The Ascription of Intentions in Gaze-Contingent Social Encounters

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Mathis Jording

aus

Köln, Deutschland

Dezember 2020

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der

Universität zu Köln im Mai 2021 angenommen

Acknowledgements

A number of people were essential to the completion of this thesis and to them I owe my deepest gratitude. First and foremost, I wish to thank my supervisors Ralph Weidner and Kai Vogeley for giving me the opportunity to pursue this project. I am very thankful for their intellectual input and the opportunity to profit by their experience and research insight. Special thanks go to Kai Vogeley as the supervisor of my group. Without his warmth, trust, and unwavering support, I would not have had the confidence and perserverance to conclude this thesis. He gave me the opportunity and financial freedom but also the confidence to present and discuss my ideas on conferences around the world.

This thesis would also not have been possible, if it were not for the extremely fruitful environment, my colleagues created and in which I was allowed to work in. These colleagues not only let me have part in their brilliant ideas, but also gave me strength when encountering obstacles and setbacks. However, most importantly, they made those last years a very enjoyable time in my life, despite the occasional hardships. Therefore, I would like to thank Ralf Tepest, Julia Proft, Astrid Gawronski, Katharina Krämer, Sebastian Lammers, Carola Bloch, Juliane Zimmermann, Julia Espelöer, Theresa Schoofs and Daniel Roth. Of outstanding relevance were Sevim Köroglu and David Vogel who stood by my side all the time, supported me and cheered me up on a weekly basis.

I also want to thank my parents for believing in me and supporting my decisions. They were the ones who sparked my interest in so many things and who taught me to stay curious. I also want to thank my brother Lennart, my cousin Judith and my other friends for their love, support and all the fun we had together. Discussions with them were the most heated and challenging but also the most fun, informative and exciting ones in my life. However, the most important influence my friends had on me was by grounding me and reminding me that there are so much more important things in this world than even science. Lastly and most importantly, I am deeply thankful to my girlfriend Anne, for her love, support and especially her patience and all the help and encouragement over all those years.

Content

List of Figures

List of Abbreviations

Abbrevaition	**Meaning**
AoI	Area of interest
ASD	Autism spectrum disorder
COM	‘Communicative’ condition
IJA	Initiating joint attention
INT	Introspective gaze state
JA	Joint attention
LOOK	‘Looking’ condition
OO	Object-oriented gaze state
PO	Partner-oriented gaze state
PRIV	‘Private’ condition
RJA	Responding joint attention
SGS	Social gaze space
VC	Virtual character

List of Studies
(Thematic ordering)

Study 1: (published) **Jording, M.,** Engemann, D., Eckert, H., Bente, G., & Vogeley, K. (2019). Distinguishing Social from Private Intentions Through the Passive Observation of Gaze Cues. Frontiers in Human Neuroscience, 13, 442.

Contribution: MJ contributed to the conception of the work and the statistical analyses, drafted the manuscript and implemented the comments of the co-authors and reviewers in the final version of the manuscript.

Study 2: (published) **Jording, M.,** Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2018). The “Social Gaze Space”: A Taxonomy for Gaze-Based Communication in Triadic Interactions. Frontiers in Psychology, 9.

Contribution: MJ contributed to the conception of the work, drafted the manuscript and implemented the comments of the co-authors and reviewers in the final version of the manuscript.

Study 3: (submitted) Hartz, A., Guth, B., **Jording, M.,** Vogeley, K., & Schulte-Rüther, M. (submitted). TriPy: An Artificial Social Agent for Systematic Investigations of Gaze-Based Communication in Triadic Interactions.

Contribution: MJ contributed to the conception of the algorithm, collected the data, assisted in analyzing the data, in implementing results as parameters in the algorithm, and drafting the manuscript, and revised the manuscript critically.

Study 4: (published) **Jording, M.,** Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2019). Inferring Interactivity From Gaze Patterns During Triadic Person-Object-Agent Interactions. Frontiers in Psychology, 10, 1913.

Contribution: MJ contributed to the conception of the work and the design of the study protocol, collected and analyzed the data, drafted the manuscript and im-plemented the comments of the co-authors and reviewers in the final version of the manuscript.

Study 5: (submitted) **Jording, M.,** Hartz, A., Vogel, D., Schulte-Rüther, M., & Vogeley, K. (submitted). Reduced Experience of Interactivity During Gaze-Based Interactions in Participants with Autism Spectrum Disorder.

Contribution: MJ contributed to the conception of the work and the design of the study protocol, collected and analyzed the data, drafted the manuscript and implemented the comments of the co-authors and reviewers in the final version of the manuscript.

Study 6: (in press) Roth, D., **Jording, M.,** Schmee, T., Kullmann, P., Navab, N., & Vogeley, K. (2020). Towards Computer Aided Diagnosis of Autism Spectrum Disorder Using Virtual Environments. 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 115–122.

Contribution: MJ contributed to the conception of the work and the design of the study protocol, collected the data, assisted in analyzing the data and drafting the manuscript, and revised the manuscript critically.

1 Introduction

While navigating the social world, we are guided by the eyes of the people we encounter. They act as points of interest, instantly and persistently attracting our attention and informing us about inner experiences and intentions of others (Heider, 1958; Yarbus, 1967). In general, nonverbal cues including body movement, postures, gestures, and facial expressions in addition to gaze behavior, do not only supplement verbal utterances, but also constitute a fundamental part of communication as a whole. Nonverbal communication is considered to convey up to 65% of social meaning in human communication (Burgoon, 1994) and to be crucial for the development not only during the individual's childhood but also of human culture and society (Tomasello et al., 2005; Vogeley & Roepstorff, 2009). Of all the nonverbal communicative channels, gaze constitutes one of the most effective and most frequently relied upon ones (Argyle & Cook, 1976; Emery, 2000). We can learn from the gaze of others about our shared environment and thus profit from their experiences and perspectives, figuratively and literally. Observing the gaze of others also helps us understand their motives and intentions and thus predict their future behavior. It offers insights into relationships between members of our social group and is important when trying to understand group dynamics. However, we also use our own eyes to communicate with others. We try to "catch" their attention by looking at them and use gaze to tell them about our view of things. We might be able to establish a relationship with them or to persuade them to help us with our own goals. In short, gaze helps us to become a member of a social group – provided we 'understand' each other.

However, correctly interpreting the gaze of other persons can actually be quite tricky for different reasons. First, eye movements, compared to other human movements, are quite circumscribed: While a change in the visual angle of a few degree can result in the perception of a whole new scene for the gazer, from an onlooker's perspective, a movement of just a few millimeters is visible. Fascinatingly, humans are highly adapted to solve this more technical problem and do so remarkably well (Emery, 2000; Gibson & Pick, 1963). However, another conceptual issue arises for gaze as a communication channel. When another person smiles at us, we might not understand what the person is trying to tell us, but at least we immediately know that the other person is trying to communicate with us (Heider, 1958). The eyes, however, are not only an instrument of communication but first and foremost one of perception. A person whose eyes we observe might be trying to communicate with us or just be interested in their surroundings. In fact, the person might not even have noticed us. Thus, when observing others, we primarily have to figure out whether their gaze behavior is meant as signal to us before determining what the other person is trying to tell us. For persons with difficulties in understanding and interpreting the intentions of others, as is the case in autism spectrum disorder (ASD), this situation can be even more challenging.

This thesis addressed the problem of how we are able to interpret and understand the gaze behavior of others in order to communicate successfully. As a first step, we examined which kinds of inferrences regarding the general intentions of another person are possible from the passive observation of gaze behavior (**study 1**). Turning towards gaze interactions, we then outlined a theoretical concept and a taxonomy of social gaze, allowing for holistic considerations of ongoing gaze encounters (**study 2**). The practical implementation of this approach in form of the new agent interaction platform TriPy was presented in **study 3**. Subsequently, an investigation of the inference of communicative intentions in ongoing gaze interactions was performed **(study 4**). **Study 5** then compared the performance of healthy participants with that of persons with ASD, finding that the latter especially have trouble in interactive situations. As an outlook to the future direction of social gaze investigations and their application in clinical contexts, **study 6** introduced a new technical system for avatar-mediated communication between two persons combined with machine learning based data analysis.

Before these studies are presented in more detail, a theoretical background is provided by introducing central phenomena of social gaze as well as describing impairments in gaze communication in ASD. In addition, methodological requirements and challenges in the investigation of social gaze are elucidated. After presenting the individual studies, the results obtained in this thesis are integrated in a general discussion focusing on our understanding of social gaze, its clinical implications, as well as potential future directions in social gaze research.

2 Theoretical Background

2.1 The Eyes as a Source of Information

In humans, gaze constitutes an important channel for learning from and interacting with each other. The pure morphology of the human eye is already believed to point to the evolutionary importance of gaze communication. The high visibility and contrast due to the white sclera is unique to human eyes (Kobayashi & Kohshima, 1997, 2001), suggesting an adaption to the demands of understanding and being understood by others when living in groups (Emery, 2000).

Starting in early infancy (Haith et al., 1977) the eyes are among the earliest and most frequently attended features of faces (Walker-Smith et al., 1977; Yarbus, 1967). From the eye region we infer gender, age, and personality of a person (George & Conty, 2008; Itier & Batty, 2009) as well as attentional and emotional states (Baron-Cohen et al., 1997; Emery, 2000). The informational density of eyes can even lead to avoiding looking at the eyes in order to lower cognitive demands (Doherty-Sneddon & Phelps, 2005; Glenberg et al., 1998; Markson & Paterson, 2009; Phelps et al., 2006). In addition to information about inner states of the other person, we also learn about the person's relationship and intentions towards objects in their, respectively our shared environment. Starting from the age of 8 month, infants seem to expect the gaze of others to be directed at specific objects (Csibra & Volein, 2008). From the age of two, children can infer desires of other persons by observing their gaze behavior (Lee et al., 1998). Even older children also take into account how long a person looks at objects when trying to assess the person's preferences (Einav & Hood, 2006; Montgomery et al., 1998). It seems that from the mere observation of someone looking at an object we anticipate their behavior towards the object and when their view is distracted, the kinematics of our own motor actions toward the object change as well (Castiello, 2003). In addition to our behavior, observing someone else looking at objects also changes our attitudes towards these objects, making them appear more familiar (Reid et al., 2004; Reid & Striano, 2005) and likeable (Bayliss et al., 2006; Landes et al., 2016; Ulloa et al., 2015).

Gaze processing, i.e. the process, in which we recognize the gaze direction of another person and deduce from it the probable focus of attention, seems to take place automatically and involuntary. When seeing a face with averted gaze the attention of an onlooker reflexively shifts towards the point that is focused by the observed eyes (Friesen & Kingstone, 1998). This effect is referred to as gaze cueing (cf. Frischen et al., 2007) and it can override effects of higher psychophysical saliency (Borji et al., 2014). Gaze cueing can be observed, even when the cue is actually uninformative or counterpredictive (Bayliss & Tipper, 2006; Driver et al., 1999; Friesen & Kingstone, 1998). In addition to the attentional shift, onlookers often follow the observed gaze with their own eyes (i.e. "gaze following"), resulting in joint attention (JA), the situation in which both persons look towards the same object (Emery, 2000). Gaze following seems to be of high importance to the development of social cognition in humans and is believed to be the basis for human cooperation (Tomasello et al., 2007). The ability to follow another person's gaze develops starts at 6 month of age (Senju & Csibra, 2008) and predicts the later development of a theory-of-mind and language (Charman et al., 2000; Morales et al., 1998). It also seems to be connected to intelligence, depth of information processing and self-regulation (Mundy & Newell, 2007) and to be a prerequisite of reinforcement learning (Vernetti et al., 2017).

2.2 Direct Gaze and Eye Contact

Of specific relevance to onlookers are eyes directed at them. They are detected faster than averted gaze (Conty et al., 2006; Senju et al., 2008; von Griinau & Anston, 1995), even by newborns (Farroni et al., 2002), and automatically capture attention (Dalmaso et al., 2017; Senju & Hasegawa, 2005). In accordance, it seems that the processing of direct and averted gaze is instantiated by at least partially distinct neural pathways (Gale et al., 1975; Hietanen et al., 2008; Senju et al., 2005). The differentiation of these pathways has been observed as early as 160ms after stimulus presentation (Conty et al., 2007). This suggests a distinction between eyes directed at or averted from the observer on the bases of low-level stimuli features.

However, in addition to the physical properties of eye stimuli, it seems that the social dimension distinguishes direct from averted gaze. Eyes directed at oneself can convey the feeling of being observed and can enhance compliance with social norms (Bateson et al., 2006, 2013; but also see Carbon & Hesslinger, 2011) or in other cases might be experienced as akward or threatening (Ellsworth et al., 1972). However, the effects of direct gaze seem to be modulated by ones believe to be facing a real human. Early-stage processing of facial features is enhanced for real and dynamic compared to static stimuli (Pönkänen et al., 2011a, 2011b). Similarly, seing eyes directed at oneself is associated with a stronger increase in the cognitive demand (Doherty-Sneddon & Phelps, 2005; Markson & Paterson, 2009) and physiological arousal (Hietanen et al., 2008; Pönkänen et al., 2011a, 2011b) in face-to-face communication compared to static or video stimuli.

Interestingly, in situations in which we believe we are being observed, we also control our gaze behavior for social adequacy (Risko & Kingstone, 2011; Wu et al., 2013). This leads to a unique feature of social gaze within nonverbal communication. Contrary to other nonverbal communication channels, communication via gaze is bidirectional and the eyes serve simultaneously as input and output device. While we perceive the environment through our eyes, others can deduce our focus of attention from observing our eyes. In our daily encounters with others, we are aware of and incorporate their ability to follow our gaze by controlling and actively using our gaze as a signaling device. Gaze is therefore said to serve a 'dual function' for humans (Gobel et al., 2015; Jarick & Kingstone, 2015). Against this background, eye contact constitutes a very powerful and complex social signal. During eye contact, both interactants are aware of the existence of the bidirectionality and know of or rather expect each other's awareness. Accordingly, eye contact is often associated with the intention to initiate interactions and to communicate (Cary, 1978; Emery, 2000; Kleinke, 1986). Eye-contact was found to enhance the chances for compliance with a request (Guéguen & Jacob, 2002), to convey intimacy and to moderate interpersonal distance (Argyle & Dean, 1965). Furthermore, establishing eye contact with another person seems to change how we think about and interact with this person. The ”eye contact effect“ (Senju & Johnson, 2009a) entails the improvment of person recognition (Hood et al., 2003), person memory (Mason et al., 2004), gender discrimination (Macrae et al., 2002), and emotional-empathy (Schulte-Rüther et al., 2007). Depending on the duration of gaze towards the observer, the person might also be experienced as more potent (Brooks et al., 1986), having more self-esteem (Droney & Brooks, 1993), or appears as generally more attractive (Mason et al., 2005), likeable (Argyle et al., 1974; Mason et al., 2005) or approachable (Stass & Willis, 1967).

Eye contact was also described as the starting point and a central element of the coordination of social interactions in different contexts (Argyle & Cook, 1976; Emery, 2000). However, the exact role of eye contact and whether it is a prerequisite or a cause of the emergence of gaze interactions is unclear as of yet. Therefore, it is essential to better understand the experience of interactants during eye contact and the relationship between the occurance of eye contact and the ascription of intentions.

2.3 Gaze Coordination of Social Interactions

Humans use gaze to coordinate social interactions in dyadic as well as triadic encounters. In dyadic encounters, i.e. an interaction between two persons, gaze playes a pivotal role as a supplement to speech. During conversations, interlocutors convey their understanding of the current allocation of speaker and listener roles through their gaze behavior (Argyle et al., 1974; Argyle & Cook, 1976): While the listener looks at the speaker for most of the time in order to express attentive listening, the speaker frequently avoids eye contact and directs their gaze away from the listener. This behavior seems to serve two purposes, avoiding cognitive demands from eye contact, but also as a signal to the conversation partner. By avoiding eye contact, the speaker informs the partner that she has not yet concluded expressing their thoughts and that she intends to keep on talking. This way the speaker can avoid being interrupted even during short pauses. Only as soon as the speaker looks back at the partner, she is expecting a replay or utterance by the partner.

In triadic encounters, i.e. situations constituted by two interactants and one or several additional objects (cf. Lee et al., 1998), gaze predominantly serves the purpose of coordinating the allocation of attention. As discussed earlier, the term JA denotes the situations in which one interactant followed the other interactants' gaze towards an object with the effect of both converging on the same object (Emery, 2000). Emery further distinguishes between JA and 'shared attention' based on the interactants awareness of the situation. Shared attention, compared to JA, requires both participants to be aware of their joint focus on the object. For the sake of clarity and consistency, shared attention will not be distinguished from JA in the following, since both terms are often used interchangibly in the literature (cf. Pfeiffer et al., 2012). Furthermore, the identification of shared attention would require surveys of both interactants' assumptions of the situation. Since many experimental setups do not allow and studies do not report such surveys, this requirement is not met for the majority of the existing literature. Instead, the related but behaviorally defined and more widely applied distinction between responding to joint attention (RJA) and initiating joint attention (IJA) will be adopted here. RJA denotes the process already described, in which one of the partners follows the other one towards an object, whereas, IJA means that one of the interactants deliberately leads their partners' gaze towards an object.

Despite the behavioral definition of RJA and IJA, both processes also require different levels of knowledge of the situation and the partner. Therefore, they allow inferences about the mentalizing capacities of the actors. RJA is based soley on gaze following and only requires the following person to understand the other person's gaze behavior as being goal-directed. In theory, learning that following another person's gaze potentially reveals valuable information about the environment suffices to engage in RJA. It does not require any acknowlegdement of the partners' cognitive capacities or intentionality. It is therefore not surprising, that human infants from the age of 6 months start following another person's gaze (Morales et al., 2000), especially if encouraged by additional communicative cues (Senju & Csibra, 2008). Gaze following also does not seem to be unique to humans but was observed in chimpanzees as well (Tomasello & Carpenter, 2005).

IJA on the otherhand requires more elaborate levels of cognitive processing. The initiation of JA requires understanding, that the own gaze is visible to others and informs them about ones own focus of attention (Gobel et al., 2015; Jarick & Kingstone, 2015). Secondly, the establishment of JA has to be identified as a crucial step in the initiation of a cooperation (Tomasello et al., 2005). Finally, cooperation with others has to be recognized as a potentially profitable or rewarding endeavour. Consequently, chimpanzees that successfully engaged in gaze following, did not show signs of IJA (Tomasello & Carpenter, 2005). In humans, IJA also seem to develop later in life compared to RJA, with first reports stemming from the second year of life (Mundy et al., 2007; Mundy & Newell, 2007). Based on these empirical findings and differences in cognitive requirements, it has been hyopthesized that RJA and IJA are realized through different cognitive mechanisms (Mundy & Newell, 2007). Interestingly, the IJA compared to RJA was found to more strongly activate reward related circutries (Oberwelland et al., 2016; Pfeiffer et al., 2014; Schilbach et al., 2010). Humans seem to expect others to follow their gaze (Pfeiffer et al., 2011) but they also react with increased interest to people that had followed their gaze previously (Bayliss et al., 2012).

Taken together, it becomes clear that the coordination of attention via gaze is by no means a trivial operation, requiring many different skills, but that it is crucial not only to the socio-cognitive development of the individual but also to the cooperation between humans in general. Aim of this thesis is to elucidate the individual steps, the course and the underlying mechanisms of the emergence of gaze interactions. Furthermore, it is important to understand, to which extent interactants develop an awareness of this situation. This includes the question whether this process is cognitively penetrable or whether interactants predominantly act based on intuitions. This knowledge is of paramount importance when trying to understand impairments in gaze communication as in the example of persons with autism, discussed in the following.

2.4 Impairments in Gaze Interactions in Autism

ASD[1] describes a pervasive neurodevelopmental disorder with deficits in communication, social interactions or restricted patterns of behavior and interests (American Psychiatric Association, 2013). One key feature are impairments in the socio-cognitive domain while non-social cognitive faculties are often preserved (Klin, 2006). These impairments are especially prevalent in nonverbal communication, including the detection (Dratsch et al., 2013; Senju et al., 2005) and interpretation (Baron-Cohen et al., 1997; Uljarevic & Hamilton, 2013) of nonverbal cues. The impairments in social interactions and the difficulties in inferring mental states from the eye region in ASD (Baron-Cohen et al., 1997, 2001), make this group especially interesting to social gaze research. Comparisons with healthy participants can foster the understanding of the fundamentals of social gaze while at the same time advancing knowledge about the background and potential therapies of ASD.

Persons with ASD have often been reported to show less visual attention towards socially relevant features of a scene. Autistic teenagers and adolescents compared to same aged control participants were found to focus less on faces than other objects (Bird et al., 2011; Riby & Hancock, 2009a, 2009b; Wilson et al., 2010). Additionally, persons with ASD have been reported to show pronounced fixation of the mouth region and diminished fixation of the eye region (Klin et al., 2002a, 2002b, 2003; Neumann et al., 2006). These findings promoted the assumption that ASD might be characterized by a lack of interest and motivation for social encounters (Chevallier et al., 2012).

Interestingly, the picture of attention for social stimuli in ASD has become more nuanced recently. There is a growing consensus that the social context of the situation under investigation, with regard to complexity, dynamics, and interactional affordances can have a strong mediating influence (Chita-Tegmark, 2016; Falck-Ytter & von Hofsten, 2011; Frazier et al., 2017; Guillon et al., 2014). Differences between autistic participants and control participants are especially apparent in paradigms using dynamic (Bird et al., 2011; Riby & Hancock, 2009a; Speer et al., 2007) or socially complex stimuli (Riby & Hancock, 2008; Wilson et al., 2010), while static or less complex stimuli often do not elicit equivalent effects (Fletcher-Watson et al., 2009; Freeth et al., 2010). Persons with ASD have also been shown to avoid looking at another person more

[1] This thesis uses the term ‚Autism Spectrum Disorder‘ since it is currently the official clinical term (American Psychiatric Association, 2013) and according to a recent survey in the UK there is no other english term for which there is a wider agreement in the autistic community (Kenny et al., 2016). Similary, persons with ASD are refered to as ‚autistic persons‘ as it is by far the most accepted term in this community (Kenny et al., 2016). Participants without an ASD diagnosis with whom autistic participants are compared in experimental studies are refered to as ‚non-autistic persons‘, ‚control participants‘, or ‚control group‘.

strongly when the person looks back at them and thus the situation entails the potential for an interaction (Freeth & Bugembe, 2019). Interestingly in children with ASD even the social content of a conversation might be relevant to their gaze behavior with seemingly more pronounced avoidance of eye contact when talking about ‘how people feel’ compared to ‘what people do’ (Hutchins & Brien, 2016).

Despite differences in the allocation of attention in social contexts in ASD, there is evidence for impairments in other domains of social gaze as well. However, again, the results are partially contradicting and the picture is less clear than generaly assumed. Persons with ASD have been reported to be less sensitive to subtle variations in the gaze behavior of others (Georgescu et al., 2013) and do not exibit the ‘eye contact effect’ mentioned earlier (Senju & Johnson, 2009b). However, whether gaze direction processing is generally impaired in ASD is unclear. Some studies have found evidence for impairments in the ability to recognize gaze directions of another person (Ashwin et al., 2009; Howard et al., 2000). In another study adolescents and adults with ASD did not differ from control participants in their ability to detect gaze direction changes (Fletcher-Watson et al., 2008). Children with ASD were also able to recognize the gaze direction of another person at least when using their fingers to trace the observed person’s line of sight (Leekam et al., 1997). Gaze cueing seems to be generally intact in person with ASD at the age of ten (Kylliainen & Hietanen, 2004; Senju et al., 2004), but see Frischen et al. (2007) for a more detailed analysis. When observing gaze shifts in others, autistic children exhibit reflexive gaze following comparable to not affected children (Chawarska et al., 2003; Kylliainen & Hietanen, 2004), thus they are able to respond to JA. However, autistic children seem to be less inclined to initate JA themselves (MacDonald et al., 2006; Mundy, 2003; Mundy et al., 1994; Whalen & Schreibman, 2003).

In summary, it seems that a broad range of skills related to the processing of social gaze and the behavior in gaze interactions can be affected in ASD. However, these impairments do not seem to be general and comprehensive but appear to be individual, age-, and context-dependent instead. In addition, despite the wealth of research about gaze behavior in ASD (cf. Ames & Fletcher-Watson, 2010; Cañigueral & Hamilton, 2019; Falck-Ytter & von Hofsten, 2011; Guillon et al., 2014; Nation & Penny, 2008), many studies are restricted to mere visual attention or to the use of non-interactive paradigms. This thesis aims at elucidating, how gaze behavior of persons with ASD looks like in more interactive, complex situations with a special focus on alterations of the experience and understanding of these situations.

2.5 Theoretical and Technological Approaches Towards Studying Gaze Interactions

Traditionally, much of the knowledge in the field of social gaze is based on findings from studies investigating the distribution of visual attention in participants passively observing specific social stimuli. Early eyetracking studies impressively found social elements in different scenes that reliably attracted the gaze of onlookers (Yarbus, 1967). Although the importance of bidirectionality in gaze interactions was recognized right from the beginning (Argyle & Cook, 1976; Gibson & Pick, 1963), the state-of-the-art of eyetracking of that time made investigations of free gaze interactions almost impossible. Early eyetracking techniques were either not precise enough or strenuous and potentially even painful for participants (Yarbus, 1967). These methods were too cumbersome for the application in face-to-face interactions, especially when aiming at preferably naturalistic situations. Consequently, the field of social gaze research was predominated by static stimuli and explicit instructions of participants to observe the pictures while their gaze behavior was recorded. Occasionally, ongoing gaze interactions were studied by observing and manually recording the gaze behavior through coders seated behind the interactants and concealed by one-way mirrors (Argyle & Cook, 1976). However, these methods are inconvenient for their lack of spatial and temporal accuracy in data acquisition, experimental control and ecological validity.

Fueled by technological advancements, new methodological approaches and experimental designs became available. Modern eyetracking devices precisely measure gaze behavior without irritating participants and can easily be integrated in laboratory and field experiments or even neuroimaging studies (Pfeiffer et al., 2013b). An additional development that promoted social cognition research was the combination of virtual characters (VCs) and eyetracking devices (Georgescu et al., 2014; Vogeley & Bente, 2010). By creating computer algorithms that can process eyeposition data in real-time, it was possible to investigate ongoing social interactions: e.g. an algorithm by Wilms and colleagues (Schilbach et al., 2010; Wilms et al., 2010) processes the participant's gaze behavior and controls a VC accordingly, thus enabling the VC to react to the participant by following with its gaze or deliberately adverting its gaze. These paradigms are therefore referred to as gaze-contingent, since the agent's behavior is contingent upon the participant's behavior. With this platform is was possible to investigate dynamic elements of gaze interactions like the establishment of JA and its neural correlates (Pfeiffer et al., 2011, 2012, 2014; Schilbach et al., 2010). Thus for the first time it was possible to conduct controlled experiments with ongoing gaze interactions (cf. Pfeiffer et al., 2013b). However, these studies still did not allow for the complete unfolding of extended or unrestricted gaze interactions, as these paradigms were based on a course of discrete trials of a few seconds, i.e. treating gaze interactions as a series of isolated events instead of phenomena emerging over time. In addition, in many cases, participants were instructed explicitly as to what to do and agents behaved in a predictable, often monotonic fashion. However, social interactions are inherently dynamic and implicit in character, which has to be taken into account in the pursuit of higher ecological validity (Pfeiffer et al., 2013a; Risko et al., 2012, 2016; Schilbach, 2010).

