the missing link to scales of the SPQ, the EmpaToM-Y is a valid and appropriate tool for the assessment of empathic affect sharing and ToM in adolescent samples.

General discussion

The present study introduces a novel instrument for the simultaneous assessment of empathic affect sharing and ToM in adolescent samples. In experiment 1, we successfully validated the new task on an established measure of social cognition in a group of adults. In experiment 2, we demonstrated the feasibility of the procedure in the intended age group. The EmpaToM-Y will be a valuable tool in future research and help to close the gap of knowledge on social cognition between childhood and adulthood.

The valence manipulation of the EmpaToM-Y reliably induced empathic affect sharing in both age groups. Participants indicated to feel significantly more negative after videos with negative valence and experiment 1 revealed a high correlation between affect ratings in our new task and an established measure of empathic affect sharing.

In experiment 1, we found significant correlations of both ToM and nonToM questions between the EmpaToM-Y and an established and thoroughly validated ToM task. While a direct and systematic comparison between experiments 1 and 2 is precarious due to the strong heterogeneity in context variables, there are some noticeable differences that could inspire further research. Our first experiment indicated that both ToM and nonToM questions of our new task were equally demanding for adults. However, adolescents in experiment 2 seemed to find ToM questions more difficult, which is in line with the finding that social cognition is not yet fully emerged in late childhood but instead continues to develop until early adulthood (Blakemore, 2008; Blakemore et al., 2007; Sebastian et al., 2012;

Symeonidou et al., 2020; Wang et al., 2006). Note that the very same questions were posed in both experiments of this study, which indicates similar difficulty of ToM and nonToM questions when ToM is fully developed. Our results suggest that this was true already for a small proportion of the adolescent sample, making the EmpaToM-Y a valuable tool to assess the developmental status of ToM beyond childhood. Future studies should apply the present task to larger and representative participant samples to gain a more comprehensive understanding of social affect and social cognition development in adolescents.

While social cognition is under-investigated even in the healthy teenage population, the research demands in adolescents with mental disorders are even higher. In some disorders, including schizophrenia (Bourgou et al., 2016; Li et al., 2017), ASD and Asperger's syndrome (Kaland et al., 2008), ToM has been investigated more thoroughly even in adolescent samples. For other conditions, such as social anxiety disorder (Öztürk et al., 2020), conduct disorder (Arango Tobón et al., 2017), personality disorders (Sharp et al., 2011) or bipolar disorder (Schenkel et al., 2008), the evidence is still very limited and more research is urgently needed. In all cases, however, systematic investigation of the relationship between social cognition and disease onset and progression are missing. Especially in combination with the EmpaToM (Kanske et al., 2015b), the EmpaToM-Y constitutes a promising basis for longitudinal studies assessing empathic affect sharing, ToM and their interplay – an opportunity that, to the best of our knowledge, is given by no other task to the present date. Due to known differences between the sexes in brain development and the incidence of many mental disorders, the role of sex in adolescent social cognition should receive special attention in future research.

In conclusion, we introduce a promising novel task for the assessment of empathic affect sharing and ToM as well as their interaction in adolescents. With its naturalistic setting, the EmpaToM-Y provides the opportunity of capturing inherently interactive capacities in their complexity and studying social understanding in a more realistic and ecologically valid setting. The short implementation duration and stimulating character make the EmpaToM-Y a measure that is particularly suitable for the assessment of social cognition in teenagers. Future studies could use this task to investigate inter-individual variability of socio-cognitive and socio-affective capacities as well as their precursors and outcomes in healthy minors and clinical populations. A first application of the novel task in a healthy sample adds evidence to the notion of an ongoing development of ToM throughout adolescence and a wide range of inter-individual differences in social cognition. This is important groundwork towards a more sophisticated understanding of the developmental trajectory of empathy and ToM beyond childhood and an important extension to our knowledge of social cognition across the lifespan.

Acknowledgements

This work was supported by the Emmy Noether Programme of the German Research Foundation (grant number BO 4962/1-1). We thank Caroline Beyer for her help with data acquisition.