A very first step towards this goal is the development of a more holistic understanding of gaze interactions. As apparent in the eye contact effect and explained by the dual function of social gaze, the mere presence of another person can already shape cognitive processes as well as behavior in others (see above). When two persons see each other's eyes, they will notice every change in their counterparts gaze and potentially react to it, as subtle as it may be. In this case, the whole encounter entails the possibility of reciprocal interferences and the emergence of complex interactions. Therefore, investigations of gaze communication should not be limited to short, deliberate bits of interaction, e.g. single attempts to establish JA as described in earlier taxonomies of social gaze (e.g. Emery, 2000). Instead, every situation in which two persons can see each other's eyes has to be treated as a gaze encounter worth investigating. This warrants a new unifying taxonomy of social gaze, encompassing all potential interactive situations and all facets of gaze interactions. Common understanding in the field can only be accomplished, if the taxonomy provides a precise and specific terminology and entails detailed characterisations of individual elements of gaze interactions.

This new understanding and approach towards social gaze has to be accompanied by technical innovations in the field of gaze-contingent agent platforms. Specifically, requirements from three areas have to be met by a human-agent interaction platform in order to allow investigations of naturalistic interactions. First, participants should be allowed to interact freely with their partner, with as few restrictions in space and time as possible. This entails that instructions are easily and intuitively understandable and do not require narrow specifications of the target behavior. Secondly, the agent has to be able to display a broad behavioral repertoire and react in a flexible manner to the participant without creating the impression of repetitive or deterministic behavior. The agent's behavior has to be empirically informed and appear naturalistic while also granting the researcher sufficient experimental control. Lastly, the system has to allow for detailed and comprehensive observations and measurements of the participant's behavior during the unfolding of the interaction. An interaction platform that fulfills these requirements would allow investigating the ascription of intentions in unrestricted and temporally extended gaze interactions. In addition, it could elucidate the complex dynamics of reciprocal interferences in gaze encounters and thus help us to retrace the emergence of gaze interactions. With our growing understanding of gaze interactions, the complexity of the interactions under investigation could then also be increased incrementally.

Ultimately, the obvious goal is to investigate gaze interactions between two real persons. This requires the development of an environment in which ongoing gaze interactions can be monitored and recorded in real-time without too many restrictions for the interactants. In addition, powerful analytical tools are warrented that can deal with the wealth of data that can be collected in these situations. A very promising candidate are machine-learning approaches that make use of increases in computational power and that are able to detect complex patterns in multi-dimensional data. In summary, the goal to investigate real two-person interactions is certainly very ambitious and will require more technical developments and extensive research and testing. However, its potential to deepen our understanding of gaze interactions and social cognition in general justifies the efforts.

3 Thesis' Research Agenda

The aim of this thesis was to improve our understanding of how humans infer and understand intentions of others from their gaze behavior. This was accomplished by subsequently increasing the complexity of the interactions under investigation while simultaneasously developing a new theoretical concept and a new technological system, incorporating recent results from empirical studies. This approach was applied for the field of mental disorders by investigating the behavior and experiences of persons with ASD during unrestricted gaze interactions.

The first step was an investigation of the principles of ascribing intentions from the passive observation of another person's gaze (**study 1**). The central question here was, whether observers can generate the impression of being able to infer intentional states of others by the mere passive observation of gaze behavior. This allowed the identification of gaze patterns that prompt observers to ascribe either private or social intentions to the observed person. Furthermore, the robustness of the effects gave some indication as to the certainty of the participants' judgements and thus to the potential reliance on these more basic gaze patterns in everyday live interactions. **Study 2** then introduced the "Social Gaze Space (SGS)" as a new taxonomy of social gaze. The SGS comprises and describes the most fundamental mental states during gaze interactions and allows disentangling the complex unfolding of gaze interactions into distinct elements and processes. It is thus perfectly suited to investigate unrestricted and temporally extended gaze interactions and provided the framework for the subsequent investigations. **Study 3** described the development of the new gaze-contingent human-agent interaction platform TriPy that can account for and reproduce the complex behaviors during gaze interactions. Data from a pilot study informed about the different parameters and characteristics of the different mental states of the SGS and provided high degrees of realism in the agent's behavior. **Study 4** then turned towards observing gaze interactions of longer durations and with higher degrees of behavioral freedom. It focused on whether participants are able to identify attempts to initiate interactions via gaze by another person. The second objective was to identify the strategies that allow participants to solve the task. In **study 5**, a group of persons with ASD was compared to a group of healthy subjects with regard to their experiences during these gaze interactions. Revealing more fundamental experiental aspects of ASD during social interactions extended the understanding of the disorder beyond mere behavioral descriptions. As an outlook to the future of social gaze research and applications, **study 6** introduced a system for real two-person interactions. It demonstrated the potential of a machine-learning based classification of avatar-mediated interactions in investigations into social cognition as well as clinical diagnosis.

3.1 Study 1: Distinguishing Social From Private Intentions Through the Passive Observation of Gaze Cues (**Jording, M.**, Engemann, D., Eckert, H., Bente, G., & Vogeley, K, 2019. Frontiers in Human Neuroscience, 13, 442.)

When trying to understand another person's inner experiences and predict their future actions, the gaze almost reflexively focuses on the eye region of the person. The prevalence of this behavior in situations of social uncertainty reflects the high informative value of cues provided by the eye region. However, the eyes are not primarily a communication device and cues glimpsed from them are ambiguos and context-dependent. This study investigated the relationship between a person's gaze cues and the intentional inferences they elicit in an observer. The study described these associations and assessed their consistency and reliability. From the perspective of the observer, the information, whether or not the actions of another person are likely to be directed towards the observer, are most crucial. Therefore, the study distinguished between private and communicative intentions based on whether they are directed at the observer or more specifically whether they are directed towards any kind of interaction with the observer.

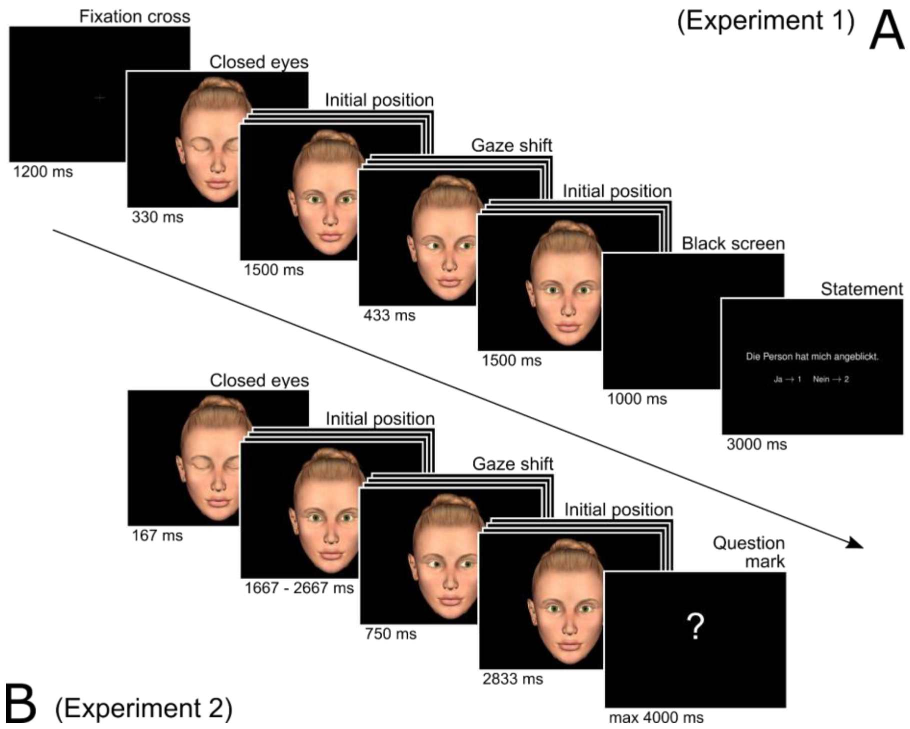


Figure 1. Trial courses of study 1, experiment 1 & 2 (A & B). The stack of images for initial position and shift position indicate that in each trial, one out of four possible images was displayed. Note that after the shift, the VC always returned to the same initial position it had started from. In this example, the initial position 1 (direct gaze) and the shift position 4 are depicted. The question mark indicated the prompt for participants to give their ratings. (Adapted from Jording et al., 2019)

In two experiments (Figure 1), participants repeatedly observed an agent whos gaze was initially directed at either the participant, thus establishing eye contact or who's initial gaze direction diverged from direct gaze by different degrees. The agent then shifted its gaze to one of the sides with different amplitudes. In experiment 1 (Figure 1A) participants were asked to either assess whether the agent was looking at them ("LOOK" condition) or whether the agent was trying to show them something ("COM", e.g. communicative condition). In experiment 2 (Figure 1B) participants had to assess whether the agent was interested in something ("PRIV", e.g. private condition) or again whether the agent was trying to show them something (COM condition). Experiment 1 was conducted as an online study with a large number of participants while experiment 2 was conducted in a lab with a lower number of participants but under control of the experimental environment. Results suggested the impression of being looked at (Fig-

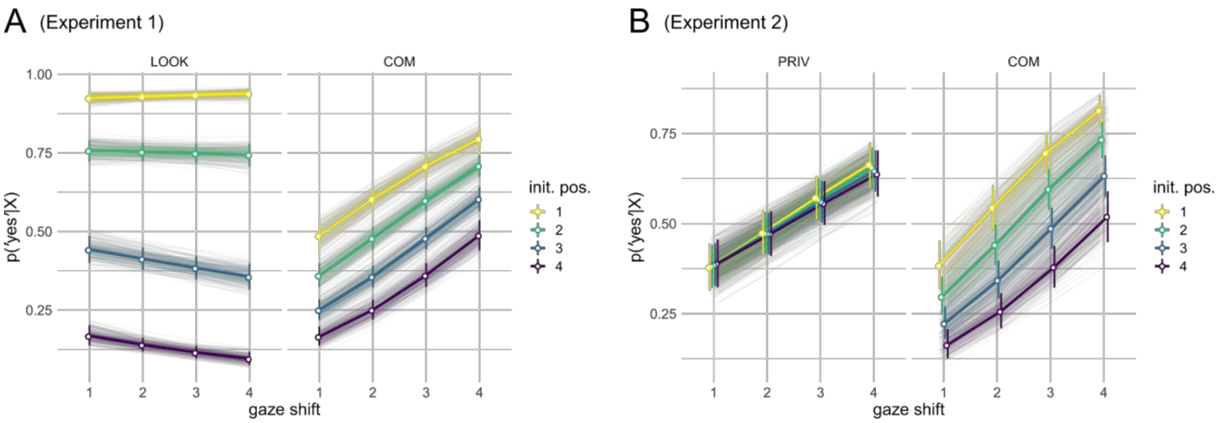


Figure 2. Influences of initial gaze position and gaze shift on the ascription of intentions in experiment 1 (A) and experiment 2 (B). Posterior predictions of the influence of initial position ("init. pos.") and gaze shift amplitude ("gaze shift") in the COM condition ("the person wanted to show me something") and the LOOK condition ("the person looked at me") in experiment 1 (A) and in the COM condition and the PRIV condition ("the person was interested in something") in experiment 2 (B). For the initial position, "1" corresponds to direct gaze and "4" to a maximally (vertically) averted position. For the shift amplitude, "1" corresponds to the smallest and "4" to the largest possible shifts. (Adapted from Jording et al., 2019)

ure 2A) to be almost exclusively determined by the degree of initial eye contact. Whereas the inference of private intentions (Figure 2B) depended mostly on the amplitude of the subsequent gaze shift, with larger shifts increasing the tendency to ascribe private intentions, the inference of communicative intentions (Figure 2A&B) depended on both aspects of the gaze behavior and was highest for a combination of initial eye contact and large subsequent gaze shifts. All effects were comparably decisive, implying that participants were used to rely on simple gaze despite their inherent ambiguity when ascribing intentions. Results from the LOOK condition corresponded to earlier studies, finding a high sensibility to eyes directed towards the observer (Senju & Johnson, 2009a; von Griinau & Anston, 1995). While counterintuitively the likelihood of ascribing private and communicative intentions both increased with the amplitude of the subsequent gaze shift, participants distinguished the two on basis of the initial eye contact. Encounters that were initiated with eye contact were more often experienced as communicative, i.e. as a situation in which the agent was trying to show something to the participant. This supported accounts claiming a facilitating role of initial eye contact as 'ostensive' cue in situations in which one person is trying to teach something to another person (Csibra & Gergely, 2009, 2011).

3.2 Study 2: The "Social Gaze Space": A Taxonomy for Gaze-Based Communication in Triadic Interactions (**Jording, M.**, Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K., 2018. Frontiers in Psychology, 9.)

Despite the dynamic and complex character of gaze communication, much of the previous social gaze research was actually based on quite restricted und circumscribed situations and encounters. As demonstrated in study 1, this does not necessarily diminish the informative value of these studies. However, it becomes increasingly apparent that these approaches tend to overlook vital aspects of communication.

The most influential accounts of social gaze communication that shaped social gaze research (e.g. Emery, 2000) are mainly based on behavioral observations. They structure and understand gaze communication as a series of isolated elements identifiable and distinguishable purely along observable, behavioral criteria. Taxonomies derived from these accounts made this field accessible to systematic controlled experimental studies and opened it to the first neuroscientific investigations. However, the reductionistic approaches reach their limit when trying to understand the relationship and connection between the individual elements of gaze communication. This requires a more holistic perspective that understands gaze communication as a dynamic phenomenon that only emerges over time out of the interaction of two persons.

The SGS pursues this approach by focusing on intentional and experiential aspects of gaze communication. Based on theoretical considerations, it structures gaze communication by distinguishing the most basic internal states a person in triadic gaze interaction can adopt (Figure 3). **Study 2** comprehensively described these different states and systematically summarized empirical evidence showing the inherently interactive character of all of these situations: 1. In the partner-oriented state (PO) a person focusses solely on the partner. The eyes of the partner will automatically attract a person's eyes while being looked at will also have subtle effects on

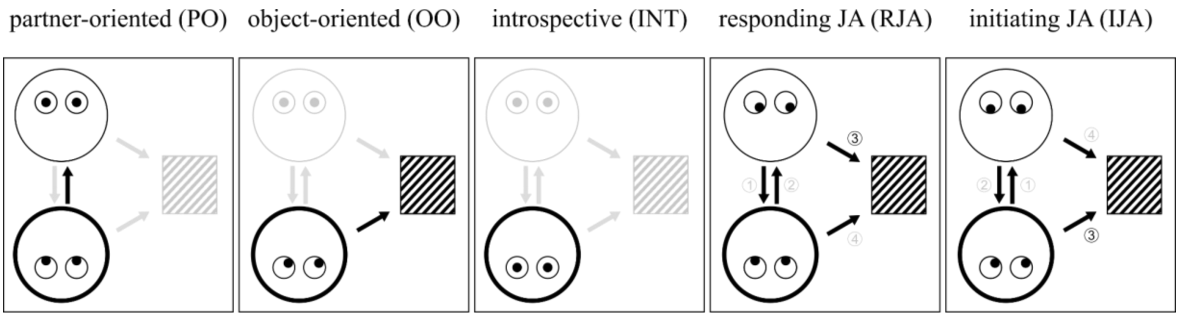


Figure 3. Illustration of the five interactional states of the SGS (illustration in alignment with Emery, 2000) from the perspective of index person A (always the bold face at the bottom) in interaction with person B. (1) Partner-oriented (PO): Attentional focus of A is directed toward B without deliberate attempts to interact of any of the two interactants. (2) Object-oriented (OO): Attentional focus of A is directed toward an object within the shared environment. (3) Introspective (INT): The attention is directed toward A's own inner experience. (4) Responding JA (RJA): A follows B's gaze toward an object. (5) Initiating JA (IJA): A tries to shift B's attention toward an object. (From Jording et al., 2018)

the partner herself; 2. In the object-oriented state (OO), the attention is focused on some objects in the environment and not directly on the partner. However, even in these situations the interacting partners are not detached from each other and they will mutually influence each other's dealings with the environment, although potentially unintentionally and unconsciously; 3. In the introspective state (INT) the person focusses on inner or bodily experiences. Although the attentional focus here does not lie in the environment, INT is accompanied by subtle changes in occulomotiric behavior. Empirical evidence suggests that these behavioral changes indicate active attempts to disengage from the interaction with the partner; 4. In the RJA state the person is trying to interact with the partner by following their gaze. This behavior often occurs automatically and allows the person to learn from the partner about the shared environment; 5. In the IJA state the person is trying to lead the partner to some object in their shared environment. The initiation of gaze interactions is not only rewarding in itself, but is suggested to be a powerful skill for establishing cooperation and relationships. However, ontogenetic and phylogenetic evidence indicates that it also requires more advanced sociocognitive skills.

With the SGS **study 2** provided a comprehensive taxonomy and terminology to structure gaze interactions while accounting for their dynamic character. The different states described and categorized the behavior of each interaction partner and provided a framework for systematic investigations of the interplay between both partners as a two-dimensional state-space.

3.3 Study 3: TriPy: A Cross-Platform Toolbox for Systematic Investigation of Gaze-Based Communication in Triadic Interactions with a Virtual Agent (Hartz, A., Guth, B., **Jording, M.**, Vogeley, K., & Schulte-Rüther, M., submitted)

The combination of VCs and real-time processing of eye-tracking data allowes to generate gaze-contingent agent algorithms for the investigation of gaze interactions (Wilms et al., 2010). The combination of ecological validity and experimental control makes these setups a valuable tool for the growing field of social neuroscience (Pfeiffer et al., 2013b), especially since they can be applied in fMRI investigations (Pfeiffer et al., 2014; Schilbach et al., 2010) and patient groups (Georgescu et al., 2014). However, so far, application of most of these systems was very restricted due to various technical and conceptual reasons. Eye-trackers required the restraining of participants by using chin-rests or frequent recalibrations in order to ensure data quality. Furthermore, research focused on specific elements of gaze communication and tasks were tailored to recreate these. This required participants' instructions to be very explicit and agents to behave in deterministic fashion, which in turn resulted in linear and repetitive courses of interactions.

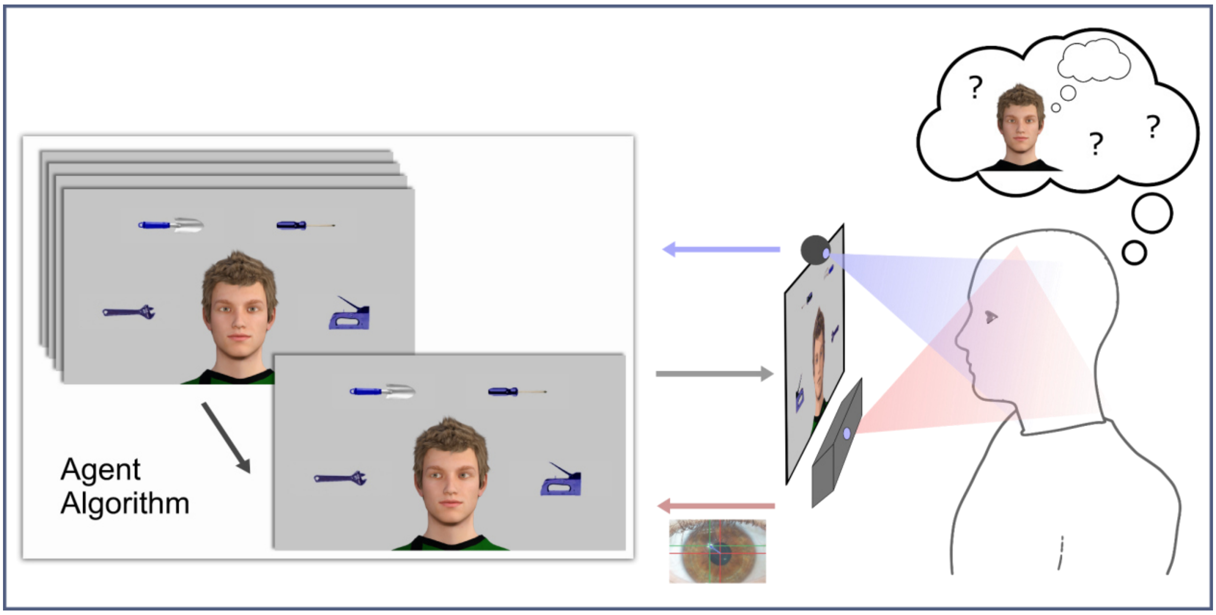


Figure 4. Technological architecture of TriPy: Eye tracker and video camera for data acquisition, computer screen for stimulus presentation, and computer for gaze- and emotion-contingent agent algorithm, systemintegration, and data collection. (From Hartz et al., submitted)

In **study 3** we developed TriPy (Figure 4) as a new system which allows for more naturalistic interactions and has more widespread applications. It provides high data quality without requiring to physically restrain participants with a chin-rest. Instead, participants can sit comfortably in front of the screen. The agent's behavior is based on probability matrices instead of relying on pre-determined series of actions. Thus, encounters are less predictable but emerge and develop spontaneously. However, the general behavioral pattern of the agent is governed by so-called macro-states over which the user (i.e. the experimenter) has full experimental control. Macro-states describe the overall distribution of the agent's visual attention as well as its responsiveness towards the participant. Already implemented in TriPy are the five states as specified in the SGS (**study 2**). The behavioral parameters corresponding to these states were empirically informed in a pilot study. Due to its generic structure, TriPy also allows to integrate additional states according to the needs of the individual research design. TriPy was developed independently of specific hardware, platforms or operating systems and allows the use of different eye-tracking systems. It also can integrate information from additional recording devices like cameras for the capturing of facial expressions. All data are recorded, logged and preprocessed automatically.

In summary, TriPy is a valuable tool for the investigation of social gaze in naturalistic interactions. It grants researchers high flexibility and experimental control while allowing for detailed observations and recordings in numerous settings.

3.4 Study 4: Inferring Interactivity From Gaze Patterns During Triadic Person-Object-Agent Interactions (**Jording, M.**, Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2019). Frontiers in Psychology, 10, 1913.)

As demonstrated in **study 1**, participants are adept in ascribing mental states based on the mere passive observation of gaze behavior. Communicative cues like eye contact and gaze shifts of certain amplitudes seem to inform these decisions. However, these cues are not biuniquely informative, i.e. depending on the situation, the same cues could be interpreted as signalling different mental states. Thus, so far, it is unclear whether participants infer mental states correctly and if so, which cues and techniques they use to accomplish this. In truly interactive gaze encounters, interactants have more information at their disposal and can use their own gaze for back channeling. Thus, participants potentially can resolve the ambiguity of gaze by directly communicating with their partner via gaze. **Study 4** investigated whether in these situations, participants can actually infer mental states correctly and whether their gaze behavior can reveal the strategies used.

In interactive gaze encounters with a partner, participants had to decide whether their partner was trying to interact with them. While participants believed their partner to be a real human (confederate), the VC displayed on the screen was in fact an agent controlled by TriPy (**study 3**; Figure 3). The agent would engage in pseudo-randomly chosen gaze states, entailing different attentional foci and responsiveness towards the participants' actions (**study 2**; Figure 3). Participants could freely interact with the agent for a maximum duration of 30 sec per trial in which they had to assess the agent's interactive intentions.

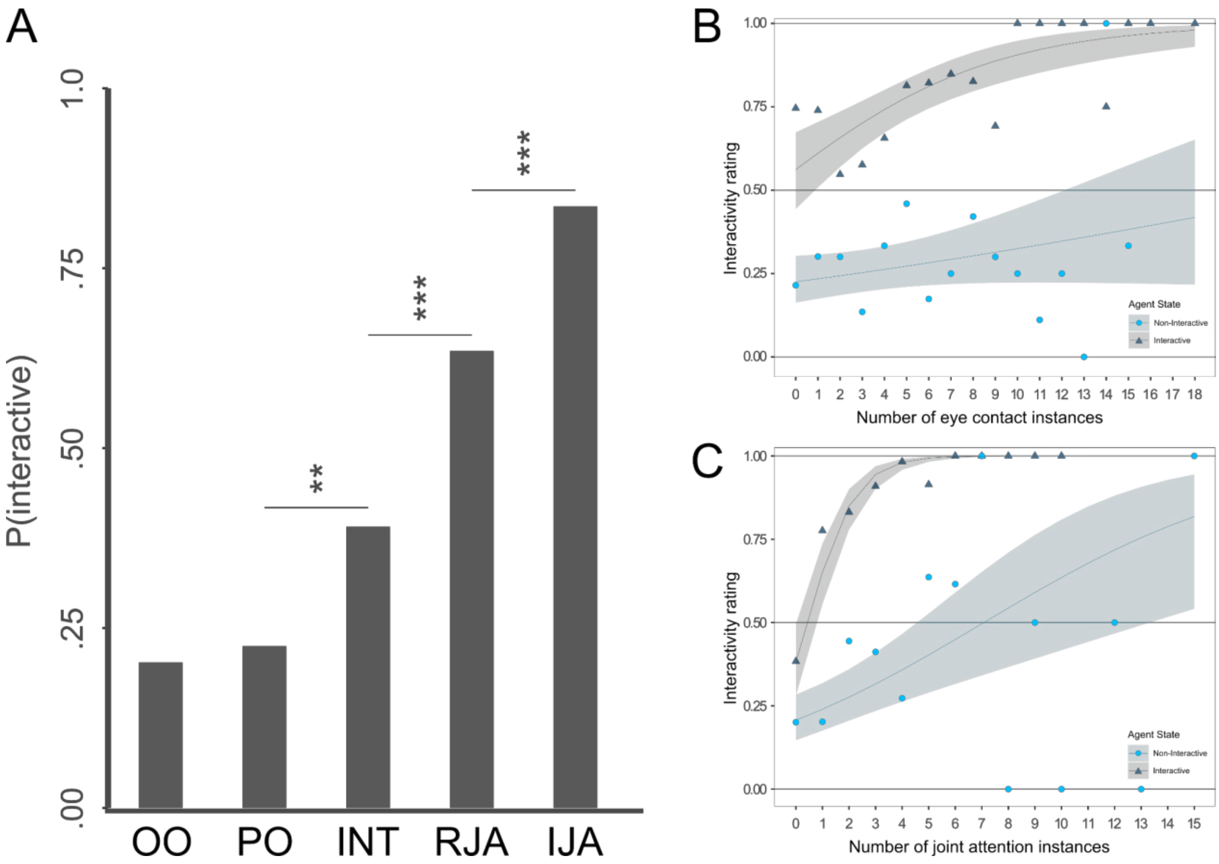


Figure 5. Interactivity ratings and joint focus in different gaze states. Plots of mean interactivity ratings separately for the different gaze states and as a function of the frequency of eye contact and joint attention. A: Mean interactivity ratings for different agent states. Asterisks indicate significant differences between neighboring states (when ranked in ascending order) in post-hoc tests (* <.05; ** <.01, *** <.001); B: Mean rates (circles and triangles) and model predictions with 95% confidence intervals (lines and bands) of interactivity ratings for differing numbers of eye contact instances per trial, separately for an agent behaving non-interactively (light blue) vs. interactively (dark blue); C: Mean rates (circles and triangles) and model predictions with 95% confidence intervals (lines and ribbons) of interactivity ratings for differing numbers of joint attention instances per trial, separately for an agent behaving non-interactively (light blue) vs. interactively (dark blue). (Adpated from Jording et al., 2019)

Results showed that participants overall clearly distinguished interactive and non-interactive states of the agent (Figure 5A). Different recognition rates within the interactive as well as the non-interactive states revealed that participants further differentiated states based on the distribution of the agent's visual attention. For the distinction between interactive and non-interactive states, a detailed analysis of the interaction between participant and agent hinted at an important role of eye contact and JA. Although it was not possible to prove a causal link in this explorative setup, data suggested that the occurrence of eye contact (Figure 5B) or JA (Figure 5C) was interpreted as a strong indicator for the agent's intentions. With increasing rates of eye contact or JA, the probability that the agent was rated interactive also increased. Interestingly, this effect was stronger in interactive trials when the agent was actually responsive as compared to unresponsive non-interactive trials. This showed that participants noticed whether they alone had caused instances of eye contact or JA or whether these were the result of a joint effort of both interactants. It was suspected that participants used the information as a basis for the decision whether the agent was actively trying to interact with them or not.

3.5 Study 5: Reduced Experience of Interactivity During Gaze-Based Interactions in Participants with Autism Spectrum Disorder (**Jording, M.**, Hartz, A., Vogel, D., Schulte-Rüther, M., & Vogeley, K., submitted)

Impairments of social cognition in persons with ASD extend to numerous different contexts and are considered generally pervasive. Reports about reduced visual attention for social stimuli have sparked the notion that persons with ASD are less motivated to engage socially and thus lack opportunities to develop their social skills. However, a growing number of recent studies draws a more complex picture of visual attention and social engagement in persons with ASD.

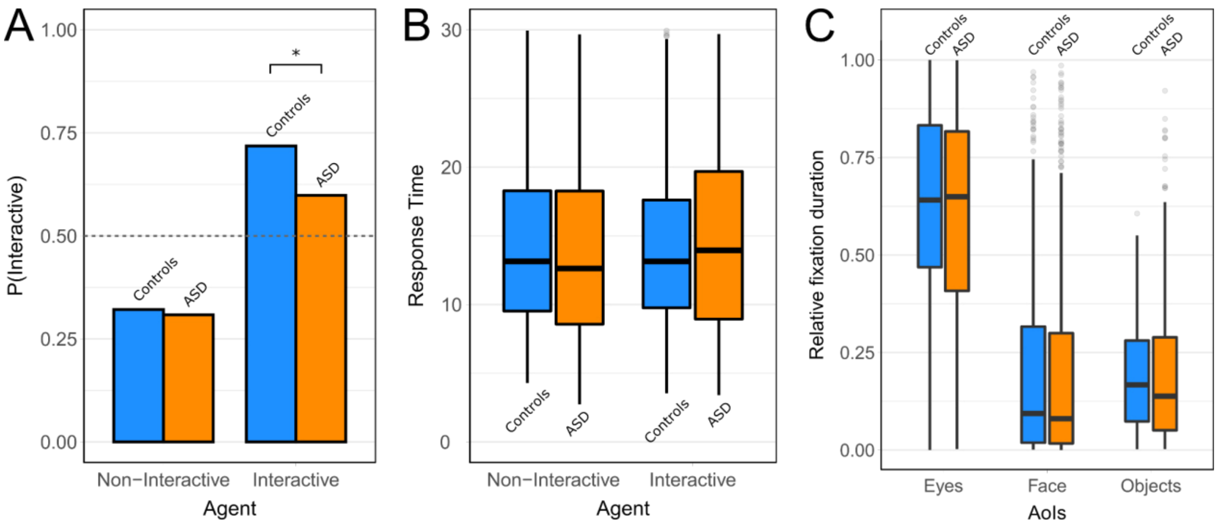


Figure 6. Interactivity ratings, response times and visual attention in diagnostic groups. Plots of mean interactivity ratings (A), mean response times (in ms, B) and the distribution of the participant´s visual attention (C) for diagnostic groups (blue: control participants / orange: ASD participants). A & B: non-interactive (left) vs. interactive (right) agent states, asterisks indicate significant post-hoc comparisons (* <.05; ** <.01, *** <.001), dashed line indicates the 50% guessing rate). ASD participants have significantly smaller detection rates for interactive states compared to control participants. Diagnostic groups do not differ with regard to non-interactive states or response times. C: relative fixation durations as the portion of time spent on the different AoIs (eyes, face, objects) per trial. Diagnostic groups do not differ significantly in their distribution of visual attention. (Adapted from Jording et al., submitted).

It seems that specifically interactive, dynamic or complex situations present challenges to persons with ASD while no differences can be observed in many more static experimental designs. Here, we systematically compared the establishment of a triadic gaze interaction with a gaze-contingent agent between healthy control participants and participants with ASD. The primary question was whether participants with ASD were able to recognize the intentions of their counterpart. Additionally, we were interested in attentional or behavioral differences between the groups that could potentially inform about the basics of interaction deficits in ASD. Therefore, we let a group of participants with ASD interact with the TriPy agent and infer its' intentions, in accordance with the design of **study 4**. This group was then compared to a group of healthy control subjects. We were interested in potential group differences in the ability to infer the partner's intentions but also in group differences concerning gaze behavior.

Results revealed impairments in the recognition of interactive agent states in persons with ASD (Figure 6A). Compared to the control group, these states were recognized significantly less often. No differences were observable in the response times (Figure 6B). Interestingly, the distribution of visual attention on the different areas of interest (AoIs) eyes, face, and objects did

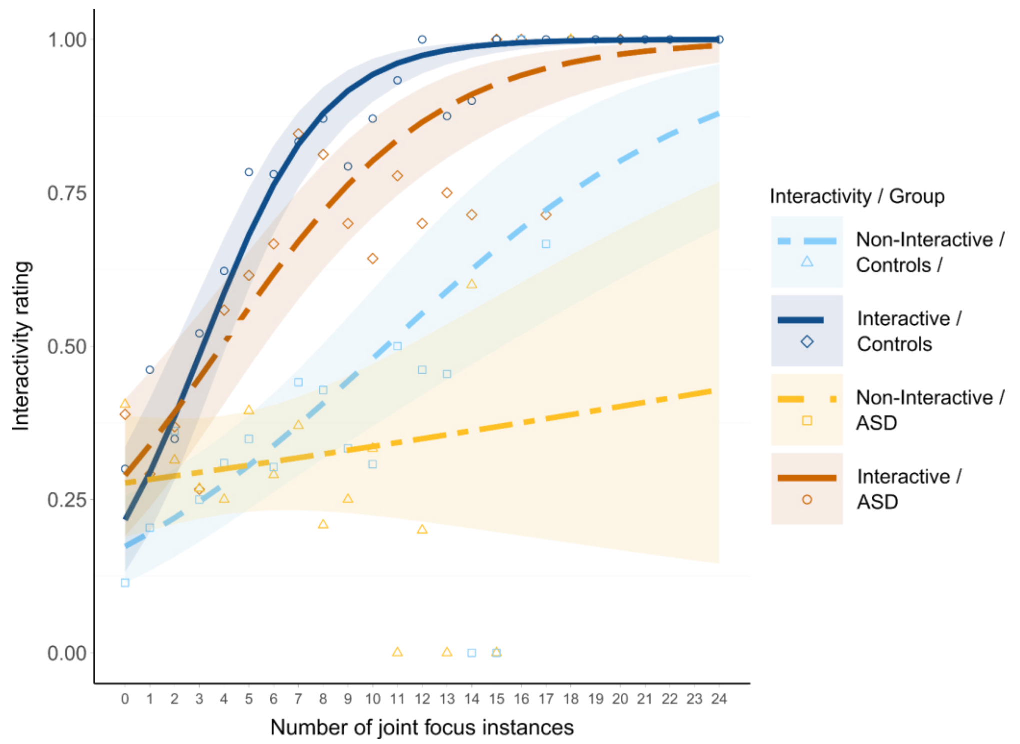


Figure 7. Relation between joint focus and interactivity ratings in diagnostic groups. Illustration of the distribution of instances of shared foci (eye contact or joint attention) between participant and agent, separately for a non-interactive agent (light colors) vs. an interactive agent (dark colors) in both diagnostic groups (blue: control participants / orange: ASD participants). A: Shared foci frequencies per trial for control participants (top) and ASD participants (bottom); B: Mean rates (triangles, diamonds, squares, and circles) and model predictions (lines) with 95% confidence intervals (ribbons) of interactivity ratings for differing numbers of shared foci instances per trial. The establishment of shared foci on average predicts the participants' interactivity ratings. This effect is especially strong for interactive compared to non-interactive agents and for control participants compared to ASD participants. (Adpated from Jording et al., submitted)

not differ between the two groups, neither in interactive nor non-interactive states (Figure 6C). Similiarly, both groups did not seem to differ in their ability to establish joint focus with their counterpart, i.e. eye contact or JA. However, the relationship between the establishment of joint focus and the impression of interactive intentions in the agent was weaker for the ASD group compared to the control group (Figure 7). These results suggested that a general impairment exists in persons with ASD to correctly identify intentions in gaze interactions. However, this impairment does not seem to be caused by reduced attention for socially relevant cues in the situations. Instead, it seems to be the more complex cognitive evaluations and the experience of the social phenomena emerging out of these situations like eye contact and JA that are disturbed in ASD.

3.6 Study 6: Towards Computer Aided Diagnosis of Autism Spectrum Disorder Using Virtual Environments (Roth, D., **Jording, M.**, Schmee, T., Kullmann, P., Navab, N., and Vogeley, K., in press)

The next step in the incremental increase of ecological validity in social gaze research will be the investigation of gaze interactions between two real persons. In addition, eventually the study of gaze communication has to be extended to other (nonverbal) communication channels. The major challenge is to create an environment for unrestricted real-time interactions that feel natural to the interaction partners but that can be recorded with high resolution.

Study 6 introduced a new system that allows two persons to interact with each other via gaze and speech. The interaction partners were seated in separate rooms and could speak to each

Figure 8. Exemplary setup of study 6. Two users sitting in remote rooms interact with each other via avatars and headsets. (Adapted from Roth et al., in press)

other via headsets while seeing each other's avatars on a computer screen (Figure 8). Each avatar mimicks the gaze and head movement of the person in the other room respectively and simulates lip movements when the person speaks. All data from head and gaze movements as well as audio data were recorded by the system for later analysis. We invited 9 dyads of participants with ASD and compared them to 10 dyads of persons without ASD. From each dyad, we recorded three conversations of 10 min duration each, the topics of which were given to the participants beforehand (1. "Introduction and getting to know each other"; 2. "Agree upon five items to take to a desert island"; 3. "Plan a five-course menu solely from ingredients they both dislike"). For the analysis, data from all conversations was separated into splits of 2.5 min and treated separately. For each split, 14 features were extracted, including the average dwell times on the AoIs eyes, mouth, small, medium and large head padding (Figure 9A), average horizontal and vertical shifts in gaze position on the screen, as well as the average head movement and rotation in all three dimensions. Three different classifiers (logistic regression model, Support Vector Machine (SVM), and Multilayer Perceptron neural network (MLP)) were then trained on the data and tested as to whether they were able to differentiate splits from ASD dyads and control dyads based on the 14 extracted features.

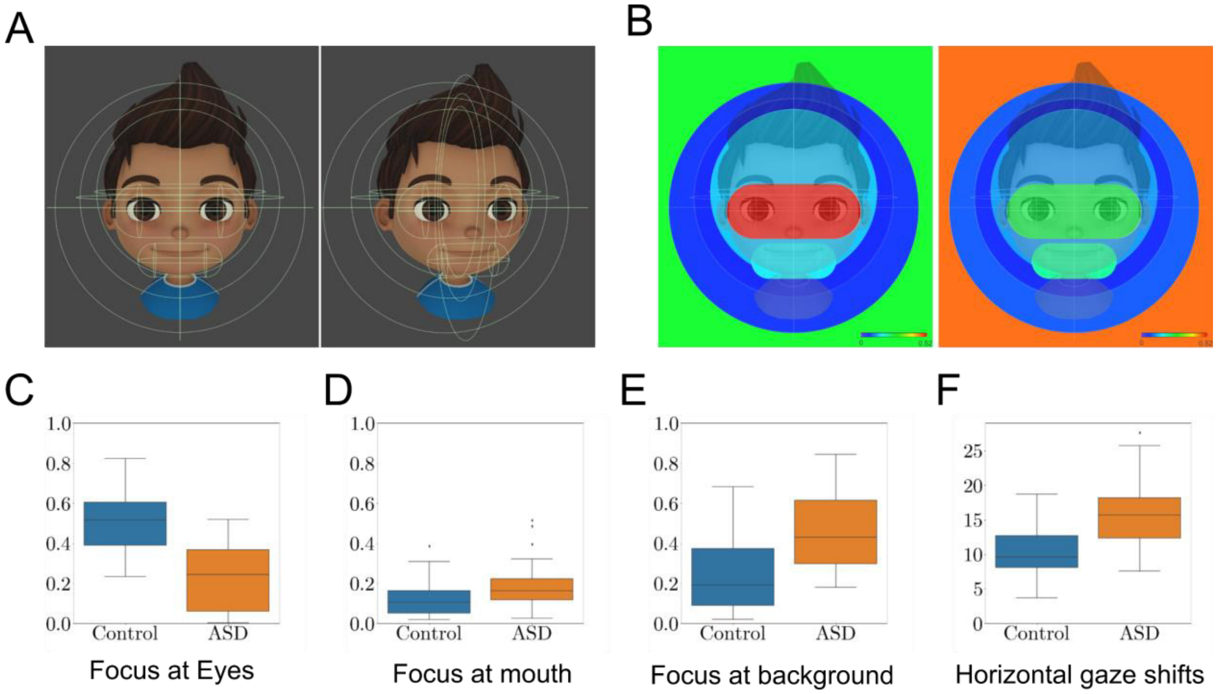


Figure 9. AoIs and group differences in the four most informative extracted features. A: Dynamic spatial AoIs of eyes, mouth, and large, medium, and small padded head. B: The heatmaps of average dwell times for control participants (left) show longer dwell times on the eyes AoI compared to participants with ASD (right). c) Whisker plots for the four features that were most informative for the differentiation of control (blue) and ASD (orange) splits: proportional dwell times on the AoIs eyes (C), mouth (D), and background (E) and average horizontal gaze shifts (F).

Results revealed that ASD participants and control participants exhibited distinct patterns of visual fixation (Figure 9B). While control participants focused significantly more on the eyes of their interaction partner (Figure 9C), ASD participants focused significantly more on the mouth (Figure 9D) and background (Figure 9E). Furthermore, ASD participants made significantly more horizontal gaze shifts (Figure 9F). Testing the three classifiers in their ability to differentiate splits from both groups of participants, the logistic regression model reached an accuracy of 88.5%, the SVM reached and accuracy of 88.6% and the MLP reached and accuracy of 92.9%.

The fixation patterns we observed for both groups are in accordance with previous results finding diminished eye focus and access mouth focus in persons with ASD (Klin et al., 2002a, 2002b, 2003; Neumann et al., 2006). The accuracy of the different classifiers was very promising with the MLPs accuracy of 92.9% by far exceeding that of traditional screening instruments for autistic symptoms (Conner et al., 2019). However, several limitations apply because of which these results should be treated very cautiously until they have been replicated. The most important ones are differences in the age and gender distribution of the two groups under investigation and limited data set for classification due to small sample sizes and rigorous data exlusion during preprocessing.

4 General Discussion

4.1 Ascribing Intentions from Passive Observation

Results from **study 1** showed that participants readily differentiated intentional states of a person by observing the person's eyes. Specific combinations of eye contact and gaze shifts away from the observer proved to elicit reliably the ascription of distinct intentions. Interestingly, data also demonstrate that participants were very sensitive even to subtle gaze cues and comparably small variations in gaze behavior, suggesting a strong disposition to turn to the eyes as a source of information about persons. Humans are used to and quite adept in observing other persons eyes for any clues that might help in predicting their future behavior or entail information about other objects in the environment. These observations happen constantly and automatically, not necessarily with any specific intention or conscious processing, i.e. we often observe and interpret our social environment without intending to or even being aware of it. As mentioned earlier, from these observations we gain information about a person's longlasting traits (George & Conty, 2008; Itier & Batty, 2009) as well as more ephemeral states (Baron-Cohen et al., 1997; Emery, 2000). In addition, we try to retrace their attentional focus and are thus able to make assumptions about their preferences or intentions towards objects in the environment. As discussed in more detail earlier, gaze processing as well as following an observed person's gaze angle towards an object happen reflexively. It also seems that these skills are learned very early in life, starting from birth. Given that these skills are prerequisites for the understanding and learning of more complex social skills, the apparent effortlessness with which they are applied even by infants is very plausible. In these abilities, humans exceed their closest animal relatives (Gibson & Pick, 1963) and several studies have identified specialized neural regions and circuits in humans that are predisposed to the recognition of eyes or the processing of gaze directions (cf. Carlin & Calder, 2013; Haxby et al., 2002; Itier & Batty, 2009; Nummenmaa & Calder, 2009).

However, **study 1** also showed that these passively observable gaze cues are not necessarily biuniquely informative but can be ambiguous. Depending on the context, the same cue could be interpreted to indicate different intentions. Communicative intentions were ascribed based on initial eye contact and subsequent extensive gaze shifts, whereas eye contact alone was taken as a sign for mere interest in the observer and extensive gaze shifts as a signal for private intentions. However, neither did eye contact impede the ascription of private intentions nor did larger subsequent gaze shifts decrease the probability of assuming an interest in the observer. This raises the question to which degree humans can rely on gaze cues in everyday life. In the following, I will elaborate on human mechanisms and techniques in dealing with the ambiguity of gaze cues as illustrated in **study 1**. However, I argue that at the core, the ambiguity roots in the unidirectionality of the situation and that ultimately none of these techniques can resolve this. In the subsequent part, I will therefore focus on the ascription of intentions in bidirectional gaze communication in interactive setups.

4.2 The Ambiguity of Unidirectional Gaze Communication

Interpreting gaze cues entails two problems, the reconstruction of another person's visual focus and the inference of the motivation driving it. As mentioned earlier, humans are in general quite proficient in retracing gaze and inferring an observed person's focus of attention. However, since the geometrical complexity of this problem does not allow for absolut accuracy, we have developed different mechanisms, allowing us to cope with the remaining uncertainty in specific circumstances. This is most obvious when processing gaze directed in our direction. In these situations, we tolerate deviations of up to several degrees, interpreting also slightly averted gaze as being directed at us (Jenkins et al., 2006; Mareschal et al., 2013, 2014). Similarily, when gaze points in the general direction of some object, we perceive it at being focused at the objects even when in fact being slightly diverted (Lobmaier et al., 2006). In particular, more decisive gaze shifts seem to trigger the expectation of being object-directed, as suggested by an increased gaze cueing effect for more distant cueing positions (Qian et al., 2013).

When having identified the focus of attention, we have to infer the motivation "behind" the behavior. It seems that here again, we partially rely on simple heuristics. As mentioned earlier, seeing another person looking at an object increased the likeability of the object (Bayliss et al., 2006; Landes et al., 2016; Ulloa et al., 2015). One explanation is that observers interpreted the observed person's gaze as a sign of interest and a positive attitude towards the object. This then in turn influenced their own attitude towards the object. Fittingly, observing a person looking in the direction of objects only increased the liking of the objects, when participants had the impression the observed person could actually see the object (Manera et al., 2014). When the observed person's gaze towards the object was obscured, it had no effect on the liking of the object. In addition, seeing a person pointing at an object, in contrast to looking at it, did not change the observer's preferences (Ulloa et al., 2015). This suggests that looking at an object was interpreted as a sign of genuine interest in the object, while pointing at it was understood as a mere communicative signal.

However, as discussed earlier, gaze can serve both private and communicative purposes (Gobel et al., 2015; Jarick & Kingstone, 2015) and looking at an object could either demonstrate personal interest or could be meant as a signal to others. Thus, the question remains of how observers can differentiate the two. Csibra and Gergely (2009, 2011) have speculated that humans use and are especially sensitive to additional “ostensive cues” as meta-communicative signals that convey their intentions. This theory was based on the observation that infants only followed their parents gaze when it was preceded by eye contact (Senju & Csibra, 2008). Accordingly, eye contact had been used to signal a “teaching” situation in which infants had expected being refered to something of specific interest (Csibra & Gergely, 2009, 2011).

However, recently infants have been shown to also follow with their gaze without preceeding eye contact (Gredebäck et al., 2018; Szufnarowska et al., 2014). Conversely, in **study 1** we found that depending on the situation the same gaze behavior was interpreted as signaling different intentions. Thus, it seems that there are no biuniquely informative cues that in each situation unequivocally inform observers about the actor’s intentions. As long as the observer is not able to reinsure with the actor, whether the interpretation of the intentions was correct, the ascription of intentions stays ambiguous. Only in truly interactive settings allowing for a dynamic exchange is it possible to jointly establish the meaning and purpose of the relationship (De Jaegher et al., 2010; De Jaegher & Di Paolo, 2007). In the following, the implications of direct exchange in “online” interactions for social cognition and specifically nonverbal communication research will therefore be illustrated.

4.3 New Developments in the Investigating of Gaze Interactions

Lately, it has become more and more evident that social cognition emerges out of social interactions and can only be investigated in truly interactive settings (De Jaegher et al., 2010; De Jaegher & Di Paolo, 2007; Pfeiffer et al., 2013a, 2013b; Risko et al., 2016; Schilbach, 2010). Many of the phenomena of social gaze discussed so far have been shown to rely heavily on the unrestricted exchange between interactants and do not occur when presented with static stimuli. This starts with a facilitated processing and increased physiological arousal (Hietanen et al., 2008; Pönkänen et al., 2011a, 2011b) for dynamic compared to static face stimuli, and increases in cognitive demand (Markson & Paterson, 2009) and self awareness (Hietanen & Hietanen,

2017) for face-to-face contact as opposed to non-interactive stimuli. With regard to gaze cueing and gaze following, the picture is less consistent but still emphasizes, at least partially, the importance of interactive designs. Neural responses to observing an interaction partner follow their gaze increased when participants believed their partner to be another human as compared to a computer algorithm (Caruana et al., 2017). Conversely, in another study, participants looked less at an interaction partner and followed less with their gaze when they believed to be engaged in a live interaction compared to a video recording (Gregory et al., 2015). Lachat and colleagues (2012) on the other hand found similar gaze cueing effects in live-interactions compared to studies with non-interactive designs. Interestingly, the effect of real interactions does not seem to stem from higher perceptual richness compared to other stimuli but to be based on the participant's believes about the reciprocity of the situation. In one study, the autonomic arousal accompanying eye-contact only increased when participants believed that their counterpart in a face-to-face interaction was able to see them as well (Myllyneva & Hietanen, 2015). In another study participants averted their eyes more in situations when there was the potential for an interaction compared to the passive observation of a dynamic video animation (Gregory & Antolin, 2019).

As of yet, we have not understood the full extent and implications of bidirectionality in gaze communication and the importance of more interactive research designs is becoming more and more evident. Aiming at investigating the ascription of intentions in an ecologically more valid setting, we therefore developed a new conceptual and methodological concept for social gaze research. The SGS (**study 2**) is a new taxonomy of social gaze that for the first time allows describing gaze interactions in their entirety. It identifies the different attentional states of the interactants during the encounter and structures interactions along these states. This allows for more systematic investigations of gaze interactions and provides a common terminology and classification system. Furthermore, the SGS describes the behavior associated with the different states in terms of probabilities instead of fixed courses and thus is more inclusive and can be applied more flexible. So far, investigations of social gaze have focused on a small number of fixed, short behavioral sequences, e.g. as categorized by Emery (2000). The SGS also allows predictions as to the dynamic enfolding between the interactants as well as to transitions between attentional states. With the SGS researchers are not limited anymore to gaze exchanges of a few seconds but can systematically investigate extended periods of interactions.

With TriPy, **study 3** provided the experimental platform that integrates the SGS for extended gaze-contingent human-agent interactions. The TriPy agent is able to display a broad repertoire of behaviors either in response to specific actions of the participant or independent of the participant. The temporal parameters of these behaviors are empirically informed based on previous studies as well as a pilot study (**study 3**). The experimenter can define various macro states for the agent with specific probabilities for the different behaviors. These macro states do not determine the exact course of the interaction but describe the overall distribution of the agent's visual attention and responsiveness towards the participant. For the states of the SGS, we measured the specific distributions of attention and responsiveness individually and implemented these parameters in TriPy (**study 3**). Thus, the agent is now able to display the most fundamental attentional and interactional states during triadic gaze encounters. TriPy can be used to investigate gaze interactions in dynamic, extended human-agent encounters with an unrivalled combination of ecological validity and experimental control. However, it is also possible to include additional or alternative states with other parameters in TriPy in order to costumize the agent's behavior.

Even higher levels of ecological validity can be achieved in setups for interactions between two real persons as described in **study 6**. In this configuration, people were able to speak to each other while seeing each other's head and gaze movements mapped to avatars. Mediating the interactions via avatars of course still diminished the ecological validity compared to real-world face-to-face interactions. However, the loss was outweighed by the benefits in terms of experimental control that come with avatar-mediated designs. With avatar-mediated communication it can be ensured that all cues and signals that transpire between participants are recorded and can subsequently be analyzed. Furthermore, the use of avatars rules out effects of individual differences in appearances and physiognomy and it allows the experimenter to focus on specific nonverbal communication channels while "switching off" others. Especially when investigating gaze communication in ASD confounding effects of impairments in other nonverbal communication channels have to be ruled out. For example, impairments have been reported for the recognition (Humphreys et al., 2007), as well as the production (Trevisan et al., 2018) of facial expressions in ASD, although the exact extent of these impairments is still under dispute (Keating & Cook, 2020). Furthermore, there is also evidence that the coordination and synchronization between different communication channels might be disturbed in ASD (de Marchena & Eigsti, 2010).

At the current state, analysing and interpreting the complex data from interactions between two persons is still very challenging due to our limited understanding of these interactions. Here, more controlled experiments using human-agent interaction platforms as in **studies 1**, **4** and **5** can help develop hypothesis and models of gaze interactions. However, in the future, these hypothesis and models will have to be tested in interactions between two real persons. Study 6 suggested that machine-learning approaches could be of great assistance in this as well as in the development of new hypothesis.

4.4 The Ascription of Intentions in Gaze Interactions

In **study 4** we investigated the ascription of intentions in interactive encounters with a gaze contingent agent that was able react dynamically to the participant´s behavior. The agent (**study 3**) displayed different gaze states (**study 2**), in some of which it was actively responding to the participants' gaze and some in which its behavior was not contingent upon the participants' gaze. The participants, while believing to be interacting with a real human (confederate), were instructed to ascertain whether their counterpart was trying to interact with them. Results show that participants were able to reliably distinguish interactive from non-interactive agent states. Thus, given a setting that allows for a more or less unrestricted interaction, participants can infer intentions from mere gaze exchange.

Furthermore, participants even distinguished between the individual non-interactive states. This corroborates the findings from **study 1**, suggesting a very high sensitivity to subtle cues when having to rely on passively observing the gaze of another person. Participants did not rate the person oriented agent as more interactive than the other two non-interactive agents, despite higher rates of eye contact and its close association to the initiation of interactions (Cary, 1978; Emery, 2000; Kleinke, 1986). However, this is in accordance with studies showing high sensitivity to situations allowing for an interactive exchange as supposed to non-interactive encounters (Hietanen & Hietanen, 2017; Markson & Paterson, 2009).

We had expected that an agent that followed the participants' gaze would be more easily recognized as interactive than an agent that was trying to lead the participant's gaze. This hypothesis was based on the assumption that in the first case, participants only had to notice that the agent was following them while in the latter, they themselves had to actively follow the agent's gaze in order to keep the interaction going. In addition, humans seem to generally expect others to follow their gaze (Pfeiffer et al., 2011) and thus might already await this form of interaction. However, the results show that situations in which the participants had to follow the agent's gaze (RJA) were more often recognized as interactive then situations in which participants had to initiate JA (IJA). Two different explanations are possible: The first is based on the conceptually more complex nature of IJA (see above) as suggested by studies finding RJA but not IJA in chimpansees (Tomasello & Carpenter, 2005) and an earlier onset of RJA (Morales et al., 2000; Senju & Csibra, 2008) compared to IJA (Mundy et al., 2007; Mundy & Newell, 2007) in human children. The second explanation is that participants did not simply try to establish JA but also wanted to make sure, that the agent was aware of the situation. Detecting these instances of shared attention (Emery, 2000), requires a more elaborate interaction between the interactants then is the case for JA (Pfeiffer et al., 2012). With an agent who is taking the initiative and then waits for the participant to react this might be easier than when the participant has to wait for the agent's reaction. This second explanation also seems to be supported by results from the analysis of instances of joint focus (subsuming states of eye contact and JA). In general, the number of joint focus instances was increased for interactive compared to non-interactive states. Consequently, there was a correlation between the number of joint focus instances and the probability of the agent being rated as interactive. Interestingly, this tendency was increased when the agent was actually responsive as compared to non-responsive encounters. Thus, it seems as if participants indeed were able to assess whether the agent had an active part in the establishment of a joint focus.

Two additional findings of study 4 are worth mentioning here, as they are relevant to the discussion of methodological and technological approaches towards social gaze. First, the average time it took participants to decide whether the agent was trying to interact with them was well above 10 seconds. This is important, considering that many previous studies of related phenomena have used experimental designs which restricted encounters to a few seconds (Oberwelland et al., 2016, 2017; Pfeiffer et al., 2011, 2012; Wilms et al., 2010). Our data now suggest, that for the sake of ecological validity, research should treat gaze interactions as a phenomenon emerging over an extended period of time. This is also in line with more recent reflections on the temporal extension of cognition (Vogel et al., 2020). Secondly, there was no evidence for a correlation between the participants' performance and their experience of task difficulty. This suggests that the problem at hand and potentially other phenomena of social gaze interactions might not be cognitively accessible. Similarly, it has been shown that in certain situations participants detected and also reacted to an agent following their gaze without becoming aware of it (Courgeon et al., 2014; Grynszpan et al., 2017). This is thus another example of the automatic nature of the perception and processing of nonverbal cues (Choi et al., 2005), a notion that has to be taken into account by future studies of social gaze.

4.5 Gaze-Based Ascription of Intentions in ASD

In accordance with the procedures from **study 4**, in **study 5** we investigated the ability of persons with ASD to ascribe intentions in a triadic interaction with a gaze contingent agent (**study 3**). To this extend we compared a group of participants with ASD to a group of healthy participants. Results showed that participants with ASD were specifically impaired in their ability to recognize the agent's intention to interact. This corresponds with findings about deficits in ASD in inferring mental states from the passive observation of static eye region stimuli (Baron-Cohen et al., 1997, 2001). However, there was no general decrease in the rate of correctly recognizing non-interactive state. This suggests that there is no general bias to evaluate interaction partners as being less interactive in ASD but rather that specifically communicative cues are either not perceived or misinterpreted. Additionally, there was no evidence of differences in response time between participants with ASD and the control group. Thus, it seems that the diminished performance in ASD is not merely the result of reduced processing speed.

Interestingly, no significant differences in the distribution of visual attention, neither generally nor specifically for interactive or non-interactive states were observed. This seems counterintuitive, given reports about the attenuation of impairments in ASD in more interactive and complex settings as compared to static stimuli (cf. Chita-Tegmark, 2016; Falck-Ytter & von Hofsten, 2011; Frazier et al., 2017; Guillon et al., 2014). However, we instructed participants explicitly to pay attention to their partner. In similar situations in studies about facial processing (Schulte-Rüther et al., 2013, 2017) or gaze processing (Oberwelland et al., 2017) participants with ASD were better able to allocate attention to the relevant elements of the scene. Differences in the allocation of attention in ASD might be specific to situations that do not require visual attention to selective features of the scene. In **study 6**, were participants had no other task then to talk to their partner, we found very pronounced differences in the allocation of visual attention between participants with and without ASD. However, in the context of **study 5**, attentional differences can be ruled out as the cause for the impairments in the ascription of intentions in ASD.

Instead, results suggested that it is the evaluation of communicative signals that might be impaired in ASD. The analysis of instances of joint focus revealed no significant difference in the occurence of such events between the two groups. Furthermore, instances of joint focus increased the probability of the ascription of interactive intentions. However, this relationship was significantly stronger for the control group compared to the ASD group. This suggests that participants from the control group associated the occurence of joint focus more strongly with communicative intentions of their counterpart. This raises the question, whether ASD participants were less sensitive to these situations or whether they interpreted them differently. Some previous studies suggest an impairment in the recognition of social contingencies in ASD (Gergely, 2001; Klin, 2000), while another study did not find differences (Zapata-Fonseca et al., 2018).

Another explanation would be that participants with ASD did not have the same understanding or did not attribute the same importance to their partner's behavior. It has been argued that rather than the perception of social cues themselves, it is the extraction of socially relevant information from these cues that is impaired in ASD (Nation & Penny, 2008; Ristic et al., 2005; Senju & Johnson, 2009b). Several different studies seem to support this hypothesis. For example participants with ASD were able to detect social stimuli but their evaluation was impaired in studies about person perception (Georgescu et al., 2013) or animacy experience (Kuzmanovic et al., 2014). In non-autistic persons sharing attention and the emerging interaction on their own have a rewarding effect (Pfeiffer et al., 2014; Schilbach et al., 2010) which might not be the case in autistic persons. Similary, gaze following also seems to be an inherently social act in non-autistic persons but not necessarily for autistic persons. While control subjects shifted their attention especially as a reaction to observing a gaze shift in a eye stimulus, autistic participants did not differentiate between eye stimuli and geometric or arrow cues (Ristic et al., 2005; Senju et al., 2004; Vlamings et al., 2005). In general, it appears that the social character of the situation is less of a factor for autistic persons compared to non-autistic persons. Consequently, it has been suggested, that the core of ASD impairments in social cognition lies not in the behavior but in the experience during social interactions (Froese et al., 2013; Fuchs, 2015; Gallagher, 2004).

4.6 Future Directions of Social Gaze Applications in Clinical Contexts

As our understanding of gaze communication and impairments in mental disorders as in ASD grows, new potential applications in diagnostic and therapeutic contexts become tangible. The automatic classification of avatar-mediated communication through machine-learning algorithms demonstrated in **study 6** was able to reliably identify interlocutors with ASD. Several limitations apply, especially regarding the sample selection and data availability after data exlusion. Therefore, results at the current state should be considered a proof-of-concept and will have to be replicated with larger and more homogenous group samples. However, if results can be confirmed, the system introduced by **study 6** can provide a valuable extension to current diagnostic procedures.

The prevalence rate of ASD has continuously increased over the last years and is currently estimated up to 1 in 59 (Baio et al., 2018). However, the capacities for diagnosis of ASD did not keep up with this development. Many children with ASD receive the diagnosis only many month or even years after the onset of symptoms (Shattuck et al., 2009; Wiggins et al., 2006). The parents educational background and socioeconomic factors can further influence the time before the diagnosis is received (Fountain et al., 2011; Mazurek et al., 2014). Long waiting times, high costs and overall limited accessibility of ASD diagnosis can even prevent parents from having symptomatic children checked (Lewis, 2017). It is estimated that many persons with ASD, especially those with less severe symptoms or without developmental delay, are not earlier diagnosed than in adulthood (Lai & Baron-Cohen, 2015). This is especially serious as interventions in early childhood are considered most effective (Corsello, 2005). However, even later in life ASD diagnosis can have various beneficial effects (Stagg & Belcher, 2019). The ASD diagnosis can increase the awareness and thus foster treatment of comorbid psychiatric conditions that often accompany ASD like depression and anxiety disorders (Hollocks et al., 2019) or increases in suicide rates (Richa et al., 2014; Segers & Rawana, 2014). Furthermore, different skill trainings are available that can improve quality of life of affected persons. Persons with ASD have been found to be especially often affected by unemployment (Howlin, 2013) but specialized support can improve this situation drastically (Brooke et al., 2018). In conclusion, the increase of capacities and accessibility of ASD diagnosis is of high importance.

By definition, ASD can only be diagnosed in clinical interviews by trained clinicians. Still, several instruments were developed for the screening of patients, as well as as a supplement to clinical interviews. Unfortunately, many of the most frequently used screening and diagnostic tools lacked evidence or proofed to be of disappointing diagnostic validity in adults (Ashwood et al., 2016; Baghdadli et al., 2017; Conner et al., 2019; Hirota et al., 2018). The system introduced in **study 6** showed very promising results in the identification of persons with ASD based on their nonverbal behavior during a conversation. Two potential applications are conceivable: as a screening tool before the clinical interview or as a tool to assess nonverbal behavior automatically during the clinical interview. In the first case, it could pre-sort patients seeking a diagnosis. In the latter it could be integrated in the clinical interview and supplement the clinician's observations by standardized markers for nonverbal behavior. Both of these cases profit by the full automization of the systems as well as the simplicity of its application.

It seems likely that the approach profited from the interactive nature of the situation and pronouncement of ASD specific behaviors in these situations (Chita-Tegmark, 2016; Falck-Ytter & von Hofsten, 2011; Frazier et al., 2017; Guillon et al., 2014). However, in principle, the approach is not limited to conversations but could also be applied for assessments of nonverbal behavior in other situations. If it was combined with head-mounted VR displays for example, patients could be tested in a range of different standardized virtual scenarios.

4.7 Conclusion

In conclusion, we were able to show that humans reliably ascribe intentions based on others' gaze behavior. However, gaze cues can be ambigues, which can not be resolved through passive observations but only by interacting with the other person. Persons with ASD were considerably impaired in their ability to infer intentions in a gaze encounters with another person. In particular, communicative cues intended to establish an interaction seem to have been misinterpreted. This can potentially have severe consequences given the importance of nonverbal communication and specifically gaze communication in everyday life. Interestingly, it appears that instead of attentional or motivational factors, the participants' experiences of the situation were crucial here. This underlines the importance of extending social cognition research by taking into account not only behavioral but also experiential aspects of the situation. Incorporating the experiental dimension will improve our understanding of ASD including the underlying mechanisms and – given the conception of many psychiatric conditions as disorders of social cognition (Crespi & Badcock, 2008; Moutoussis et al., 2014; Vogeley, 2018; Vogeley & Newen, 2009) – of the broader field of clinical psychology and psychiatry. New VR based communication plattforms and machine-learning assisted data analysis can contribute to the development of applications in diagnostic and therapeutic contexts.

5 Summary

Understanding the intentions and actions of others, especially of other group members, is of paramount importance for our survival as individuals as well as for the whole group. To this end, nonverbal cues and specifically gaze cues of others provide an essential source of information. This thesis therefore investigated the question, how humans ascribe intentions to others either through passive observation or during interactive encounters. It included the development of a new conceptual and a new technological approach to social gaze research and additionally investigated impairments in the ascription of intentions in healthy control subjects as well as subjects with ASD as an application in the clinical domain.

First, we demonstrated that humans are accustomed to and proficient in ascribing intentions based on subtle gaze cues. In two experiments (**study 1**), we could show that participants developed a precise idea of the intentions of another person simply by observing their gaze behavior. However, the results also illustrated the ambiguities inherent to unidirectional gaze communication during passive observation. For the investigation of social gaze in more interactive settings with higher ecological validity, we then developed the SGS as a new taxonomy of social gaze (**study 2**). It comprises all central states of triadic gaze interactions from the perspective of one of the interactants. The SGS for the first time allowed holistic descriptions of unrestricted and temporally extended triadic gaze interactions. We then developed the human-agent interaction platform TriPy for triadic gaze encounters that implements the different gaze states (**study 3**). Thus, the agent can display and dynamically change between different states and behave in an empirically informed fashion. We were then able to investigate the ascription of intentions in unrestricted and temporally extended gaze encounters between healthy participants and the TriPy agent (**study 4**). Results demonstrated the proficiency of participants in recognizing the intentions of their counterpart through gaze interactions. It was suggested that participants based their decision on the establishment of eye contact and JA. Lastly, we examined the gaze based ascription of intentions in persons with ASD (**study 5**) showing a diminished ability to recognize interactive intentions in their counterpart compared to the control group. Interestingly, these impairments did not seem to be based in attentional or behavioral deficits but in the evaluation of cues and the experience of the interaction. Overall, the results substantiated the claim for more interactive and ecological valid approaches in social cognition research and the consideration of experiential factors in addition to behavior in social and clinical psychology and social neuroscience. As an outlook, this thesis concludes with **study 6,** which introduced a new setup for the automatic classification of ASD based on avatar-mediated communication between two persons. Results suggest that this setup could serve as a screening tool and corroborate the diagnosis of ASD by quantifiable markers of nonverbal communication.

6 References

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Publishing.

Ames, C., & Fletcher-Watson, S. (2010). A review of methods in the study of attention in autism. *Developmental Review*, *30*(1), 52–73. https://doi.org/10.1016/j.dr.2009.12.003

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

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

Argyle, M., Lefebvre, L., & Cook, M. (1974). The meaning of five patterns of gaze. *European Journal of Social Psychology*, *4*(2), 125–136. https://doi.org/10.1002/ejsp.2420040202

Ashwin, C., Ricciardelli, P., & Baron-Cohen, S. (2009). Positive and negative gaze perception in autism spectrum conditions. *Social Neuroscience*, *4*(2), 153–164. https://doi.org/10.1080/17470910802337902

Ashwood, K. L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F. S., Findon, J., Eklund, H., Spain, D., Wilson, C. E., Cadman, T., Young, S., Stoencheva, V., Murphy, C. M., Robertson, D., Charman, T., Bolton, P., Glaser, K., Asherson, P., … Murphy, D. G. (2016). Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological Medicine*, *46*(12), 2595–2604. https://doi.org/10/gg6bt4

Baghdadli, A., Russet, F., & Mottron, L. (2017). Measurement properties of screening and diagnostic tools for autism spectrum adults of mean normal intelligence: A systematic review. *European Psychiatry*, *44*, 104–124. https://doi.org/10/gbsfrw

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson, C., Rosenberg, White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L.-C., Harrington, R., Lopez, M., Fitzgerald, R. T., … Dowling, N. F. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR. Surveillance Summaries*, *67*(6), 1–23. https://doi.org/10/gfx5sc

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

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

Bateson, M., Callow, L., Holmes, J. R., Redmond Roche, M. L., & Nettle, D. (2013). Do Images of 'Watching Eyes' Induce Behaviour That Is More Pro-Social or More Normative? A Field Experiment on Littering. *PLoS ONE*, *8*(12), e82055. https://doi.org/10/ggjg92

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414. https://doi.org/10.1098/rsbl.2006.0509

Bayliss, A. P., Murphy, E., Naughtin, C. K., Kritikos, A., Schilbach, L., & Becker, S. I. (2012). "Gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General*, *142*(1), 76–92. https://doi.org/10.1037/a0029286

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review*, *13*(6), 1061–1066. https://doi.org/10.3758/BF03213926

Bayliss, A. P., & Tipper, S. P. (2006). Predictive Gaze Cues and Personality Judgments: Should Eye Trust You? *Psychological Science*, *17*(6), 514–520. https://doi.org/10.1111/j.1467-9280.2006.01737.x

Bird, G., Press, C., & Richardson, D. C. (2011). The Role of Alexithymia in Reduced Eye-Fixation in Autism Spectrum Conditions. *Journal of Autism and Developmental Disorders*, *41*(11), 1556–1564. https://doi.org/10.1007/s10803-011-1183-3

Borji, A., Parks, D., & Itti, L. (2014). Complementary effects of gaze direction and early saliency in guiding fixations during free viewing. *Journal of Vision*, *14*(13), 1–32. https://doi.org/10.1167/14.13.3

Brooke, V., Brooke, A. M., Schall, C., Wehman, P., McDonough, J., Thompson, K., & Smith, J. (2018). Employees with Autism Spectrum Disorder Achieving Long-Term Employment Success: A Retrospective Review of Employment Retention and Intervention. *Research and Practice for Persons with Severe Disabilities*, *43*(3), 181–193. https://doi.org/10/gd3fz7

Brooks, C. I., Chruch, M. A., & Fraser, L. (1986). Effects of Duration of Eye Contact on Judgments of Personality Characteristics. *The Journal of Socal Psychology*, *126*(1), 71–78. https://doi.org/10.1080/00224545.1986.9713572

Burgoon, J. K. (1994). Nonverbal signals. In M. L. Knapp & J. A. Daly (Eds.), *The Sage handbook of interpersonal communication* (4th ed, pp. 450–507). SAGE.

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*. https://doi.org/10.3389/fpsyg.2019.00560

Carbon, C.-C., & Hesslinger, V. M. (2011). Bateson et al.'s (2006) Cues-of-Being-Watched Paradigm Revisited. *Swiss Journal of Psychology*, *70*(4), 203–210. https://doi.org/10.1024/1421-0185/a000058

Carlin, J. D., & Calder, A. J. (2013). The neural basis of eye gaze processing. *Current Opinion in Neurobiology*, *23*(3), 450–455. https://doi.org/10.1016/j.conb.2012.11.014

Caruana, N., de Lissa, P., & McArthur, G. (2017). Beliefs about human agency influence the neural processing of gaze during joint attention. *Social Neuroscience*, *12*(2), 194–206. https://doi.org/10.1080/17470919.2016.1160953

Cary, M. S. (1978). The Role of Gaze in the Initiation of Conversation. *Social Psychology*, *41*(3), 269–271. https://doi.org/10.2307/3033565

Castiello, U. (2003). Understanding other people's actions: Intention and attention. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 416–430. https://doi.org/10.1037/0096-1523.29.2.416

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

Chawarska, K., Klin, A., & Volkmar, F. (2003). Automatic Attention Cueing Through Eye Movement in 2-Year-Old Children With Autism. *Child Development*, *74*(4), 1108–1122. https://doi.org/10.1111/1467-8624.00595

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

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, *48*, 79–93. https://doi.org/10/f76qfz

Choi, Y. S., Gray, H. M., & Ambady, N. (2005). The glimpsed world: Unintended communication and unintended perception. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The New Unconscious* (pp. 309–333). Oxford University Press.

Conner, C. M., Cramer, R. D., & McGonigle, J. J. (2019). Examining the Diagnostic Validity of Autism Measures Among Adults in an Outpatient Clinic Sample. *Autism in Adulthood*, *1*(1), 60–68. https://doi.org/10/gg6bt5

Conty, L., N'Diaye, K., Tijus, C., & George, N. (2007). When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing. *Neuropsychologia*, *45*(13), 3024–3037. https://doi.org/10/c3hd2j

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

Corsello, C. M. (2005). Early Intervention in Autism. *Infants & Young Children*, *18*(2), 74–85. https://doi.org/10.1097/00001163-200504000-00002

Courgeon, M., Rautureau, G., Martin, J.-C., & Grynszpan, O. (2014). Joint Attention Simulation Using Eye-Tracking and Virtual Humans. *IEEE Transactions on Affective Computing*, *5*(3), 238–250. https://doi.org/10/ggv7m6

Crespi, B., & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences*, *31*(3), 241–261. https://doi.org/10/bnd8g4

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

Csibra, G., & Gergely, G. (2011). Natural Pedagogy as Evolutionary Adaptation. In A. Whiten, R. A. Hinde, C. B. Stringer, & K. N. Laland (Eds.), *Culture Evolves* (pp. 377–392). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199608966.003.0022

Csibra, G., & Volein, Á. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, *26*(1), 1–11. https://doi.org/10/cmk3mx

Dalmaso, M., Castelli, L., Scatturin, P., & Galfano, G. (2017). Trajectories of social vision: Eye contact increases saccadic curvature. *Visual Cognition*, *25*(1–3), 1–8. https://doi.org/10.1080/13506285.2017.1290727

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507. https://doi.org/10.1007/s11097-007-9076-9

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

de Marchena, A., & Eigsti, I.-M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research*, *3*(6), 311–322. https://doi.org/10.1002/aur.159

Doherty-Sneddon, G., & Phelps, F. G. (2005). Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition*, *33*(4), 727–733. https://doi.org/10.3758/BF03195338

Dratsch, T., Schwartz, C., Yanev, K., Schilbach, L., Vogeley, K., & Bente, G. (2013). Getting a Grip on Social Gaze: Control over Others' Gaze Helps Gaze Detection in High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *43*(2), 286–300. https://doi.org/10.1007/s10803-012-1569-x

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

Droney, J. M., & Brooks, C. I. (1993). Attributions of Self-Esteem as a Function of Duration of Eye Contact. *The Journal of Social Psychology*, *133*(5), 715–722. https://doi.org/10/bhrwdh

Einav, S., & Hood, B. M. (2006). Children's use of the temporal dimension of gaze for inferring preference. *Developmental Psychology*, *42*(1), 142–152. https://doi.org/10/b62zf8

Ellsworth, P. C., Carlsmith, J. M., & Henson, A. (1972). The stare as a stimulus to flight in human subjects: A series of field experiments. *Journal of Personality and Social Psychology*, *21*(3), 302–311. https://doi.org/10/ch7dh4

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

Falck-Ytter, T., & von Hofsten, C. (2011). How special is social looking in ASD: a review. *Progress in Brain Research*, *189*, 209–222. https://doi.org/10.1016/B978-0-444-53884-0.00026-9

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

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

Fletcher-Watson, S., Leekam, S. R., Findlay, J. M., & Stanton, E. C. (2008). Brief Report: Young Adults with Autism Spectrum Disorder Show Normal Attention to Eye-Gaze Information—Evidence from a New Change Blindness Paradigm. *Journal of Autism and Developmental Disorders*, *38*(9), 1785–1790. https://doi.org/10.1007/s10803-008-0548-8

Fountain, C., King, M. D., & Bearman, P. S. (2011). Age of diagnosis for autism: individual and community factors across 10 birth cohorts. *Journal of Epidemiology & Community Health*, *65*(6), 503–510. https://doi.org/10/dqsshk

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A Meta-Analysis of Gaze Differences to Social and Non-social Information Between Individuals With and Without Autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(7), 546–555. https://doi.org/10/gbncjs

Freeth, M., & Bugembe, P. (2019). Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults? *Autism*, *23*(2), 503–513. https://doi.org/10/ggwj5p

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

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

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

Froese, T., Stanghellini, G., & Bertelli, M. O. (2013). Is it normal to be a principal mindreader? Revising theories of social cognition on the basis of schizophrenia and high functioning autism-spectrum disorders. *Research in Developmental Disabilities*, *34*(5), 1376–1387. https://doi.org/10.1016/j.ridd.2013.01.005

Fuchs, T. (2015). Pathologies of Intersubjectivity in Autism and Schizophrenia. *Journal of Consciousness Studies*, *22*, 191–214.

Gale, A., Spratt, G., Chapman, A. J., & Smallbone, A. (1975). EEG correlates of eye contact and interpersonal distance. *Biological Psychology*, *3*(4), 237–245. https://doi.org/10/d5s3nn

Gallagher, S. (2004). Understanding Interpersonal Problems in Autism: Interaction Theory as An Alternative to Theory of Mind. *Philosophy, Psychiatry, & Psychology*, *11*(3), 199–217. https://doi.org/10.1353/ppp.2004.0063

George, N., & Conty, L. (2008). Facing the gaze of others. *Neurophysiologie Clinique/Clinical Neurophysiology*, *38*(3), 197–207. https://doi.org/10.1016/j.neucli.2008.03.001

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., & Vogeley, K. (2014). The Use of Virtual Characters to Assess and Train Non-Verbal Communication in High-Functioning Autism. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00807

Georgescu, A. L., Kuzmanovic, B., Schilbach, L., Tepest, R., Kulbida, R., Bente, G., & Vogeley, K. (2013). Neural correlates of "social gaze" processing in high-functioning autism under systematic variation of gaze duration. *NeuroImage: Clinical*, *3*, 340–351. https://doi.org/10.1016/j.nicl.2013.08.014

Gergely, G. (2001). The obscure object of desire-'Nearly, but clearly not, like me': Contingency preference in normal children versus children with autism. *Bulletin of the Menninger Clinic*, *65*(3), 411–426. https://doi.org/10/b8mtwv

Gibson, J. J., & Pick, A. D. (1963). Perception of Another Person's Looking Behavior. *The American Journal of Psychology*, *76*(3), 386–394. https://doi.org/10.2307/1419779

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*, 651–658. https://doi.org/10.3758/BF03211385

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

Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze Following Is Not Dependent on Ostensive Cues: A Critical Test of Natural Pedagogy. *Child Development*, *86*(6), 2091–2098. https://doi.org/10.1111/cdev.13026

Gregory, N. J., & Antolin, J. V. (2019). Does social presence or the potential for interaction reduce social gaze in online social scenarios? Introducing the “live lab” paradigm. *Quarterly Journal of Experimental Psychology*, *72*(4), 779–791. https://doi.org/10/gg8m6k

Gregory, N. J., López, B., Graham, G., Marshman, P., Bate, S., & Kargas, N. (2015). Reduced Gaze Following and Attention to Heads when Viewing a “Live” Social Scene. *PLOS ONE*, *10*(4), e0121792. https://doi.org/10.1371/journal.pone.0121792

Grynszpan, O., Nadel, J., Martin, J.-C., & Fossati, P. (2017). The awareness of joint attention: A study using gaze contingent avatars. *Interaction Studies*, *18*(2), 234–253. https://doi.org/10/gcr292

Guéguen, N., & Jacob, C. (2002). Direct Look Versus Evasive Glance and Compliance With a Request. *The Journal of Social Psychology*, *142*(3), 393–396. https://doi.org/10/dfq82d

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

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye Contact and Face Scanning in Early Infancy. *Science*, *198*(4319), 853–855. https://doi.org/10.1126/science.918670

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

Heider, F. (1958). *The psychology of interpersonal relations*. Erlbaum.

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. https://doi.org/10.1016/j.concog.2017.01.019

Hirota, T., So, R., Kim, Y. S., Leventhal, B., & Epstein, R. A. (2018). A systematic review of screening tools in non-young children and adults for autism spectrum disorder. *Research in Developmental Disabilities*, *80*, 1–12. https://doi.org/10/gd4ddt

Hollocks, M. J., Lerh, J. W., Magiati, I., Meiser-Stedman, R., & Brugha, T. S. (2019). Anxiety and depression in adults with autism spectrum disorder: a systematic review and meta-analysis. *Psychological Medicine*, *49*(4), 559–572. https://doi.org/10/gd6s9p

Hood, B. M., Macrae, C. N., Cole-Davies, V., & Dias, M. (2003). Eye remember you: the effects of gaze direction on face recognition in children and adults. *Developmental Science*, *6*(1), 67–71. https://doi.org/10/fcb7cz

Howard, M. A., Cowell, P. E., Boucher, J., Broks, P., Mayes, A., Farrant, A., & Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism: *NeuroReport*, *11*(13), 2931–2935. https://doi.org/10/fm4qwn

Howlin, P. (2013). Social Disadvantage and Exclusion: Adults With Autism Lag Far Behind in Employment Prospects. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(9), 897–899. https://doi.org/10/f2xkg4

Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. *Neuropsychologia*, *45*(4), 685–695. https://doi.org/10/bqbkqp

Hutchins, T. L., & Brien, A. (2016). Conversational topic moderates social attention in autism spectrum disorder: Talking about emotions is like driving in a snowstorm. *Research in Autism Spectrum Disorders*, *26*, 99–110. https://doi.org/10.1016/j.rasd.2016.03.006

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*(6), 843–863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Jarick, M., & Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01423

Jenkins, R., Beaver, J. D., & Calder, A. J. (2006). I thought you were looking at me: Direction-specific aftereffects in gaze perception. *Psychological Science*, *17*(6), 506–513. https://doi.org/10.1111/j.1467-9280.2006.01736.x

Keating, C. T., & Cook, J. L. (2020). Facial Expression Production and Recognition in Autism Spectrum Disorders. *Child and Adolescent Psychiatric Clinics of North America*, *29*(3), 557–571. https://doi.org/10/ggx8c3

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, *20*(4), 442–462. https://doi.org/10/f8js3b

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

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The Social Attribution Task. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *41*(7), 831–846.

Klin, A. (2006). Autism and Asperger syndrome: an overview. *Revista Brasileira de Psiquiatria*, *28*(suppl 1), 3–11. https://doi.org/10/bq3g48

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 345–360. https://doi.org/10.1098/rstb.2002.1202

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002a). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002b). Defining and Quantifying the Social Phenotype in Autism. *American Journal of Psychiatry*, *159*(6), 895–908. https://doi.org/10.1176/appi.ajp.159.6.895

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

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

Kuzmanovic, B., Schilbach, L., Georgescu, A. L., Kockler, H., Santos, N. S., Shah, N. J., Bente, G., Fink, G. R., & Vogeley, K. (2014). Dissociating animacy processing in high-functioning autism: Neural correlates of stimulus properties and subjective ratings. *Social Neuroscience*, *9*(3), 309–325. https://doi.org/10.1080/17470919.2014.886618

Kylliainen, A., & Hietanen, J. K. (2004). Attention orienting by another's gaze direction in children with autism. *Journal of Child Psychology and Psychiatry*, *45*(3), 435–444. https://doi.org/10.1111/j.1469-7610.2004.00235.x

Lachat, F., Conty, L., Hugueville, L., & George, N. (2012). Gaze Cueing Effect in a Face-to-Face Situation. *Journal of Nonverbal Behavior*, *36*(3), 177–190. https://doi.org/10/gc6zb4

Lai, M.-C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry*, *2*(11), 1013–1027. https://doi.org/10/f7xjhr

Landes, T. L., Kashima, Y., & Howe, P. D. L. (2016). Investigating the Effect of Gaze Cues and Emotional Expressions on the Affective Evaluations of Unfamiliar Faces. *PLOS ONE*, *11*(9), e0162695. https://doi.org/10.1371/journal.pone.0162695

Lee, K., Eskritt, M., Symons, L. A., & Muir, D. (1998). Children's use of triadic eye gaze information for "mind reading." *Developmental Psychology*, *34*(3), 525. https://doi.org/10.1037/0012-1649.34.3.525

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, *15*(1), 77–95. https://doi.org/10/bmprjn

Lewis, L. F. (2017). A Mixed Methods Study of Barriers to Formal Diagnosis of Autism Spectrum Disorder in Adults. *Journal of Autism and Developmental Disorders*, *47*(8), 2410–2424. https://doi.org/10/gbqmnb

Lobmaier, J. S., Fischer, M. H., & Schwaninger, A. (2006). Objects Capture Perceived Gaze Direction. *Experimental Psychology*, *53*(2), 117–122. https://doi.org/10.1027/1618-3169.53.2.117

MacDonald, R., Anderson, J., Dube, W. V., Geckeler, A., Green, G., Holcomb, W., Mansfield, R., & Sanchez, J. (2006). Behavioral assessment of joint attention: A methodological report. *Research in Developmental Disabilities*, *27*(2), 138–150. https://doi.org/10.1016/j.ridd.2004.09.006

Macrae, C. N., Hood, B. M., Milne, A. B., Rowe, A. C., & Mason, M., F. (2002). Are You Looking at Me? Eye Gaze and Person Perception. *Psychological Science*, *13*(5), 460–464. https://doi.org/10.1111/1467-9280.00481

Manera, V., Elena, M. R., Bayliss, A. P., & Becchio, C. (2014). When seeing is more than looking: Intentional gaze modulates object desirability. *Emotion*, *14*(4), 824–832. https://doi.org/10.1037/a0036258

Mareschal, I., Calder, A. J., & Clifford, C. W. G. (2013). Humans Have an Expectation That Gaze Is Directed Toward Them. *Current Biology*, *23*(8), 717–721. https://doi.org/10.1016/j.cub.2013.03.030

Mareschal, I., Otsuka, Y., & Clifford, C. W. G. (2014). A generalized tendency toward direct gaze with uncertainty. *Journal of Vision*, *14*(12), 1–9. https://doi.org/10.1167/14.12.27

Markson, L., & Paterson, K. B. (2009). Effects of gaze-aversion on visual-spatial imagination. *British Journal of Psychology*, *100*(3), 553–563. https://doi.org/10.1348/000712608X371762

Mason, M., F., Hood, B., & Macrae, C. N. (2004). Look into my eyes: Gaze direction and person memory. *Memory*, *12*(5), 637–643. https://doi.org/10.1080/09658210344000152

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. https://doi.org/10.1111/j.0956-7976.2005.00809.x

Mazurek, M. O., Handen, B. L., Wodka, E. L., Nowinski, L., Butter, E., & Engelhardt, C. R. (2014). Age at First Autism Spectrum Disorder Diagnosis: The Role of Birth Cohort, Demographic Factors, and Clinical Features. *Behavioral Pediatrics*, *35*(9), 561–569. https://doi.org/doi: 10.1097/DBP.0000000000000097

Montgomery, D. E., Back, L. M., & Moran, C. (1998). Children's Use of Looking Behavior as a Cue to Detect Another's Goal. *Child Development*, *69*(3), 692–705. https://doi.org/10/b87knq

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to Joint Attention Across the 6- Through 24-Month Age Period and Early Language Acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298. https://doi.org/10/cr9sm8

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

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. *Consciousness and Cognition*, *25*, 67–76. https://doi.org/10/f5vnwf

Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry*, *44*(6), 793–809. https://doi.org/10.1111/1469-7610.00165

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual Differences and the Development of Joint Attention in Infancy. *Child Development*, *78*(3), 938–954. https://doi.org/10.1111/j.1467-8624.2007.01042.x

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

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology*, *6*(3), 389–401. https://doi.org/10.1017/S0954579400006003

Myllyneva, A., & Hietanen, J. K. (2015). There is more to eye contact than meets the eye. *Cognition*, *134*, 100–109. https://doi.org/10.1016/j.cognition.2014.09.011

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology*, *20*(1), 79–97. https://doi.org/10.1017/S0954579408000047

Neumann, D., Spezio, M. L., Piven, J., & Adolphs, R. (2006). Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social Cognitive and Affective Neuroscience*, *1*(3), 194–202. https://doi.org/10.1093/scan/nsl030

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

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

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2017). Young adolescents with autism show abnormal joint attention network: A gaze contingent fMRI study. *NeuroImage: Clinical*, *14*, 112–121. https://doi.org/10.1016/j.nicl.2017.01.006

Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., & Vogeley, K. (2012). Eyes on the Mind: Investigating the Influence of Gaze Dynamics on the Perception of Others in Real-Time Social Interaction. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00537

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, *101*, 124–137. https://doi.org/10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., & Schilbach, L. (2011). A Non-Verbal Turing Test: Differentiating Mind from Machine in Gaze-Based Social Interaction. *PLoS ONE*, *6*(11), e27591. https://doi.org/10.1371/journal.pone.0027591

Pfeiffer, U. J., Timmermans, B., Vogeley, K., Frith, C. D., & Schilbach, L. (2013a). Towards a neuroscience of social interaction. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00022

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013b). 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

Phelps, F. G., Doherty-Sneddon, G., & Warnock, H. (2006). Helping children think: Gaze aversion and teaching. *British Journal of Developmental Psychology*, *24*(3), 577–588. https://doi.org/10.1348/026151005X49872

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2011a). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, *6*(4), 486–494. https://doi.org/10.1093/scan/nsq068

Pönkänen, L. M., Peltola, M. J., & Hietanen, J. K. (2011b). The observer observed: Frontal EEG asymmetry and autonomic responses differentiate between another person's direct and averted gaze when the face is seen live. *International Journal of Psychophysiology*, *82*(2), 180–187. https://doi.org/10/dmq8ts

Qian, Q., Song, M., & Shinomori, K. (2013). Gaze cueing as a function of perceived gaze direction: Gaze cueing. *Japanese Psychological Research*, *55*(3), 264–272. https://doi.org/10.1111/jpr.12001

Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: implications for cognitive neuroscience. *European Journal of Neuroscience*, *21*(6), 1763–1766. https://doi.org/10.1111/j.1460-9568.2005.03986.x

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport*, *15*(16), 2553–2555. https://doi.org/10.1097/00001756-200411150-00025

Riby, D. M., & Hancock, P. J. B. (2008). Viewing it differently: Social scene perception in Williams syndrome and Autism. *Neuropsychologia*, *46*(11), 2855–2860. https://doi.org/10.1016/j.neuropsychologia.2008.05.003

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

Riby, D. M., & Hancock, P. J. B. (2009b). Do Faces Capture the Attention of Individuals with Williams Syndrome or Autism? Evidence from Tracking Eye Movements. *Journal of*

Autism and Developmental Disorders, *39*(3), 421–431. https://doi.org/10.1007/s10803-008-0641-z

Richa, S., Fahed, M., Khoury, E., & Mishara, B. (2014). Suicide in Autism Spectrum Disorders. *Archives of Suicide Research*, *18*(4), 327–339. https://doi.org/10/gg2h4m

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

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

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Current Directions in Psychological Science*, *25*(1), 70–74. https://doi.org/10.1177/0963721415617806

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, *24*(3), 715–718. https://doi.org/10/dnjptx

Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, *11*(6), 449–449. https://doi.org/10.1038/nrn2805-c1

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah, N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, *22*(12), 2702–2715. https://doi.org/10.1162/jocn.2009.21401

Schulte-Rüther, M., Greimel, E., Piefke, M., Kamp-Becker, I., Remschmidt, H., Fink, G. R., Herpertz-Dahlmann, B., & Konrad, K. (2013). Age-dependent changes in the neural substrates of empathy in autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *9*(8), 1118–1126. https://doi.org/10.1093/scan/nst088

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, *19*(8), 1354–1372. https://doi.org/10.1162/jocn.2007.19.8.1354

Schulte-Rüther, M., Otte, E., Adigüzel, K., Firk, C., Herpertz-Dahlmann, B., Koch, I., & Konrad, K. (2017). Intact mirror mechanisms for automatic facial emotions in children and adolescents with autism spectrum disorder: Intact mirror mechanisms in Austism. *Autism Research*, *10*(2), 298–310. https://doi.org/10/f9vkwd

Segers, M., & Rawana, J. (2014). What Do We Know About Suicidality in Autism Spectrum Disorders? A Systematic Review: Systematic review of suicidality in ASD. *Autism Research*, *7*(4), 507–521. https://doi.org/10/gfkc8x

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

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

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

Senju, A., Tojo, Y., Dairoku, H., & Hasegawa, T. (2004). Reflexive orienting in response to eye gaze and an arrow in children with and without autism. *Journal of Child Psychology and Psychiatry*, *45*(3), 445–458. https://doi.org/10.1111/j.1469-7610.2004.00236.x

Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., Lee, L.-C., Rice, C., Giarelli, E., Kirby, R., Baio, J., Pinto-Martin, J., & Cuniff, C. (2009). Timing of Identification Among Children With an Autism Spectrum Disorder: Findings From a Population-Based Surveillance Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(5), 474–483. https://doi.org/10/fn925n

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

Stagg, S. D., & Belcher, H. (2019). Living with autism without knowing: receiving a diagnosis in later life. *Health Psychology and Behavioral Medicine*, *7*(1), 348–361. https://doi.org/10/gg2v3n

Stass, J. W., & Willis, F. N. (1967). Eye contact, pupil dilation, and personal preference. *Psychonomic Science*, *7*(10), 375–376. https://doi.org/10.3758/BF03331131

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports*, *4*(1). https://doi.org/10.1038/srep05304

Tomasello, M., & Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees. *Monographs of the Society for Research in Child Development*, *70*(1), i–152. https://doi.org/10.1111/j.1540-5834.2005.00324.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(05), 675–691. https://doi.org/10.1017/S0140525X05000129

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

Trevisan, D. A., Hoskyn, M., & Birmingham, E. (2018). Facial Expression Production in Autism: A Meta-Analysis: Facial Expression Production in Autism. *Autism Research*, *11*(12), 1586–1601. https://doi.org/10/ghcrxg

Uljarevic, M., & Hamilton, A. (2013). Recognition of Emotions in Autism: A Formal Meta-Analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Ulloa, J. L., Marchetti, C., Taffou, M., & George, N. (2015). Only your eyes tell me what you like: Exploring the liking effect induced by other's gaze. *Cognition and Emotion*, *29*(3), 460–470. https://doi.org/10.1080/02699931.2014.919899

Vernetti, A., Smith, T. J., & Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1850), 20162747. https://doi.org/10.1098/rspb.2016.2747

Vlamings, P. H. J. M., Stauder, J. E. A., van Son, I. A. M., & Mottron, L. (2005). Atypical Visual Orienting to Gaze- and Arrow-Cues in Adults with High Functioning Autism. *Journal of Autism and Developmental Disorders*, *35*(3), 267–277. https://doi.org/10/dt73n9

Vogel, D. H. V., Jording, M., Kupke, C., & Vogeley, K. (2020). The Temporality of Situated Cognition. *Frontiers in Psychology*, *11*, 546212. https://doi.org/10/ghdd77

Vogeley, K. (2018). Communication as Fundamental Paradigm for Psychopathology. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 804–820). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.43

Vogeley, K., & Bente, G. (2010). "Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Networks*, *23*(8–9), 1077–1090. https://doi.org/10.1016/j.neunet.2010.06.003

Vogeley, K., & Newen, A. (2009). Consciousness of oneself and others in relation to mental disorders. In S. J. Wood, N. B. Allen, & C. Pantelis (Eds.), *The Neuropsychology of Mental Illness* (pp. 408–413). Cambridge University Press. https://doi.org/10.1017/CBO9780511642197.035

Vogeley, K., & Roepstorff, A. (2009). Contextualising culture and social cognition. *Trends in Cognitive Sciences*, *13*(12), 511–516. https://doi.org/10.1016/j.tics.2009.09.006

von Griinau, M., & Anston, C. (1995). The detection of gaze direction: A stare-in-the-crowd effect. *Perception*, *24*(11), 1297–1313. https://doi.org/10.1068/p241297

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, *6*(3), 313–326. https://doi.org/10.1068/p060313

Whalen, C., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry*, *44*(4), 456-468. https://doi.org/10.1111/1469-7610.00135

Wiggins, L. D., Baio, J., & Rice, C. (2006). Examination of the Time Between First Evaluation and First Autism Spectrum Diagnosis in a Population-based Sample: *Journal of Developmental & Behavioral Pediatrics*, *27*(Supplement 2), 79–87. https://doi.org/10/fc58tb

Wilms, M., Schilbach, L., Pfeiffer, U. J., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes - using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(1), 98–107. https://doi.org/10.1093/scan/nsq024

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

Wu, D. W.-L., Bischof, W. F., & Kingstone, A. (2013). Looking while eating: The importance of social context to social attention. *Scientific Reports*, *3*, Article 2356. https://doi.org/10.1038/srep02356

Yarbus, A. L. (1967). *Eye Movements and Vision*. Springer US. https://doi.org/10.1007/978-1-4899-5379-7

Zapata-Fonseca, L., Froese, T., Schilbach, L., Vogeley, K., & Timmermans, B. (2018). Sensitivity to Social Contingency in Adults with High-Functioning Autism during Computer-Mediated Embodied Interaction. *Behavioral Sciences*, *8*(2), 22. https://doi.org/10.3390/bs8020022

7 Appendix

Jording, M., Engemann, D., Eckert, H., Bente, G., & Vogeley, K. (2019). Distinguishing Social From Private Intentions Through the Passive Observation of Gaze Cues. Frontiers in Human Neuroscience, 13, 442.

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2018). The "Social Gaze Space": A Taxonomy for Gaze-Based Communication in Triadic Interactions. Frontiers in Psychology, 9.

Hartz, A., Guth, B., **Jording, M.,** Vogeley, K., & Schulte-Rüther, M. (submitted). TriPy: An Artificial Social Agent for Systematic Investigations of Gaze-Based Communication in Triadic Interactions.

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2019). Inferring Interactivity From Gaze Patterns During Triadic Person-Object-Agent Interactions. Frontiers in Psychology, 10, 1913.

Jording, M., Hartz, A., Vogel, D., Schulte-Rüther, M., & Vogeley, K. (submitted). Reduced Experience of Interactivity During Gaze-Based Interactions in Participants with Autism Spectrum Disorder.

Roth, D., **Jording, M.,** Schmee, T., Kullmann, P., Navab, N., & Vogeley, K. (2020). Towards Computer Aided Diagnosis of Autism Spectrum Disorder Using Virtual Environments. 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR).

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 17 December 2019
doi: 10.3389/fnhum.2019.00442

Distinguishing Social From Private Intentions Through the Passive Observation of Gaze Cues

Mathis Jording[1,2]*, Denis Engemann[1,3], Hannah Eckert[2], Gary Bente[4] and Kai Vogeley[1,2]

[1] Cognitive Neuroscience (INM-3), Institute of Neuroscience and Medicine, Research Center Jülich, Jülich, Germany, [2] Department of Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany, [3] Université Paris-Saclay, Inria, CEA, Palaiseau, France, [4] Department of Communication, Michigan State University, East Lansing, MI, United States

Observing others' gaze is most informative during social encounters between humans: We can learn about potentially salient objects in the shared environment, infer others' mental states and detect their communicative intentions. We almost automatically follow the gaze of others in order to check the relevance of the target of the other's attention. This phenomenon called gaze cueing can be conceptualized as a triadic interaction involving a gaze initiator, a gaze follower and a gaze target, i.e., an object or person of interest in the environment. Gaze cueing can occur as "gaze pointing" with a communicative or "social" intention by the initiator, telling the observer that she/he is meant to follow, or as an incidental event, in which the observer follows spontaneously without any intention of the observed person. Here, we investigate which gaze cues let an observer ascribe a social intention to the observed person's gaze and whether and to which degree previous eye contact in combination with an object fixation contributes to this ascription. We varied the orientation of the starting position of gaze toward the observer and the orientation of the end position of a lateral gaze shift. In two experiments participants had to infer from the gaze behavior either mere approach ("the person looked at me") vs. a social ("the person wanted to show me something") or a social vs. a private motivation ("the person was interested in something"). Participants differentially attributed either approach behavior, a social, or a private intention to the agent solely based on the passive observation of the two specific gaze cues of start and end position. While for the attribution of privately motivated behavior, participants relied solely on the end position of the gaze shift, the social interpretation of the observed behavior depended additionally upon initial eye contact. Implications of these results for future social gaze and social cognition research in general are discussed.

Keywords: social gaze, Bayesian multilevel models, ostension, eye contact, communicative intention, gaze cueing

Edited by:
Peter König,
University of Osnabrück, Germany

Reviewed by:
José Luis Ulloa,
University of Talca, Chile
Camila Rosa De Oliveira,
Faculdade Meridional (IMED), Brazil

***Correspondence:**
Mathis Jording
m.jording@fz-juelich.de

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *03 July 2019*
Accepted: *02 December 2019*
Published: *17 December 2019*

Citation:
Jording M, Engemann D, Eckert H, Bente G and Vogeley K (2019) Distinguishing Social From Private Intentions Through the Passive Observation of Gaze Cues. Front. Hum. Neurosci. 13:442. doi: 10.3389/fnhum.2019.00442

INTRODUCTION

The eye region displays emotional and attentional states and is a crucial element in understanding the inner experiences of others (Baron-Cohen et al., 1997; Emery, 2000). This leads to the pivotal role of gaze in social cognition research (Shepherd, 2010) because it informs not only about internal states of persons but also about their relationship to objects or persons in their environment.

Humans process the gaze direction, deduce from it the focus of attention and automatically shift their own attention accordingly. This process is called gaze cueing (Frischen et al., 2007) and is a prerequisite for joint attention, the case in which both persons visually attend the same object. Observing someone looking at objects also informs us about the environment shared by both partners. Accordingly, following someone's gaze changes the perception and processing of jointly attended objects (Becchio et al., 2008); objects, that had previously been looked at by another person are liked more (Bayliss et al., 2006). Gaze following is acquired early in life: 6 month old infants are already able to follow someone's gaze (Senju and Csibra, 2008). Proficiency in gaze following predicts the development of language "theory of mind" capacity (Morales et al., 1998), IQ, self-regulation, social competence and depth of information processing (Mundy and Newell, 2007). It is also believed to be a prerequisite component for reinforcement learning (Vernetti et al., 2017).

A key research question is whether successful gaze processing is an automatic holistic ability, or whether it can be decomposed into distinct cognitive operations, hence, taught and learned. As a clear prerequisite, the gaze angle has to be estimated and the spatial location of the partner's attention has to be inferred from the gaze vector. Compared to great apes and monkeys, humans are especially proficient in this regard (Gibson and Pick, 1963), and the neural implementation of gaze reconstruction has been intensely researched over the past decades (Itier and Batty, 2009).

A second challenge is to discern intentions underlying gaze behavior, which may be explicitly communicative or "social" in the sense that gaze partners want to convey certain information. The "dual function" of gaze comprises the perception of the environment and the signaling of the attentional focus to others (Gobel et al., 2015). I.e., we do not only use the gaze of others as a cue about their attentional focus, but we are at the same time aware that others can deduce our attentional focus from our gaze. Effects of this awareness have been demonstrated impressively in studies showing that participants control their gaze according to its social adequacy when being watched (Risko and Kingstone, 2011). In other words, humans are forced to actively avoid undesired communication by controlling their eye gaze in social contexts. Likewise, when observing another person, this person's gaze might be driven by self-centered interests or it might be an attempt to communicate or to express a "social" intention. Thus when deducing the other's intentions, perceivers have to distinguish between "private" and "communicative" intentions (Walter et al., 2004). It can be expected that this distinction fundamentally affects our relationship toward the other person. Walter et al. (2004) could show that, during mentalizing, the processing of private and communicative intentions rely on distinct neural mechanisms, even if the communicative actions are not directed toward the observer.

Csibra and Gergely (2009, 2011) speculate that humans use eye contact as an "ostensive" signal to announce situations in which they want to show or teach something to others. Being gazed at by another person is a powerful social cue to which most humans are highly sensitive (von Griinau and Anston, 1995; Senju and Johnson, 2009), and eye contact is supposed to signal communicative intents (Kleinke, 1986). Conversely, according to Csibra and Gergely (2009, 2011), infants have an innate sensibility to ostensive cues which allows them to generalize their experience in these situations in order to fully benefit from their teacher. Preceding communication indeed has been shown to facilitate subsequent gaze cueing and gaze following already in 4–6 month old infants (Farroni et al., 2003). This mechanism might also explain the strong ontogenetic link not only between gaze following and joint attention, but also between the mental and cognitive development.

Here, we present two studies that explore the link between gaze direction processing and communicative or "social" affordances. We investigate the principles of how humans deduce the attentional focus from others' gaze with regard to the tension between private and communicative or "social" intentions. The motivation for Study 1 was to study the role of eye contact in reducing the ambiguity of gaze and to identify the parameters that allow to interpret the gaze behavior of others as ostensive, i.e., a special case of communicative intention that bridges the gap between person and environment. Specifically, we aimed at the difference between situations in which we experience an interacting partner as being interested in us by visually attending to us in contrast to situations in which the partner is actively trying to communicate with us about something in the outside world by a rudimentary form of joint attention. The observation of distinct patterns of observed gaze in the two conditions lead us to the question in Study 2, whether and how participants distinguish aforementioned communicative intentions from situations in which the partner is experienced as being "privately" interested in something without involving and addressing the perceiver.

As the basic design of both studies, participants watched short videos of a virtual character (VC) looking at the participant with different degrees of vertical deviations, ranging from direct gaze (i.e., eye contact) to different degrees of downward averted gaze, before shifting the gaze to the left or to the right with different degrees of lateral deviations. (For simplicity, we will refer to the starting position of initial gaze as "initial position" and to the gaze shift to the left or to the right as "shift amplitude"). Subsequently, participants had to report their experiences based on explicit statements (see **Figure 1**). We used VCs as stimulus material, as they combine high experimental control with ecological validity (Vogeley and Bente, 2010) and are well suited for the investigation of non-verbal communication (Pfeiffer et al., 2013, 2014; Georgescu et al., 2014; Jording et al., 2019).

In the first study, we investigated the difference between situations in which participants had the impression of been looked at by the VC ("LOOK" condition) and situations in which they had the impression that the VC was trying to show them something ("COM," e.g., "communicative," condition). Besides the aforementioned empirical question, a second goal of this first study was to ensure the validity of our stimuli and the overall methods. The sensitivity of human observers to the visual stimulus of eyes directed at them is already well established (Senju and Johnson, 2009) and VCs were shown to reliably

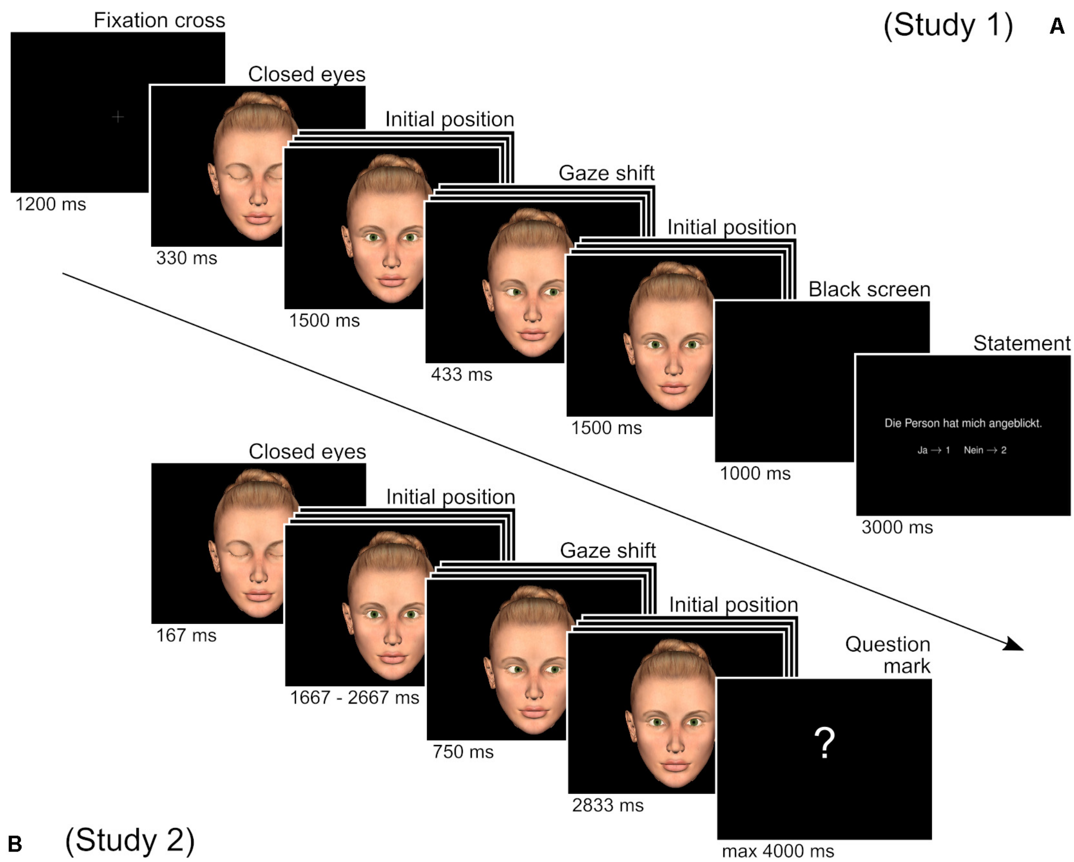


FIGURE 1 | Course of one trial of Study 1 (A) and Study 2 (B). The stack of images for initial position and shift position indicate that in each trial, one out of four possible images was displayed. Note that after the shift, the VC always returned to the same initial position it had started from. In this example, the initial position 1 (direct gaze) and the shift position 4 are depicted. The question mark indicated the prompt for participants to give their ratings.

induce the impression of social presence (Bente et al., 2007). Our stimuli can elicit the feeling of being in the attentional focus of or being addressed by the VC. Therefore, results of this first study serve as a test of these properties of the stimuli. This first study was conducted as an internet-based online survey to maximize sample size and account for possible variability in the general population. In the second study with a new sample of participants, we again studied COM in comparison to the situation in which the VC was merely privately interested in something without any social intention ("PRIV" condition). This second study had a repeated measures design and was conducted in a laboratory setting, increasing experimental control of environment and participant specific factors.

We expected the impression of being looked at to be dependent solely on the degree to which the initial gaze is directed toward the participants but not on subsequent outward-directed behavior. In COM, available evidence in the field suggested an influence of preceding eye contact for the impression of communicative intentions as well. Considering that participants were asked whether the other wanted to show them something located in the outside world, we also expected an influence of the subsequent gaze shift during COM. However, this situation by definition requires a triadic interaction between two interactants and another object in the environment. Therefore, we expected high agreement rates only for situations with direct gaze and large shift amplitudes. During PRIV, we expected an influence of the shift amplitude only. However, it was also interesting to see whether preceding eye contact might have an adverse effect. Should participants understand private and communicative intentions as mutually exclusive, they should take eye contact as an indicator of the latter, leading to an impediment of the impression of mere personal interest.

MATERIALS AND METHODS

Study 1

Participants for Study 1

Out of 555 participants, 403 participants completed the online survey. In 11 cases videos were not presented correctly, resulting in 392 remaining participants (257 female; age ranging from 17–70 years, $M = 30$, SD = 10.63). Participants were recruited via mailing lists from different German universities (University of Cologne, University of Münster, University of Bayreuth) and gave their informed consent prior to participation. There were no further exclusion criteria.

Stimuli for Study 1

One female and one male VC were created with Poser for Apple Mac OS X (Poser 8, Smith Micro Software, Inc., Columbia, SA, United States). For both VCs images were created for four different initial gaze positions and for four different gaze shift targets in two different directions. Initial gaze positions were equidistantly positioned on a central vertical line, ranging from direct gaze to clearly averted gaze. Positions after gaze shifts were equidistantly located on a horizontal line slightly below the eye level, ranging from slight central deviation up to the maximal still realistic and lifelike appearing deviation, both for the right and the left side. From these images we approximated the deviation of the visual angle from direct gaze (initial position 1) by measuring for all images the position of the iris in relation to its position in the direct gaze image. On this basis we computed angles, taking 22 mm as the average diameter of the human eye (Bekerman et al., 2014) and 12 mm as average diameter of the human iris (Thainimit et al., 2013). Averaged between VCs, the initial positions vertically deviated approximately equidistantly from direct gaze by 0°, 3°, 8°, and 12°. VC-averaged gaze positions after the shift lay on a plane 6° vertically below the eye level, horizontally deviating from direct gaze approximately equidistantly by 5°, 9°, 14°, and 18°. (For examples of all initial positions and gaze shift images and the exact values of the degree of aversion, please refer to the **Supplementary Material**.) Images of initial positions and gaze shifts were then combined to flash videos by the python 2.6 based video tool "ffmpeg 0.7.8." For both sexes of VCs videos were created for each combination of four different initial positions and four different shift amplitudes to both sides, resulting in 16 videos of gaze shifts to the right and 16 videos for gaze shifts to the left per VC and a total of 64 videos. Each video started with showing a fixation cross for 1200 ms. Afterward the VC appeared, having his/her eyes closed for 330 ms before he/she subsequently opened the eyes and looked toward the initial position for 1500 ms, then shifted toward the target for 433 ms, before returning to the initial position for 2000 ms. Afterward the screen went black for 1000 ms, before the statement and response buttons were displayed for 3000 ms as a reminder at the end of the video (see **Figure 1A** for an illustration and **Supplementary Videos S1–S4** for examples of the trial course).

Task for Study 1

Each participant watched videos of either the female or the male VC for all 16 different combinations of gaze initial positions and shift amplitudes to the left or to the right in randomized order exactly once. After each video participants had to rate the VCs behavior according to statements randomly assigned in the beginning of the experiment. Statements were either "the person looked at me" (German original: "Die Person hat mich angeblickt") or "the person wanted to show me something" (German: "Die Person wollte mir etwas zeigen"), to which participant had to respond per button press in a binary choice ("yes" or "no").

Setup and Design for Study 1

The survey was presented via the online survey tool Unipark (Questback GmbH, Cologne, Germany). During the survey, participants were informed about the procedure, the voluntary nature of their participation and the opportunity to withdraw from the study at any point in time and without providing any reasons for their decision. They further had to state their age and sex before they were pseudo-randomly assigned to one of the two VCs and one of the two rating statements. After that, participants were told which statement they had to answer and whether they would see a female or a male character. Participants were then presented with videos for all 16 combinations of initial positions and shift amplitudes in a pseudorandomized order with shifts randomly either to the right or the left. After each video the screen turned black before the statement was presented together with the binary response options (button "1" for "Yes" and button "2" for "No"). The next trial then started after the participants had given their answers.

Statistics for Study 1

The effect of different gaze shifts (initial position and shift amplitude) on the ascription of different intentions to the VC (conditions) were analyzed in a multilevel model with an inverse logit link function, in which we considered individual differences between the participants' average responses through varying intercept coefficients. Importantly, we considered the statement as experimental condition and hence constructed a joint model for both statements instead of two separate models. The model focuses on the interaction between the statement and eye gaze behavior. This approach has enabled explicitly modeling statement-specific-biases, e.g., due to difficulty or individual preferences, while, at the same time subjecting the estimated differences between the effects to statistical control through shrinkage priors (see below). The resulting logistic regression model can be expressed as:

$$y_i \sim Binomial(n = 1, p = \hat{y}_i)$$
$$\hat{y}_i = logit^{-1}(\alpha_j[i] + T[i] * \beta)$$

Where α_j is the individual intercept for each subject, T is a matrix of treatment effects, and β the unknown parameter vector that has to be learnt from the data. The treatment effects are the statement, the vertical initial gaze position and the horizontal amplitude of the gaze shift, covering all main effects as well as second and third order interactions. The statement was dummy-coded with a 0–1 predictor. We included the eye gaze as continuous predictor after z-scoring. No prior information concerning effect sizes of the initial gaze position or shift amplitude were available. We hence used the non-informative default priors from the "brms" package (Bürkner, 2017, 2018) according to which coefficient are centered around zero. These priors are shrinkage priors and are conservative. Shrinkage is used in statistics to improve generalization to new data can be thought of correcting initial estimates by pushing them toward zero. The amount of shrinkage fades out as the sample size increases. For the prior for the population variance component σ_j of the individual intercepts, we kept the conservative default prior that puts most probability mass on smaller values close to zero.

$$\beta \sim student's\ t(df = 3,\ center = 0,\ \sigma^2 = 10)$$

$$\alpha_j \sim student's\ t(df = 3,\ center = 0,\ \sigma^2 = \sigma_j)$$

$$\sigma_j \sim half\text{-}student's\ t(df = 3,\ center = 0,\ \sigma^2 = \sigma_j)$$

Note that the population variance parameter σ_j uses the upper half of the student-t distribution due to the constraint that the variance cannot be negative. Also note that σ_j is a hyper-parameter and has to be estimated from the data. Here, it controls how much the model trusts the individual intercept estimates σ_j and to which extent these will be corrected by shrinkage toward the global intercept. Smaller values for σ_j would produce stronger shrinkage. This is a core feature of the multilevel model and is also referred to as partial pooling (Gelman, 2006).

We performed prior predictive checks to ensure that the priors are approximately uninformative on the scale of the model predictions after the inverse logistic link function. Analysis revealed that the results were insensitive to the choice of the prior due to the size of the data set. Data were analyzed using the "rstan" (Stan Development Team, 2018) and "brms" (Bürkner, 2017, 2018) packages for the programing language R for statistical computing (R Development Core Team, 2008) and RStudio (R Studio Team, 2016). Model fitting was performed using a Hamilton Markov chain Monte Carlo algorithm (Hoffman and Gelman, 2014). Models were run with 1000 warmup samples and 1000 iterations in total, using four chains, yielding 4000 draws from the approximated posterior distribution. Successful convergence was assessed based on the potential scale reduction factor $\hat{R}$, also known as the Gelman-Rubin statistic. $\hat{R}$, was found to be acceptably close to 1.0 ($\pm$0.1) for every model (see **Supplementary Table S1**). Posterior distributions were visually compared to observed data in order to check consistency.

Study 2

Participants for Study 2

34 subjects (19 female; age range 21–54, M = 28.88, SD = 5.82; not out of the sample from Study 1) participated in this experiment. None of these participants met any of the exclusion criteria (depressive symptoms as indicated by BDI scores: M = 3.79, range = 0–17, cut-off $\geq$ 19; autistic traits as indicated by AQ scores: M = 10.42, range = 2–19, cut-off $\geq$ 32; general cognitive impairments as indicated by MWT: M = 112.59, range = 97–136, cut-off < 70, or KAI, M = 124,24, range = 100–143, cut-off < 70) so that all participants were included for further analysis. The mean empathy score of the resulting sample as indicated by the SPF was M = 40.64, range = 30–49. Participants were recruited via mailing lists from the University of Cologne and gave their informed consent before participating.

Stimuli for Study 2

The same VC pictures were used as in Study 1. Instead of beforehand creating animated videos, as in Study 1, images were now combined to animations within the presentation software (Python 2.6), allowing for jittering of presentation durations. As in Study 1, animations of both VCs could be presented displaying gaze shifts for all 16 possible combinations of initial positions and shift amplitudes to both directions (left and right), resulting in a total of 32 different gaze shifts per VC. Each video sequence started with the VC having its eyes closed for 167 ms (10 frames) before opening them and looking toward the gaze initial position for 1667–2667 ms (100–160 frames). Afterward the VC gaze shifted and then stayed at the new location for 750 ms (45 frames) before returning to the initial location at the end of the video for another 2833 ms (170 frames). Subsequently, a screen showing a white question mark in front of a black background requested the participants to give their answer for a maximum of 4000 ms. (Please refer to **Figure 1B** for an illustration and **Supplementary Videos 5–8** for examples of the trial course).

Task for Study 2

In accordance with Study 1, participants, after having watched a gaze shift performed by the VC, had to rate the VCs behavior according to one of two different statements per trial. The statements were either "the person wanted to show me something" (German: "Die Person wollte mir etwas zeigen") or "the person was interested in something" (German original: "Die Person interessierte sich für etwas"). Again, participants had to respond per button press in a binary choice ("Yes" or "No"), for which they had 4 s before the next trial would start.

Setup and Design for Study 2

Before the experiment started, participants general cognitive level was assessed by two tests: KAI (Lehrl et al., 1991) and MWT-B (Lehrl, 2005). The experiment was conducted on a Lenovo ThinkPad T410 (Intel Core i5-520 M, 2,4 Ghz, 4GB RAM; OS: Ubuntu Linux 12.4 LTS) and displayed on a Tobii T60 Eye Tracker (60 Hz refresh rate, 1280 $\times$ 1024 px resolution) with responses given via keypad buttons and instructions presented on the screen. For the experiment, two blocks of trials (one block per statement) were presented in a pseudorandomized fashion. In each block, the participant watched all 64 gaze shifts (four initial positions $\times$ four shift amplitudes $\times$ two directions $\times$ two VCs) resulting in a total of 128 trials per participant over the whole experiment and a total duration of approximately 20 min. Before the experiment, KAI (Lehrl et al., 1991) and MWT-B (Lehrl, 2005) were conducted to rule out general cognitive impairments. After the experiment participants completed BDI (Beck et al., 2001), and AQ (Baron-Cohen et al., 2001) to rule out depressive and autism-like syndromes, respectively. In addition, participants filled out the empathy inventory SPF (Paulus, 2009) to potentially allow the matching with patient samples in future clinical studies.

Statistics for Study 2

The same statistical procedures where applied as in Study 1 (for $\hat{R}$ values see **Supplementary Table S2**). Note that the multilevel approach has allowed us to use the same model specification for Study 2, as this kind of model is robust to the structure of repeated observations and can be applied to a wide array of between or within-subject designs (see McElreath, 2016, Chapter 12, box on pp. 371 for discussion).

RESULTS

Interpreting multilevel models solely based on their coefficients is known to be notoriously difficult, especially for generalized linear models with non-Gaussian probability models (Ai and Norton, 2003). As is common practice, we therefore considered posterior predictions (**Figure 2** for Study 1; **Figure 4** for Study 2) in addition to model coefficients (**Figure 3**, Study 1; **Figure 5**, Study 2). The posterior predictions contain the uncertainty of the model and can be readily interpreted in terms of the probability of the responses given the model and the data. They conveniently support statistical inference and can be analyzed in terms of percentiles or subtracted from another to form contrasts. For the effect of the individual

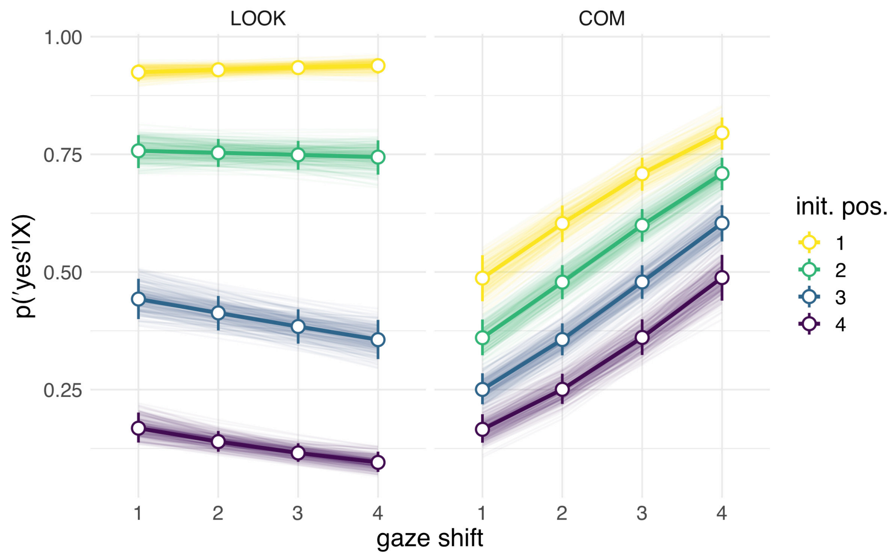


FIGURE 2 | Posterior predictions of the influence of initial position („init. pos.") and gaze shift amplitude („gaze shift") in Study 1 in the LOOK condition ("the person looked at me") and the COM condition ("the person wanted to show me something"). For the initial position, „1" corresponds to direct gaze and „4" to a maximally (vertically) averted position. For the shift amplitude, „1" corresponds to the smallest and „4" to the largest possible shifts.

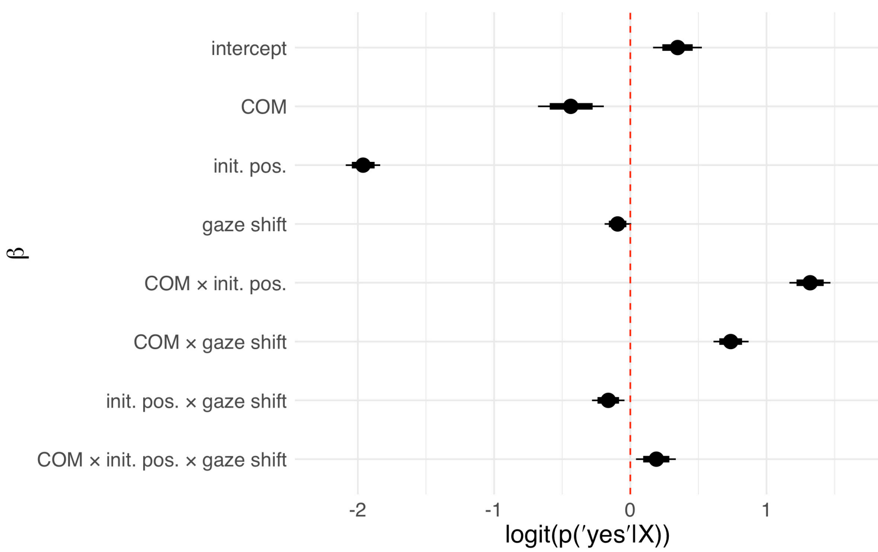


FIGURE 3 | Coefficients sampled from the approximate posterior distribution in Study 1 for the influence of condition, initial position, shift amplitude, and their respective interactions. Circles depict the posterior mean, horizontal bars and lines denote the 80 and 95% posterior compatibility intervals, respectively. The COM coefficient describes the effect of the COM condition in contrast to the LOOK condition. The coefficient for initial positions depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze). The coefficient of shift amplitude depicts the stepwise effect of increasing the shift amplitude. For additional statistics see **Supplementary Table S1**; Note that although not apparent here, the 95% confidence interval of the gaze shift coefficient does include zero.

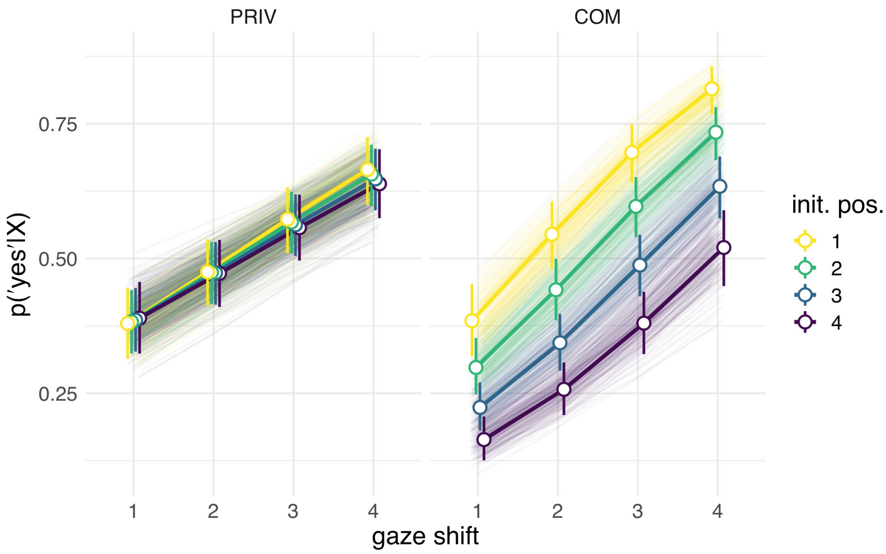


FIGURE 4 | Posterior predictions of the influence of initial position („init. pos.") and gaze shift amplitude („gaze shift") in Study 2 in the PRIV condition ("the person was interested in something") and the COM condition ("the person wanted to show me something"). For the initial position, „1" corresponds to direct gaze and „4" to a maximally (vertically) averted position. For the shift amplitude, „1" corresponds to the smallest and „4" to the largest possible shifts.

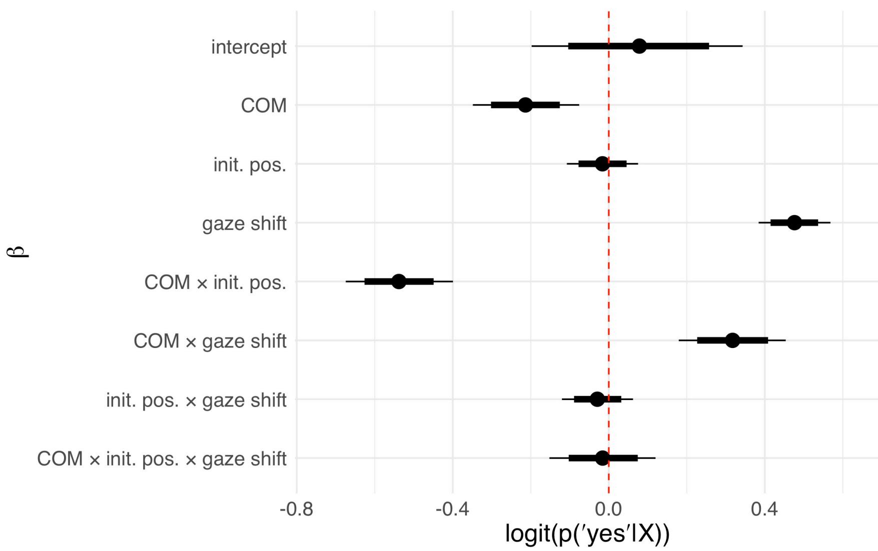


FIGURE 5 | Coefficients sampled from the approximate posterior distribution in Study 2 for the influence of condition, initial position, shift amplitude, and their respective interactions. Circles depict the posterior mean, horizontal bars and lines denote the 80 and 95% posterior compatibility intervals, respectively. The COM coefficient describes the effect of the COM condition in contrast to the PRIV condition. The coefficient for initial positions depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze). The coefficient of shift amplitude depicts the stepwise effect of increasing the shift amplitude. For additional statistics see **Supplementary Table S2**.

predictors, beta coefficients (as well as the respective 80 and 95% posterior probability distribution intervals) are reported in the **Figures 3**, **5**, additional statistics can be found in the **Supplementary Tables S1**, **S2**. This approach was chosen in order to increase the comparability to traditional reports of frequentist statistical methods with 0.05 significance levels. The intercepts for Study 1 and 2 refer to the COM condition, coefficients for the LOOK condition (Study 1) or the PRIV condition (Study 2) describe the change in coefficients compared to this intercept.

Study 1

In Study 1 (online study) 198 participants (134 female, 64 male; age: 17–66 years, M = 29.37, SD = 10.69) participated in the LOOK condition and 194 participants (123 female, 71 male; age: 18–70 years, M = 30.07, SD = 10.59) participated in the COM condition. We compared posterior predictions for agreements to LOOK ("the person looked at me") and COM ("the person tried to show me something") statements (**Figure 2**). Posterior predictions revealed that participants discriminated the two conditions based on the two gaze dimensions, initial position and shift amplitude (**Figure 2**). In the COM condition, the effect of initial position as well as shift amplitude had substantial effects with the probability of agreement to the statement "the person tried to show me something" increasing with initial positions closer to eye contact and larger shift amplitudes. In comparison, in the LOOK condition, the effect of the initial position was even more pronounced while the shift amplitude did not show any considerable effect on the probability of "the person looked at me" statement. In addition, a slight tendency to higher overall agreements to the LOOK compared to the COM statements is visible. These results are reflected in the configuration of the model coefficients (**Figure 3**) which uncovered higher order interaction effects between condition and the dimension of gaze shifts.

Study 2

In Study 2 (Lab Study) all 34 subjects participated in both conditions (COM and PRIV) in a repeated measures design. Here, we tested whether results from the COM condition in Study 1 could be replicated and how they would compare to the PRIV condition. In posterior predictions (**Figure 4**) for the COM condition the same pattern as in Study 1 arose with the probability of agreeing with the statement "the person tried to show me something" increasing with initial positions closer to eye contact and with larger shift amplitudes. Corroborating results of Study 1, no considerable interaction effect between initial position and shift amplitude was observed. In comparison, posterior predictions for the PRIV condition revealed that the overall tendency to agree with the statement "the person was interested in something" was slightly higher. Larger shift amplitude enhanced the probability of agreement even further, although this effect was less pronounced in PRIV compared to COM. Neither the initial position nor the interaction between initial position and shift amplitude had considerable effects in PRIV. Results correspond to the configuration of model coefficients (**Figure 5**), which uncovered simple but no higher order interactions.

DISCUSSION

The present study focused on the interplay of person-related and environment-related aspects of gaze behavior and how they influence our tendency to ascribe communicative or "social" and "private" intentions. The impression of being looked at (LOOK) has proved to be highly relying on initial eye contact for only in the conditions of direct gaze (or only slightly diverted gaze) ratings reached at least 75% agreement rates, while in cases of more diversion, agreement decreased substantially. Given the high sensitivity of humans to eye contact (von Grünau and Anston, 1995; Senju and Johnson, 2009) and its close link to intimacy (Argyle and Dean, 1965) this finding appears highly plausible. The amplitude of the subsequent gaze shift had no decisive influence, which corresponded also with our expectation.

The communicative condition (COM) revealed substantially the same results in the online study as in the laboratory study. Here, direct gaze or starting points close to it during the initial gaze and large gaze shifts significantly fostered the impression of being shown something. This matches the role of eye contact conveying communicative intentions (Kleinke, 1986) and nicely fits accounts of eye contact being used as ostensive cue. However, the ostensive situation also extends beyond the dyadic interaction of the two persons to the outside world. This is represented in the increasing effect of the assumed goal-directedness of the gaze shift. In other words, gaze contact with the viewer is only one component, the other component that makes this gaze behavior ostensive, is obviously the gaze shift directed toward an invisibly target in the environment. This result also ties in with other findings showing that infants as young as 9 month are not only sensitive to ostensive gaze cues, but they also expect object directed gaze shifts in these situations (Senju et al., 2008). Similarly, we had expected that participants would experience communicative intentions only when the triadic nature of the situation was apparent in the agents' gaze behavior. Accordingly, we expected to find an interaction effect between the degree of eye contact and shift amplitude for the COM condition. However, this interaction effect proved to be negligible compared to the observable main effects. Thus, in our initial hypothesis we overestimated the component to which participants considered contextual factors when inferring communicative intentions. The question therefore remains, to which extent the effect of ostensive signals facilitating gaze cueing can be ascribed to more fundamental levels of processing. When investigating the reallocation of attention in a similar situation, Bristow et al. (2007) were able to identify a corresponding interaction effect. BOLD-responses in the parieto-frontal attentional network indicated a stronger reallocation of attention for the observation of gaze shifts toward empty space vs. an object when the observed face had previously looked at the participant in contrast to an averted gaze condition. The authors assumed that the enhanced (visual) saliency of eyes directed at the viewer might have increased the gaze cueing effect.

When participants had to rate whether or not the VC appeared to be interested in something (PRIV), only shift amplitude had a notable effect with larger gaze shifts eliciting higher approval rates. We assume that participants tended to perceive small gaze shifts as still directed toward them. Despite the human general acuteness in retracing gaze vectors and directions, they show a surprising tolerance when identifying gaze directed at them with deviations up to several degrees (Gibson and Pick, 1963; Jenkins et al., 2006; Mareschal et al., 2013). Interestingly, this tendency is even stronger for participants that had experienced social exclusion prior to the experiment (Lyyra et al., 2017). We, however, did neither induce or ask explicitly for the experience of social exclusion.

It makes sense that participants, when asked whether the other one was interested in something, assumed this something in the outside world and took more decisive gaze shifts as reflecting this interest. In general, humans, when observing another persons' gaze, express some flexibility not only with regard to gaze directed at them, but also when it is directed at objects. We perceive a person as looking directly toward an object even in case of an actual divergence between gaze vector and object (Lobmaier et al., 2006). Unfortunately, research on the effect of the target position and shift amplitude in gaze cueing is still sparse. To the best of our knowledge, only one study investigated the gaze cueing effect as a function of the cued position, reporting higher effects for more distant positions (Qian et al., 2013). Our data now suggest, that when gaze shifts were more pronounced, participants more strongly imagined the existence of objects in their shared environment, even though not visible to them. However, due to the still insufficient knowledge about the underlying mechanisms this notion remains speculative.

It is interesting that the initial gaze does not influence the judgment. Even when initially eye contact was established, this did not impede the impression of privately motivated behavior so that the interpretation of the same behavior either as communicative or as private crucially depends on the instruction or the "mindset." Obviously, private and communicative intentions are not mutually exclusive, a person can be interested in something and therefore try to show it to others. However, at least in this highly reductionistic quasi-"social" context, participants did not or were not able to distinguish between those two situations.

Taken together, results corroborate that the combination of mere eye contact and lateral gaze shift together can already signal communicative intentions in a very robust way and can serve as powerful ostensive cue. However, data suggest that eye contact itself and even in combination with the subsequent gaze shift are not sufficient to biuniquely discern intentions from social gaze. The impression of communicative intentions was most prevalent in, but not limited to, the most profiled triadic situations, defined by initial eye contact and large gaze shift amplitudes. This is in line with results showing that ostensive gaze cues do not necessarily seem to be a prerequisite for gaze following in infants (Szufnarowska et al., 2014; Gredebäck et al., 2018). Conversely, eye contact did not inhibit the impression of private intentions. With regard to the differentiation between communicative and private intentions, this means that eye contact neither seems to constitute a highly predictive nor selective signal. Thus, the question remains, which other signals or processes might be used discern intentions from gaze.

Here, the highly reductionist approach of this study clearly reaches its limits. While it was warranted for elucidating the relationship between the most basic aspects of ostensive gaze behavior, its limitations have to be considered as well. First: Non-verbal communication in general was already pointed out to have a high procedural and dimensional complexity meaning that individual non-verbal cues are not isolated units but always part of a stream of cues from different non-verbal channels (Vogeley and Bente, 2010). Regarding the investigation of gaze behavior it is thus advisable not to limit the analysis to short chunks of gaze communication and potentially to include other non-verbal channels as well (Jording et al., 2018). Second: The context or environment has to be taken into account when investigating gaze processing (Hamilton, 2016). Adding and systematically varying objects to the setup as a focus point for the ostensive gaze cues would thus constitute another interesting variation of this study. Third: Closely linked to environmental aspects are factors regarding our knowledge about the other person. Although gaze cueing and gaze following can happen automatically, it is also influenced by our perception and beliefs about the other person as well as our relationship toward this person (Gobel et al., 2017). Thus, systematically manipulating the participants believes about of the observed agent (e.g., personality or preferences) might influence their interpretation of the observed gaze behavior.

CONCLUSION

In conclusion, although the two studies on gaze behavior presented here are highly minimalistic, they nevertheless substantially deepen our understanding of the powerful potential of social gaze in initiating interactions, referencing and displaying attention and thus allow a glimpse through the "window into social cognition" that social gaze can provide (Shepherd, 2010). Eye contact has again been proven to be a powerful tool in imparting communicative intents and fostering the impression that someone else is actively trying to show us something. However, it also becomes evident that eye contact itself is obviously not sufficient to discern intentions from social gaze biuniquely. Humans most likely make use of additional, e.g., temporal characteristics of gaze or they take other non-verbal or verbal signals into account; further investigations on this topic are therefore warranted. In practice, this study can inform us about the fundamental processes that underlie the perception and potentially production of gaze behavior and their functional roles in communication. Technically, these insights may help develop applications in the field of interaction and communication sciences by making use of anthropomorphic virtual agents and humanoids (Pfeiffer et al., 2013). In order for cognitive robots to become accepted as interaction partners by humans they have to share the human ability to generate and interpret informative gaze behavior as a two-way communicative act (Pönkänen et al., 2011; Gobel et al., 2015; Jording et al., 2018). A more thorough understanding of how humans convey and ascribe intentions as supplied here is therefore essential. In the long-run this approach might then also foster the development of more sophisticated agent-based diagnostic and therapeutic instruments for communication disorders like autism spectrum disorders (Georgescu et al., 2014).

DATA AVAILABILITY STATEMENT

Data will be made available in a public repository upon publication and till then can be accessed via https://osf.io/avu5w/?view_only=bd507466e05544589eac294c33253e8c.

ETHICS STATEMENT

This study followed the WMA Declaration of Helsinki (Ethical Principles for Medical Research Involving Human Subjects) and was presented to and approved by the Ethics Committee of the Medical Faculty of the University Hospital Cologne, Germany.

AUTHOR CONTRIBUTIONS

All authors substantially contributed to the conception of the work. DE programed the code for data collection in the online and in the lab study. HE recruited participants and conducted the lab experiment. MJ and DE conducted the statistical analyses. MJ drafted the manuscript. HE, DE, GB, and KV revised the manuscript critically.

FUNDING

KV received funding from the Federal German Ministry of Education and Research (Grant No.: 16SV7244) and from the EC, Horizon 2020 Framework Program, FET Proactive (Grant Agreement ID: 824128).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00442/full#supplementary-material

FIGURE S1 | Illustration of the female avatar's eyes and the measurement of the iris' position for gaze angle calculation. Middle column: Eye section from the female avatar stimuli for the initial gaze position (top four) and the position after the gaze shift (bottom four). Red circles with a centered cross mark the position of the iris as measured for the calculation of the gaze angle. Right column: lateral and horizontal deviations of the gaze angle from direct gaze. Note that depicted here are only gaze shifts to the left side; for shifts to the right side avatar stimuli were mirrored.

FIGURE S2 | Illustration of the male avatar's eyes and the measurement of the iris' position for gaze angle calculation. Middle column: Eye section from the female avatar stimuli for the initial gaze position (top four) and the position after the gaze shift (bottom four). Red circles with a centered cross mark the position of the iris as measured for the calculation of the gaze angle. Right column: lateral and horizontal deviations of the gaze angle from direct gaze. Note that depicted here are only gaze shifts to the left side; for shifts to the right side avatar stimuli were mirrored.

TABLE S1 | Coefficients sampled from the approximate posterior distribution in Study 1 for the influence of the COM condition, initial position, shift amplitude, and their respective interactions. The COM coefficient describes the effect of the COM condition in contrast to the LOOK condition; init. pos. depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze); gaze shift depicts the stepwise effect of increasing the shift amplitude. Reported are estimates (Estimate) and estimated errors (Est.Error) for the coefficients, the lower (l-95% CI) and the upper (u-95% CI) border of the 95% posterior compatibility intervals, the effective sample size (Eff.Sample) and the potential scale reduction factor $\hat{R}$ or Gelman-Rubin statistic ($\hat{R}$).

TABLE S2 | Coefficients sampled from the approximate posterior distribution in study 2 for the influence of the COM condition, initial position, shift amplitude, and their respective interactions. The COM coefficient describes the effect of the COM condition in contrast to the PRIV condition; init. pos. depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze); gaze shift depicts the stepwise effect of increasing the shift amplitude. Reported are estimates (Estimate) and estimated errors (Est.Error) for the coefficients, the lower (l-95% CI) and the upper (u-95% CI) border of the 95% posterior compatibility intervals, the effective sample size (Eff.Sample) and the potential scale reduction factor $\hat{R}$ or Gelman-Rubin statistic ($\hat{R}$).

VIDEO S1 | Example Study 1_init.1_shift.4.

VIDEO S2 | Example Study 1_init.2_shift.3.

VIDEO S3 | Example Study 1_init.3_shift.2.

VIDEO S4 | Example Study 1_init.4_shift.1.

VIDEO S5 | Example Study 2_init.1_shift.4.

VIDEO S6 | Example Study 2_init.2_shift.3.

VIDEO S7 | Example Study 2_init.3_shift.2.

VIDEO S8 | Example Study 2_init.4_shift.1.

REFERENCES

Ai, C., and Norton, E. C. (2003). Interaction terms in logit and probit models. *Econ. Lett.* 80, 123–129. doi: 10.1016/S0165-1765(03)00032-36

Argyle, M., and Dean, J. (1965). Eye-Contact, distance and affiliation. *Sociometry* 28, 289–304. doi: 10.2307/2786027

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a "Language of the Eyes"? Evidence from normal adults, and adults with autism or asperger syndrome. *Vis. Cogn.* 4, 311–331. doi: 10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17.

Bayliss, A. P., Paul, M. A., Cannon, P. R., and Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: i like what you look at. *Psychon. Bull. Rev.* 13, 1061–1066. doi: 10.3758/BF03213926

Becchio, C., Bertone, C., and Castiello, U. (2008). How the gaze of others influences object processing. *Trends Cogn. Sci.* 12, 254–258. doi: 10.1016/j.tics.2008.04.005

Beck, A. T., Steer, R. A., and Hautzinger, M. (2001). *Beck-Depressions-Inventar (BDI): Testhandbuch. 2., überarb. Aufl., 1. Nachdr*. Bern: Huber.

Bekerman, I., Gottlieb, P., and Vaiman, M. (2014). Variations in Eyeball Diameters of the Healthy Adults. *J. Ophthalmol.* 2014, 1–5. doi: 10.1155/2014/503645

Bente, G., Eschenburg, F., and Krämer, N. C. (2007). "Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations," in *Virtual reality*. ed R. Shumaker (Berlin: Springer), 185–194. doi: 10.1007/978-3-540-73335-5_21

Bristow, D., Rees, G., and Frith, C. D. (2007). Social interaction modifies neural response to gaze shifts. *Soc. Cogn. Affect. Neurosci.* 2, 52–61. doi: 10.1093/scan/nsl036

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R Package brms. *R J.* 10, 395–411. doi: 10.32614/RJ-2018-2017

Bürkner, P.-C. (2017). brms?: an R Package for bayesian multilevel models using stan. *J. Stat. Softw.* 80, 1–18. doi: 10.18637/jss.v080.i01

Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153. doi: 10.1016/j.tics.2009.01.005

Csibra, G., and Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1149–1157. doi: 10.1098/rstb.2010.0319

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-27

Farroni, T., Mansfield, E. M., Lai, C., and Johnson, M. H. (2003). Infants perceiving and acting on the eyes: tests of an evolutionary hypothesis. *J. Exp. Child Psychol.* 85, 199–212. doi: 10.1016/S0022-0965(03)00022-25

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

Gelman, A. (2006). Multilevel (Hierarchical) modeling: what it can and cannot do. *Technometrics* 48, 432–435. doi: 10.1198/004017005000000661

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., and Vogeley, K. (2014). The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Front. Hum. Neurosci*. 8:807. doi: 10.3389/fnhum.2014.00807

Gibson, J. J., and Pick, A. D. (1963). Perception of another person's looking behavior. *Am. J. Psychol.* 76, 386–394. doi: 10.2307/1419779

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

Gobel, M. S., Tufft, M. R. A., and Richardson, D. C. (2017). Social beliefs and visual attention: how the social relevance of a cue influences spatial orienting. *Cogn. Sci.* 42(Suppl. 1), 161–185. doi: 10.1111/cogs.12529

Gredebäck, G., Astor, K., and Fawcett, C. (2018). Gaze following is not dependent on ostensive cues: a critical test of natural pedagogy. *Child Dev.* 89, 2091–2098. doi: 10.1111/cdev.13026

Hamilton, A. F. de C. (2016). Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150080. doi: 10.1098/rstb.2015.0080

Hoffman, M. D., and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.* 31, 1593–1623.

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Jenkins, R., Beaver, J. D., and Calder, A. J. (2006). I thought you were looking at me: direction-specific aftereffects in gaze perception. *Psychol. Sci.* 17, 506–513. doi: 10.1111/j.1467-9280.2006.01736.x

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2018). The "Social Gaze Space": a taxonomy for gaze-based communication in triadic interactions. *Front. Psychol.* 9:229. doi: 10.3389/fpsyg.2018.00226

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2019). Inferring interactivity from gaze patterns during triadic person-object-agent interactions. *Front. Psychol*. 10:1913. doi: 10.3389/fpsyg.2019.01913

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychol. Bull.* 100, 78–100. doi: 10.1037//0033-2909.100.1.78

Lehrl, S. (2005). *Manual Zum MWT-B: Mehrfachwahl-Wortschatz-Intelligenztest. 5., unveränd. Aufl*. Balingen: Spitta-Verl.

Lehrl, S., Gallwitz, A., and Blaha, L. (1991). *Theorie und Messung Der Geistigen Leistungsfähigkeit Mit Dem Kurztest KAI: Die Allgemeinen Basisgrößen der Informationsverarbeitung.* Vless: . Fischer Ebersberg.

Lobmaier, J. S., Fischer, M. H., and Schwaninger, A. (2006). Objects capture perceived gaze direction. *Exp. Psychol.* 53, 117–122. doi: 10.1027/1618-3169.53.2.117

Lyyra, P., Wirth, J. H., and Hietanen, J. K. (2017). Are you looking my way? Ostracism widens the cone of gaze. *Q. J. Exp. Psychol*. 70, 1713–1721. doi: 10.1080/17470218.2016.1204327

Mareschal, I., Calder, A. J., and Clifford, C. W. G. (2013). Humans have an expectation that gaze is directed toward them. *Curr. Biol.* 23, 717–721. doi: 10.1016/j.cub.2013.03.030

McElreath, R. (2016). *Statistical Rethinking: a Bayesian course with Examples in R and Stan*. Boca Raton: CRC Press/Taylor & Francis Group.

Morales, M., Mundy, P., and Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behav. Dev*. 21, 373–377. doi: 10.1016/S0163-6383(98)90014-90015

Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. *Curr. Dir. Psychol. Sci.* 16, 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF(IRI) Zur Messung Von Empathie: Psychometrische Evaluation Der deutschen Version Des Interpersonal Reactivity Index. Available at: http://hdl.handle.net/20.500.11780/3343 (accessed December 05, 2019).

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., et al. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *NeuroImage* 101, 124–137. doi: 10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Vogeley, K., and Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev.* 37, 2516–2528. doi: 10.1016/j.neubiorev.2013.07.017

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., and Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? *An ERP study. Soc. Cogn. Affect. Neurosci.* 6, 486–494. doi: 10.1093/scan/nsq068

Qian, Q., Song, M., and Shinomori, K. (2013). Gaze cueing as a function of perceived gaze direction: gaze cueing. *Jpn. Psychol. Res.* 55, 264–272. doi: 10.1111/jpr.12001

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

R Studio Team (2016). *RStudio: Integrated Development for R.* Boston, MA: RStudio, Inc.

Risko, E. F., and Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Atten. Percept. Psychophys.* 73, 291–296. doi: 10.3758/s13414-010-0042-41

Senju, A., and Csibra, G. (2008). Gaze Following in human infants depends on communicative signals. *Curr. Biol.* 18, 668–671. doi: 10.1016/j.cub.2008.03.059

Senju, A., Csibra, G., and Johnson, M. H. (2008). Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition* 108, 303–319. doi: 10.1016/j.cognition.2008.02.009

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13, 127–134. doi: 10.1016/j.tics.2008.11.009

Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Front. Integr. Neurosci.* 4:5. doi: 10.3389/fnint.2010.00005

Stan Development Team (2018). *RStan: the R interface to Stan. R package version 2.14.1*. Available at: http://mc-stan.org/ (accessed December 05, 2019).

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., and Gredebäck, G. (2014). Is ostension any more than attention? *Sci. Rep.* 4:5304. doi: 10.1038/srep05304

Thainimit, S., Alexandre, L. A., and de Almeida, V. M. N. (2013). "Iris surface deformation and normalization," in *2013 13th International Symposium on Communications and Information Technologies (ISCIT)*, (Surat Thani: IEEE), 501–506.

Vernetti, A., Smith, T. J., and Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proc. R. Soc. B Biol. Sci.* 284, 20162747. doi: 10.1098/rspb.2016.2747

Vogeley, K., and Bente, G. (2010). "Artificial humans": psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Netw.* 23, 1077–1090. doi: 10.1016/j.neunet.2010.06.003

von Griinau, M., and Anston, C. (1995). The detection of gaze direction: a stare-in-the-crowd effect. *Perception* 24, 1297–1313. doi: 10.1068/p241297

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., and Bara, B. G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J. Cogn. Neurosci.* 16, 1854–1863. doi: 10.1162/0898929042947838

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

frontiers in Psychology

PERSPECTIVE
published: 26 February 2018
doi: 10.3389/fpsyg.2018.00226

The "Social Gaze Space": A Taxonomy for Gaze-Based Communication in Triadic Interactions

Mathis Jording[1], Arne Hartz[2,3], Gary Bente[4], Martin Schulte-Rüther[2,3,5] and Kai Vogeley[1,5]*

[1] *Department of Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany,* [2] *JARA-BRAIN, Aachen, Germany,* [3] *Translational Brain Research in Psychiatry and Neurology, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, University Hospital RWTH Aachen, Aachen, Germany,* [4] *Department of Communication, Michigan State University, East Lansing, MI, United States,* [5] *Cognitive Neuroscience (INM-3), Institute of Neuroscience and Medicine, Research Center Jülich, Jülich, Germany*

Humans substantially rely on non-verbal cues in their communication and interaction with others. The eyes represent a "simultaneous input-output device": While we observe others and obtain information about their mental states (including feelings, thoughts, and intentions-to-act), our gaze simultaneously provides information about our own attention and inner experiences. This substantiates its pivotal role for the coordination of communication. The communicative and coordinative capacities – and their phylogenetic and ontogenetic impacts – become fully apparent in triadic interactions constituted in its simplest form by two persons and an object. Technological advances have sparked renewed interest in social gaze and provide new methodological approaches. Here we introduce the 'Social Gaze Space' as a new conceptual framework for the systematic study of gaze behavior during social information processing. It covers all possible categorical states, namely 'partner-oriented,' 'object-oriented,' 'introspective,' 'initiating joint attention,' and 'responding joint attention.' Different combinations of these states explain several interpersonal phenomena. We argue that this taxonomy distinguishes the most relevant interactional states along their distinctive features, and will showcase the implications for prominent social gaze phenomena. The taxonomy allows to identify research desiderates that have been neglected so far. We argue for a systematic investigation of these phenomena and discuss some related methodological issues.

Keywords: non-verbal communication, social gaze, joint attention, triadic interaction, ecological validity, taxonomy, social psychology

Edited by:
Mario Weick,
University of Kent, United Kingdom

Reviewed by:
Matthias S. Gobel,
University College London, United Kingdom
Cade Andrew McCall,
University of York, United Kingdom

****Correspondence:***
Mathis Jording
mathis.jording@uk-koeln.de

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *06 October 2017*
Accepted: *09 February 2018*
Published: *26 February 2018*

Citation:
Jording M, Hartz A, Bente G, Schulte-Rüther M and Vogeley K (2018) The "Social Gaze Space": A Taxonomy for Gaze-Based Communication in Triadic Interactions. Front. Psychol. 9:226. doi: 10.3389/fpsyg.2018.00226

SOCIAL GAZE AS SPECIAL CASE OF NON-VERBAL COMMUNICATION

Non-verbal communication does not only supplement verbal utterances but constitutes a crucial part of communication in itself. Thereby, non-verbal communication must not be treated as a series of isolated and discrete signals but as a complex and dynamic process (Burgoon et al., 1989, p. 23). In addition, the production and perception of non-verbal communication behavior are often implicit and automatic (Choi et al., 2005) – i.e., unintentional, uncontrollable processes humans are unaware of (Bargh, 1994).

Among the non-verbal cues, gaze behavior plays a pivotal role. The eyes are among the first and most frequently fixated regions in humans (Yarbus, 1967; Walker-Smith et al., 1977) from early infancy on (Haith et al., 1977), serve face and emotion recognition, and allow to identify gender, age, and personality (George and Conty, 2008; Itier and Batty, 2009).

The morphology of the human eye with its white sclera significantly enhances the visibility of the eyes and facilitates gaze recognition (Kobayashi and Kohshima, 1997, 2001), suggesting evolutionary adaptation to the increased importance of gaze-based social interaction and, eventually, social cognition in humans (Emery, 2000). Ontogenetically, attending to gaze can be considered a precursor of cooperation in young children (Tomasello et al., 2007). Both phylogenetically and ontogenetically (Grossmann, 2017) social gaze opens a "window into social cognition" (Shepherd, 2010).

In addition to coordination and management of verbal conversation (Argyle and Cook, 1976), gaze mutually coordinates attention which is a hallmark of social learning, communication, social interaction, and, finally, shared intentionality (Tomasello et al., 2007) and joint action (Sebanz and Knoblich, 2009). So-called joint attention (JA) is typically defined in the gaze domain: In triadic interactions (e.g., Lee et al., 1998), two persons can jointly attend to an object by one person following another person's gaze toward a given object or possibly a third person. JA is the basis and prerequisite of cooperation (Tomasello et al., 2007) and has been investigated in great detail (Kleinke, 1986; Emery, 2000; Frischen et al., 2007; George and Conty, 2008; Itier and Batty, 2009; Shepherd, 2010; Falck-Ytter and von Hofsten, 2011; Pfeiffer et al., 2013b; Oberwelland et al., 2016; Grossmann, 2017).

THE "SOCIAL GAZE SPACE" (SGS)

Despite the wealth of social gaze research, a unifying taxonomy of social gaze is still lacking. For the most commonly used taxonomy Emery (2000) summarized several core processes like averted gaze, mutual gaze, gaze following and JA under the term social gaze. However, this taxonomy has two major limitations: (1) the basic processes described by Emery were not considered as extended in time. Relatedly, transitions between states have not been taken into account. The taxonomy of Emery therefore lacks the complex and dynamic character of gaze encounters between two persons, which are extended in time and are based on the continuous exchange between the interactants. (2) An additional restriction of the traditional social gaze terminology and research is that they focus on explicit interactions in which at least one person deliberately tries to interact with or respond to another (Schilbach et al., 2010; Pfeiffer et al., 2014). However, already the mere presence of another person presumably strongly affects a persons' behavior even when the partner is not interactively engaged. Recent research about the dual function of social gaze demonstrates that the awareness of someone else watching oneself can change the own gaze behavior (Gobel et al., 2015; Jarick and Kingstone, 2015). In accordance with recent interactionist advances emphasizing the dynamical character of interactions and arguing for ecological validity (Risko et al., 2012, 2016; Pfeiffer et al., 2013a; Schilbach et al., 2013), it is therefore important to consider all possible states of triadic interactions in a holistic approach.

In the following, we propose a taxonomy of the "Social Gaze Space" (SGS) that comprises all internal states a person can possibly adopt in the most basic setup of a gaze-based triadic interaction, as constituted by two interaction partners and an object[1]. These states are: partner-oriented (PO), object-oriented (OO), introspective (INT), responding joint attention (RJA), and initiating joint attention (IJA). We define these states on the basis of the behavior of one interactant (**Figure 1**). A dynamic interaction involving two persons can be conceptualized as a combination of two out of five different states which need not necessarily be temporally aligned. All combinations of

[1] Although, our taxonomy explicitly comprises a set of discrete states, we use the term "Social Gaze Space" throughout the manuscript instead of the more precise term "Social Gaze State Space," for the sake of simplicity and comprehensibility.

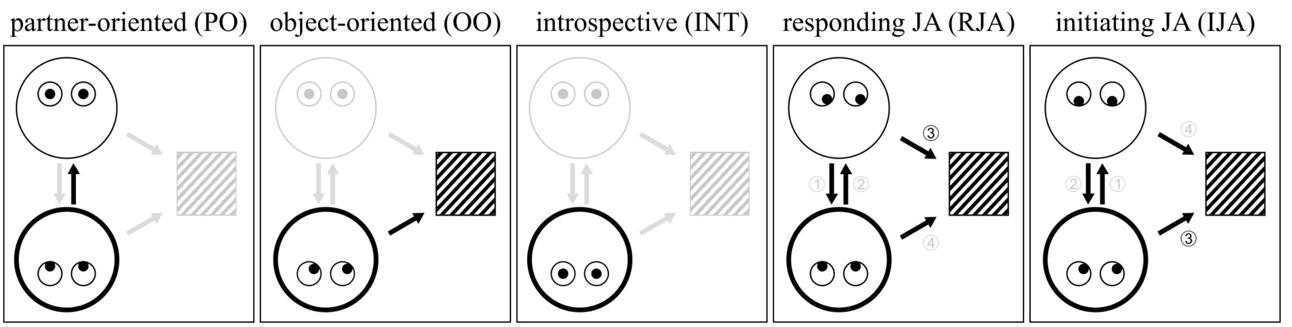


FIGURE 1 | Illustration of the five interactional states of the SGS (illustration in alignment with Emery, 2000) from the perspective of index person A (always the bold face at the bottom) in interaction with person B. (1) Partner-oriented: Attentional focus of A is directed toward B without deliberate attempts to interact of any of the two interactants. (2) Object-oriented: Attentional focus of A is directed toward an object within the shared environment. (3) Introspective: The attention is directed toward A's own inner experience. (4) Responding JA: A follows B's gaze toward an object. (5) Initiating JA: A tries to shift B's attention toward an object.

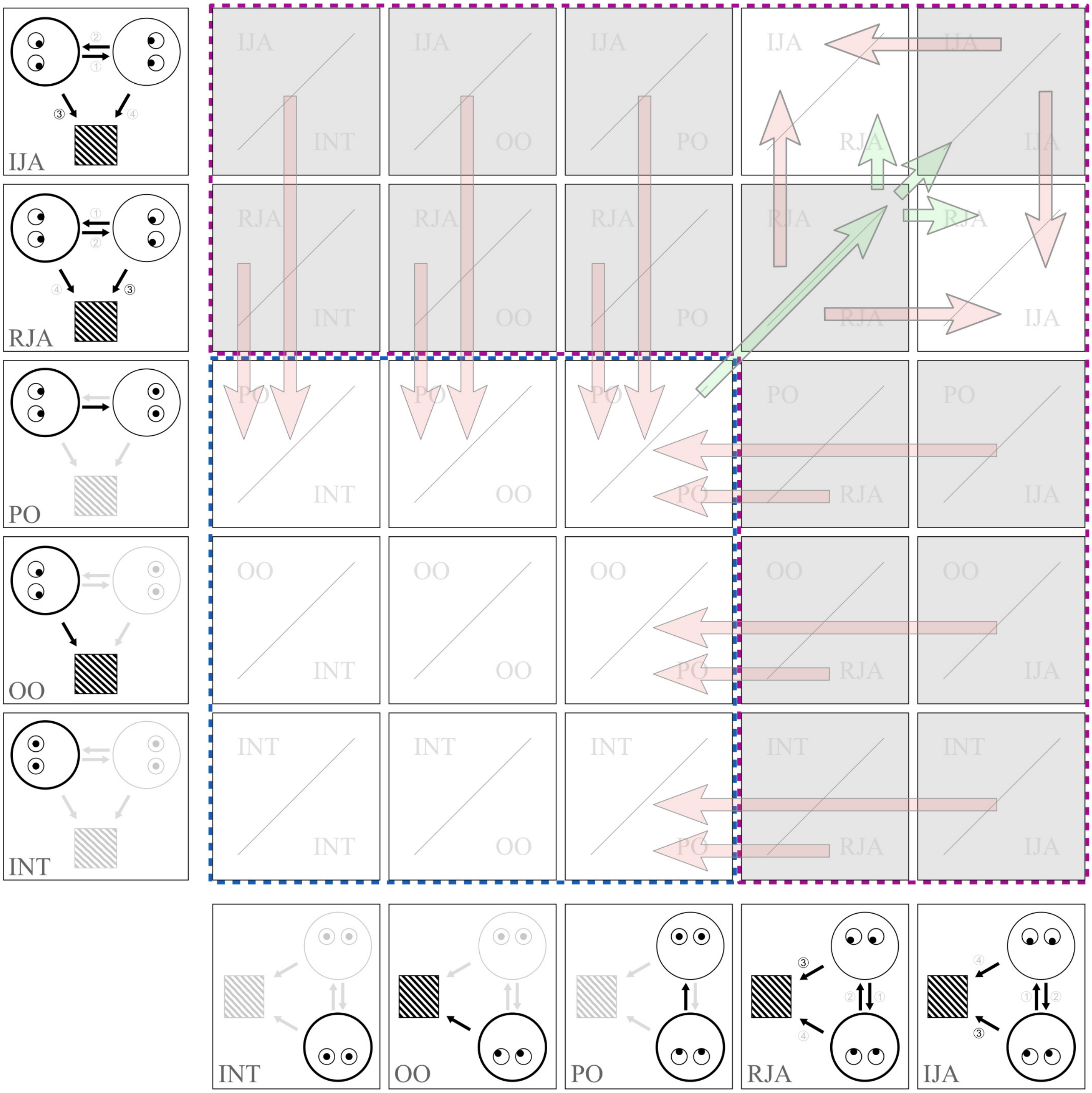


FIGURE 2 | All possible dual gaze states as a result of the combinations of gaze states of the two interactants (*x*- and *y*-axis). For illustration purposes states are presented in different order than previously introduced and as compared to **Figure 1**. Cell color indicates compatibility and stability of the states with white denoting compatible/stable states and gray denoting incompatible/unstable states. Red arrows suggest transitions from unstable to stable states. Green arrows exemplify the establishment of an interaction with a state of mutual interest serving as origin or gate (Note that arrows are not exhaustive of all possible transitions). The blue box (blue dashed line) designates states which methodologically can be inferred from a separate analysis of each participant. The purple box (purple dashed line) designates states which can only be inferred by an analysis of dynamics and interdependencies between the interactants.

states are possible and generate different types of interactional encounters that can be represented as a two-dimensional series of social gaze states evolving in time (**Figure 2**). This particularly applies to the interactive states of RJA and IJA, in which a person attempts to engage another person in an interaction which can be successful or not (see below section Triadic Interaction as a Dynamic Function of a Two-Dimensional State-Space). For this conceptualization, our focus lies on overt visual attention as deducible from gaze direction, whereas covert attention and other correlates of attention (e.g., pupil diameter, eye convergence, blinking rate) will be discussed only marginally.

THE FIVE GAZE STATES

Partner-Oriented (PO)

In the partner-oriented state person A focuses her attention on person B. The eyes automatically attract visual attention (Laidlaw et al., 2012) and possibly convey information about personal attributes including gender, age and identity (Schyns et al., 2002), as well as emotional and attentional states (Baron-Cohen et al., 1997; Emery, 2000).

Eyes that focus on the viewer will be preferentially looked at (Senju and Hasegawa, 2005) or evaluated much more positively (Stass and Willis, 1967), modulate attention (Senju and Hasegawa, 2005; Dalmaso et al., 2017), increase emotional empathy (Schulte-Rüther et al., 2007) and modulate cognition suggesting a substantial 'eye-contact-effect' for diverse aspects of socio-emotional perception (Senju and Johnson, 2009). Among distractor stimuli, viewer-directed gaze is detected easily and much faster than averted gaze (von Grünau and Anston, 1995; Conty et al., 2006; Senju et al., 2008). Profound effects of viewer-direct eye gaze on preference (Hains and Muir, 1996) and attentional modulation (Farroni et al., 2002) have also been demonstrated in infants. This is probably the most thoroughly studied gaze state.

The effect of eye contact is much stronger during dynamic interactions with real persons than when confronted with static pictures (Hietanen et al., 2008; Pönkänen et al., 2011). This requires interactive approaches with dynamic face-to-face interactions (Pfeiffer et al., 2012, 2013a; Risko et al., 2012, 2016; Schilbach et al., 2013; Schilbach, 2014; Oberwelland et al., 2016).

Object-Oriented (OO)

In the object-oriented state person A's attention is focused more or less entirely on an object in the shared environment, but not on the other person (as opposed to joint attention states described below during which person A oscillates between objects and person B). That is B's presence and behavior are likely to influence A to some level but merely coincidentally and probably without A's awareness. The exploration of different objects in a visual scenery is affected by the saliency of objects and thus the probability of persons directing their attention toward the objects (Itti and Koch, 2000). However, top-down as well as bottom-up processes are actively working together or compete for attention (Egeth and Yantis, 1997). Again, our attention and behavior toward objects are altered by actions or even the mere presence of another person looking at us (Senju and Hasegawa, 2005). Gaze cueing can automatically lead the attention toward particular objects (Frischen et al., 2007), even overriding the effect of higher psychophysical saliency (Borji et al., 2014). This brief instance of social interaction might induce a lasting attentional shift from a state of OO to the state of RJA [as examined in section Responding Joint Attention (RJA)]. However, even in the absence of any active gaze cuing, the presence of another person can attract covert attention (Kuhn et al., 2016; Laidlaw et al., 2016). Furthermore, the mere knowledge of the possibility of someone else watching their gaze lets participants control their gaze behavior with respect to its social adequacy (Risko and Kingstone, 2011).

Introspective (INT)

In this state person A neither focuses on objects nor on persons in the environment but only on his inner experience. Attentional disengagement from the outside world has been shown to correlate with a decrease in saccade frequency and an increase in saccade amplitude (Benedek et al., 2017) and, accordingly, a decrease in fixation frequency and an increase in fixation duration (Reichle et al., 2010; Benedek et al., 2017). Furthermore, in these situations blinking rate can increase (Smilek et al., 2010) and blinking duration can be prolonged (Salvi et al., 2015; Benedek et al., 2017). INT seems to show more variability in pupil diameter than episodes of directed attention to outward stimuli (Smallwood et al., 2011; Benedek et al., 2017). A higher variability of eye vergence (Benedek et al., 2017) suggests a less focused gaze (Solé Puig et al., 2013).

While it is intuitively obvious that these changes are indicative of a reduced responsiveness to events in the outside world (Smallwood et al., 2011; Benedek et al., 2017), it is an open question whether the reduced responsiveness to external stimuli and the overall change in gaze behavior are both the result and an epiphenomenon of INT, or whether changes such as a decrease in the frequency of microsaccades during INT may represent active visual disengagement as a strategy to achieve reduced responsiveness (Benedek et al., 2017). Another strategy participants adopt in situations of high cognitive load is to avoid looking into the eyes of an observer because this would entail higher demands on cognitive processing (Glenberg et al., 1998; Doherty-Sneddon and Phelps, 2005; Phelps et al., 2006; Markson and Paterson, 2009). Interestingly, the additional cognitive demands of mutual gaze do not seem to originate in the physical properties of the stimulus (e.g., the eyes) but in the interactive character inherent in this situation (Markson and Paterson, 2009). It is therefore crucial to consider introspective attentional states as potentially socially influenced by the presence of another person.

Responding Joint Attention (RJA)

In the responding JA state person A waits for B to initiate and lead the interaction, e.g., B chooses an object and A follows B's gaze toward the object. Gaze following reactions that respond to the invitation of another person thereby establishing a rudimentary form of JA appear to be deeply rooted in human behavior (Pfeiffer et al., 2011). The gaze of another person automatically cues one's own attention even when it is uninformative (Friesen and Kingstone, 1998), and participants exhibit gaze following even for forthright counter-predictive gaze cues (Driver et al., 1999; Bayliss and Tipper, 2006).

Gaze following with the aim of establishing JA constitutes a very simple though effective mechanism allowing for the inference of the attentional focus of other persons. The ability to adopt the attentional focus of another person is a prerequisite for reinforcement learning, from infants to adults (Vernetti et al., 2017). Infants at 6 months of age are already able to follow the eyes of other persons, in particular in a communicative context (Senju and Csibra, 2008). Accordingly, early proficiency in gaze following in infants predicts the development of mentalizing and emergence of language (Morales et al., 1998; Charman et al., 2000). JA and gaze following facilitate social learning, social competence, self-regulation, intelligence, and depth of information processing (Mundy and Newell, 2007).

Initiating Joint Attention (IJA)

In this state, person A takes the lead within the interaction by initiating JA. While gaze following in RJA reflects person A's understanding that B's perception and actions are goal-directed or have communicative intent, the initiation of JA is considered to require elaborate processing and insight (Tomasello and Carpenter, 2005). To initiate JA, A has to acknowledge (1) the dual function of social gaze (Gobel et al., 2015; Jarick and Kingstone, 2015) i.e., that gaze does not only serves her in perceiving but also that her gaze informs B about her focus of attention and, (2) sharing of attention is a desirable aim for mutual interaction (Tomasello et al., 2005). Whereas first elements of RJA are already evident at 6 months of age, IJA does not emerge before the second year of life (Mundy and Newell, 2007; Mundy et al., 2007). Chimpanzees followed the experimenters gaze on a frequent basis but did not try to initiate JA (Tomasello and Carpenter, 2005). Interestingly, differential development of both RJA and IJA can be observed in brain systems from childhood to adulthood (Oberwelland et al., 2016), as well as during atypical development in disorders such as autism (Oberwelland et al., 2017). In autism, IJA is typically more impaired than RJA and emerges much later than in typical development (Mundy, 2003). These empirical findings clearly point toward separate underlying cognitive systems of RJA and IJA (Mundy and Newell, 2007).

The innate tendency to expect other humans to follow their gaze (Pfeiffer et al., 2011) corresponds to the perception of successful initiation of JA as rewarding (Schilbach et al., 2010; Pfeiffer et al., 2014; Oberwelland et al., 2016). A successfully initiated instance of JA alters the consecutive interaction by increasing the tendency to look at and dwell upon the partners face (Bayliss et al., 2013).

Triadic Interaction as a Dynamic Function of a Two-Dimensional State-Space

Having defined the basic states during triadic JA, the picture becomes more complex when considering that each of the two participants can adhere to any of these states during a triadic interaction unfolding in time. In theory, a dual social state may be one of 25 possible combinations (representing varying degrees of "interactivity"), spanning a two dimensional SGS (**Figure 2**; see McCall and Singer, 2015 for an alternative concept of a 2D gaze space). Some of these combinations might be more ephemeral than others: e.g., a person A might soon lose the motivation to initiate JA if person B does not respond to him adequately, person A might switch to PO very soon subsequently ('stability' of states is indicated by cell color in **Figure 2**, with gray cells indicating unstable and ephemeral states; red arrows represent subsequent shifts from unstable to stable states).

Furthermore, it is conceivable that mutual attention (PO/PO) might facilitate transitions from non-interactive to interactive states (indicated by green arrows in **Figure 2**). These transitions have yet to be empirically investigated. Only non-interactive states (blue box in **Figure 2**) can be understood on the basis of single persons whereas the study of interactive situations (purple box in **Figure 2**) requires a complex dynamic concept and experimental setup, based on the idea that the basic unit of analysis is the interaction between both interactants.

REFLECTIONS AND FUTURE DIRECTIONS

It is our goal to provide a unifying taxonomy of social gaze in triadic interactions and their respective interdependencies. This complex, dynamic and holistic approach has two major achievements. First, it facilitates the integration of existing empirical findings within one unifying framework and helps to identify research desiderates. Second, it will go beyond many of the previous studies that investigated gaze behavior in isolation and it will provide a theoretical background to study the complex dynamics of dual states including their transitions, thereby increasing the ecological validity of the empirical approaches.

This approach is in accordance with a growing number of proposals that argued in favor of "embedded" interactionist or "enactive" approaches and emphasize the importance of ecological validity in non-verbal communication and social cognition research (Kingstone, 2009; Marsh et al., 2009; De Jaegher et al., 2010; Konvalinka and Roepstorff, 2012; Risko et al., 2012, 2016; Skarratt et al., 2012; Gallagher, 2013; Pfeiffer et al., 2013b; Schilbach et al., 2013). New methodological approaches due to technological advances increasingly allow for the development of paradigms which meet those demands (Pfeiffer et al., 2013b; Oberwelland et al., 2016, 2017).

This paves the way to research questions concerning the nature of gaze communication in triadic interactions. Even in triadic encounters which are not explicitly interactive interactants are still likely to exert subtle influences on each other in many reciprocal ways: In PO, dynamic interactions elicit a much stronger eye contact effect that static pictures (Hietanen et al., 2008; Pönkänen et al., 2011); In OO, the visual attention of another person will influence object processing in an observer in multiple ways (Reid et al., 2004; Bayliss et al., 2006; Becchio et al., 2008); the oculomotor changes observable in INT might be an active form of visual disengagement (Benedek et al., 2017). Therefore, a separate examination of allegedly interactive and non-interactive states in triadic interactions is not adequate. From the new unifying perspective of the SGS the very first step must be to systematically describe and identify the characteristics of gaze behavior associated with the individual gaze states. However, given the dynamic and continuous nature of non-verbal communication (Burgoon et al., 1989) our appreciation of the interactants experience of the encounter relies on our comprehension of transitions between interactional states.

The consequential next step will then be the identification of potentially complex signifiers of these transitions in gaze behavior, yet unknown (e.g., gaze patterns characteristic for active attempts to catch the partners attention to reach a full-fledged state of JA), which can serve as indicators of these transitions in future studies.

We speculate that transitions between gaze states of the individual interactants are not independent, but are contingent upon each other to a changing degree. If these contingencies are crucial in the establishment of states of higher interactivity and phenomena like synchrony and rapport between interactants, then it should be possible to establish their causal role in experimental paradigms. The dual state of mutual attention (PO/PO) as a candidate state for a gate to higher degrees of interactivity (**Figure 2**) – as soon as its role is empirically corroborated – could be a potential starting point in these investigations.

Having established the prototypical SGS it is worth studying individual differences in the behavior and experiences in triadic gaze interactions. Questions which to the best of our knowledge have not been tackled before concern the relationship between specific personality traits and gaze behavior in triadic encounters and to which degree personality traits are ascribed on the basis of gaze behavior. Other obvious topics relate to developmental factors in the SGS and how and when children access the SGS or the effect of impairments in non-verbal communication as observable in autism have in the SGS.

AUTHOR CONTRIBUTIONS

All authors substantially contributed to the conception of the work. MJ drafted the manuscript. AH, GB, MS-R, and KV revised it critically.

FUNDING

This work is a part of the project ‘EMOTISK’ from the ‘Bringing technology to the people’ research program on human–machine interaction, funded by the German Federal Ministry of Education and Research (Grant Number for MS-R: 16SV7242, for KV: 16SV7244). MS-R received funding by the Excellence Initiative of the German federal and state governments.

ACKNOWLEDGMENTS

The authors like to thank their respective work units for their continued support and constructive discussions. The authors want to express their gratitude to Joseph Walther from Nanyang Technological University for helpful and substantial feedback at the 3rd Annual International Summer School at Michigan State University and at the University Hospital Cologne.

REFERENCES

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

Bargh, J. A. (1994). “The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition,” in *Handbook of Social Cognition: Basic Processes; Applications*, eds R. S. Wyer Jr. and T. Srull (Hillsdale, NJ: Lawrence Erlbaum), 1–40.

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger Syndrome. *Vis. Cogn.* 4, 311–331. doi: 10.1080/713756761

Bayliss, A. P., Murphy, E., Naughtin, C. K., Kritikos, A., Schilbach, L., and Becker, S. I. (2013). “Gaze leading”: initiating simulated joint attention influences eye movements and choice behavior. *J. Exp. Psychol. Gen.* 142, 76–92. doi: 10.1037/a0029286

Bayliss, A. P., Paul, M. A., Cannon, P. R., and Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychon. Bull. Rev.* 13, 1061–1066. doi: 10.3758/BF03213926

Bayliss, A. P., and Tipper, S. P. (2006). Predictive gaze cues and personality judgments: should eye trust you? *Psychol. Sci.* 17, 514–520. doi: 10.1111/j.1467-9280.2006.01737.x

Becchio, C., Bertone, C., and Castiello, U. (2008). How the gaze of others influences object processing. *Trends Cogn. Sci.* 12, 254–258. doi: 10.1016/j.tics.2008.04.005

Benedek, M., Stoiser, R., Walcher, S., and Körner, C. (2017). Eye behavior associated with internally versus externally directed cognition. *Front. Psychol.* 8:1092. doi: 10.3389/fpsyg.2017.01092

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

Burgoon, J. K., Buller, D. B., and Woodall, W. G. (1989). *Nonverbal Communication: The Unspoken Dialogue*. Manhattan, NY: Harper & Row.x

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

Choi, Y. S., Gray, H. M., and Ambady, N. (2005). “The glimpsed world: Unintended communication and unintended perception,” in *The New Unconscious*, eds R. R. Hassin, J. S. Uleman, and J. A. Bargh (New York, NY: Oxford University Press), 309–333.

Conty, L., Tijus, C., Hugueville, L., Coelho, E., and George, N. (2006). Searching for asymmetries in the detection of gaze contact versus averted gaze under different head views: a behavioural study. *Spat. Vis.* 19, 529–545. doi: 10.1163/156856806779194026

Dalmaso, M., Castelli, L., Scatturin, P., and Galfano, G. (2017). Trajectories of social vision: eye contact increases saccadic curvature. *Vis. Cogn.* 25, 358–365. doi: 10.1080/13506285.2017.1290727

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

Doherty-Sneddon, G., and Phelps, F. G. (2005). Gaze aversion: a response to cognitive or social difficulty? *Mem. Cognit.* 33, 727–733. doi: 10.3758/BF03195338

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Vis. Cogn.* 6, 509–540. doi: 10.1080/135062899394920

Egeth, H. E., and Yantis, S. (1997). Visual attention: control, representation, and time course. *Annu. Rev. Psychol.* 48, 269–297. doi: 10.1146/annurev.psych.48.1.269

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-7

Falck-Ytter, T., and von Hofsten, C. (2011). How special is social looking in ASD: a review. *Prog. Brain Res.* 189, 209–222. doi: 10.1016/B978-0-444-53884-0.00026-9

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9602–9605. doi: 10.1073/pnas.152159999

Friesen, C. K., and Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon. Bull. Rev.* 5, 490–495. doi: 10.3758/BF03208827

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

Gallagher, S. (2013). The socially extended mind. *Cogn. Syst. Res.* 2, 4–12. doi: 10.1016/j.cogsys.2013.03.008

George, N., and Conty, L. (2008). Facing the gaze of others. *Neurophysiol. Clin.* 38, 197–207. doi: 10.1016/j.neucli.2008.03.001

Glenberg, A. M., Schroeder, J. L., and Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Mem. Cognit.* 26, 651–658. doi: 10.3758/BF03211385

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

Grossmann, T. (2017). The eyes as windows into other minds: an integrative perspective. *Perspect. Psychol. Sci.* 12, 107–121. doi: 10.1177/1745691616654457

Hains, S. M. J., and Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Dev.* 67, 1940–1951. doi: 10.2307/1131602

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

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

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

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

Jarick, M., and Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Front. Psychol.* 6:1423. doi: 10.3389/fpsyg.2015.01423

Kingstone, A. (2009). Taking a real look at social attention. *Curr. Opin. Neurobiol.* 19, 52–56. doi: 10.1016/j.conb.2009.05.004

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychol. Bull.* 100, 78–100. doi: 10.1037/0033-2909.100.1.78

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

Kobayashi, H., and Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J. Hum. Evol.* 40, 419–435. doi: 10.1006/jhev.2001.0468

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci* 6:215. doi: 10.3389/fnhum.2012.00215

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

Laidlaw, K. E. W., Risko, E. F., and Kingstone, A. (2012). A new look at social attention: orienting to the eyes is not (entirely) under volitional control. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1132–1143. doi: 10.1037/a0027075

Laidlaw, K. E. W., Rothwell, A., and Kingstone, A. (2016). Camouflaged attention: covert attention is critical to social communication in natural settings. *Evol. Hum. Behav.* 37, 449–455. doi: 10.1016/j.evolhumbehav.2016.04.004

Lee, K., Eskritt, M., Symons, L. A., and Muir, D. (1998). Children's use of triadic eye gaze information for "mind reading." *Dev. Psychol.* 34, 525–539.

Markson, L., and Paterson, K. B. (2009). Effects of gaze-aversion on visual-spatial imagination. *Br. J. Psychol.* 100, 553–563. doi: 10.1348/000712608X371762

Marsh, K. L., Richardson, M. J., and Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Top. Cogn. Sci.* 1, 320–339. doi: 10.1111/j.1756-8765.2009.01022.x

McCall, C., and Singer, T. (2015). Facing off with unfair others: introducing proxemic imaging as an implicit measure of approach and avoidance during social interaction. *PLoS One* 10:e0117532. doi: 10.1371/journal.pone.0117532

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

Mundy, P. (2003). Annotation: the neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *J. Child Psychol. Psychiatry* 44, 793–809. doi: 10.1111/1469-7610.00165

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., and Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Dev.* 78, 938–954. doi: 10.1111/j.1467-8624.2007.01042.x

Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. *Curr. Dir. Psychol. Sci.* 16, 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., et al. (2016). Look into my eyes: investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *Neuroimage* 130, 248–260. doi: 10.1016/j.neuroimage.2016.02.026

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., et al. (2017). Young adolescents with autism show abnormal joint attention network: a gaze contingent fMRI study. *Neuroimage Clin.* 14, 112–121. doi: 10.1016/j.nicl.2017.01.006

Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., and Vogeley, K. (2012). Eyes on the mind: investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Front. Psychol.* 3:537. doi: 10.3389/fpsyg.2012.00537

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., et al. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *Neuroimage* 101, 124–137. doi: 10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., and Schilbach, L. (2011). A non-verbal turing test: differentiating mind from machine in gaze-based social interaction. *PLoS One* 6:e27591. doi: 10.1371/journal.pone.0027591

Pfeiffer, U. J., Timmermans, B., Vogeley, K., Frith, C. D., and Schilbach, L. (2013a). Towards a neuroscience of social interaction. *Front. Hum. Neurosci* 7:22. doi: 10.3389/fnhum.2013.00022

Pfeiffer, U. J., Vogeley, K., and Schilbach, L. (2013b). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev.* 37, 2516–2528. doi: 10.1016/j.neubiorev.2013.07.017

Phelps, F. G., Doherty-Sneddon, G., and Warnock, H. (2006). Helping children think: gaze aversion and teaching. *Br. J. Dev. Psychol.* 24, 577–588. doi: 10.1348/026151005X49872

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., and Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Soc. Cogn. Affect. Neurosci.* 6, 486–494. doi: 10.1093/scan/nsq068

Reichle, E. D., Reineberg, A. E., and Schooler, J. W. (2010). Eye movements during mindless reading. *Psychol. Sci.* 21, 1300–1310. doi: 10.1177/0956797610378686

Reid, V. M., Striano, T., Kaufman, J., and Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport* 15, 2553–2555. doi: 10.1097/00001756-200411150-00025

Risko, E. F., and Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Atten. Percept. Psychophys.* 73, 291–296. doi: 10.3758/s13414-010-0042-1

Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., and Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Front. Hum. Neurosci* 6:143. doi: 10.3389/fnhum.2012.00143

Risko, E. F., Richardson, D. C., and Kingstone, A. (2016). Breaking the fourth wall of cognitive science: real-world social attention and the dual function of gaze. *Curr. Dir. Psychol. Sci.* 25, 70–74. doi: 10.1177/0963721415617806

Salvi, C., Bricolo, E., Franconeri, S. L., Kounios, J., and Beeman, M. (2015). Sudden insight is associated with shutting out visual inputs. *Psychon. Bull. Rev.* 22, 1814–1819. doi: 10.3758/s13423-015-0845-0

Schilbach, L. (2014). On the relationship of online and offline social cognition. *Front. Hum. Neurosci.* 8:278. doi: 10.3389/fnhum.2014.00278

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22, 2702–2715. doi: 10.1162/jocn.2009.21401

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., and Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *J. Cogn. Neurosci.* 19, 1354–1372. doi: 10.1162/jocn.2007.19.8.1354

Schyns, P. G., Bonnar, L., and Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychol. Sci.* 13, 402–409. doi: 10.1111/1467-9280.00472

Sebanz, N., and Knoblich, G. (2009). Prediction in joint action: what, when, and where. *Top. Cogn. Sci.* 1, 353–367. doi: 10.1111/j.1756-8765.2009.01024.x

Senju, A., and Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Curr. Biol.* 18, 668–671. doi: 10.1016/j.cub.2008.03.059

Senju, A., and Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Vis. Cogn.* 12, 127–144. doi: 10.1080/13506280444000157

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13, 127–134. doi: 10.1016/j.tics.2008.11.009

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

Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Front. Integr. Neurosci.* 4:5. doi: 10.3389/fnint.2010.00005

Skarratt, P. A., Cole, G. G., and Kuhn, G. (2012). Visual cognition during real social interaction. *Front. Hum. Neurosci.* 6:196. doi: 10.3389/fnhum.2012.00196

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., et al. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS One* 6:e18298. doi: 10.1371/journal.pone.0018298

Smilek, D., Carriere, J. S. A., and Cheyne, J. A. (2010). Out of mind, out of sight: eye blinking as indicator and embodiment of mind wandering. *Psychol. Sci.* 21, 786–789. doi: 10.1177/0956797610368063

Solé Puig, M., Pérez Zapata, L., Aznar-Casanova, J. A., and Supèr, H. (2013). A role of eye vergence in covert attention. *PLoS One* 8:e52955. doi: 10.1371/journal.pone.0052955

Stass, J. W., and Willis, F. N. (1967). Eye contact, pupil dilation, and personal preference. *Psychon. Sci.* 7, 375–376. doi: 10.3758/BF03331131

Tomasello, M., and Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees. *Monogr. Soc. Res. Child Dev.* 70, vii-132. doi: 10.1111/j.1540-5834.2005.00324.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi: 10.1017/S0140525X05000129

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

Vernetti, A., Smith, T. J., and Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proc. R. Soc. B Biol. Sci.* 284: 20162747. doi: 10.1098/rspb.2016.2747

von Grünau, M., and Anston, C. (1995). The detection of gaze direction: a stare-in-the-crowd effect. *Perception* 24, 1297–1313. doi: 10.1068/p241297

Walker-Smith, G. J., Gale, A. G., and Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception* 6, 313–326. doi: 10.1068/p060313

Yarbus, A. L. (1967). *Eye Movements and Vision*. Boston, MA: Springer, doi: 10.1007/978-1-4899-5379-7

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

Behavior Research Methods manuscript No.
(will be inserted by the editor)

TriPy: A cross-platform toolbox for systematic investigation of gaze-based communication in triadic interactions with a virtual agent

Arne Hartz · Björn Guth · Mathis Jording · Kai Vogeley · Martin Schulte-Rüther

Received: date / Accepted: date

Abstract To navigate the social world, humans utilize nuanced reciprocal, nonverbal communication for successful social interaction and inference. Traditional paradigms in social gaze research typically rely on static images or prerecorded videos and often lack gaze-contingency. Such approaches neglect the highly complex and dynamical character of human gaze behavior during on-going social interaction.

TriPy, the Python toolbox presented in this paper, tackles this issue by creating a shared virtual environment on a computer screen for a human interactant and an algorithmically controlled virtual agent to investigate extended triadic gaze encounters (i.e. instances of joint attention between agent and human interactant and a simultaneously presented object). TriPy's implementation is characterized by a modular and flexible behavioral repertoire of the agent which is responsive to human gaze behavior and facial expressions. It can easily be extended and adapted to create customized experimental paradigms.
Such experiments allow for a highly controlled and fine-grained investigation of the precise timings and interactivity of human social gaze behavior in joint-attention settings. As reciprocal gaze is a core feature of human communication, a deeper understanding of these aspects is of great importance to research in social gaze behavior and its deviations in psychiatric conditions such as autism and for the development of algorithmically controlled robots and virtual agents designed to interact with humans.

Keywords Eye tracking · Gaze contingency · Social gaze · Triadic interaction · Joint attention · Ecological validity · human-agent interaction · PyGaze · PsychoPy · Python · Software

A. Hartz
Translational Brain Research, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Uniklinik RWTH Aachen, Germany
JARA-BRAIN, Jülich-Aachen Research Alliance, Germany
E-mail: ahartz@ukaachen.de

B. Guth
Translational Brain Research, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Uniklinik RWTH Aachen, Germany

M. Jording
Department of Psychiatry and Psychotherapy, Universitätsklinikum Köln, Germany
Cognitive Neuroscience (INM-3), Institut für Neurowissenschaften und Medizin (INM), Forschungszentrum Jülich, Germany

K. Vogeley
Department of Psychiatry and Psychotherapy, Universitätsklinikum Köln, Germany
Cognitive Neuroscience (INM-3), Institut für Neurowissenschaften und Medizin (INM), Forschungszentrum Jülich, Germany

M. Schulte-Rüther
Translational Brain Research, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Uniklinik RWTH Aachen, Germany
JARA-BRAIN, Jülich-Aachen Research Alliance, Germany
Cognitive Neuroscience (INM-3), Institut für Neurowissenschaften und Medizin (INM), Forschungszentrum Jülich, Germany
E-mail: mschulte@ukaachen.de

1 Introduction

To navigate the social world, humans utilize a nuanced reciprocal, nonverbal communication for successful social interaction (Fiske & Taylor, 2013) and inference (Moutoussis, Fearon, El-Deredy, Dolan, & Friston, 2014). A particularly important aspect of non-verbal communication and social cognition is social gaze behavior: Humans use their eyes to obtain social information about others, but also to convey aspects of their own inner mental state (Gibson & Pick, 1963).

It therefore constitutes a powerful communicative tool shaping the phylogenetic and ontogenetic development of human social cognition (Grossmann, 2017); e.g. the unique morphology of the human eye (Kobayashi & Kohshima, 1997, 2001) was suggested as an evolutionary imprint of the role of gaze in human communication (Emery, 2000) and has been described as a "window into social cognition" (Shepherd, 2010) with a long-lasting research tradition (for early, seminal examples, see (Gibson & Pick, 1963; Kendon, 1967; Yarbus, 1967; Argyle & Cook, 1976; Kleinke, 1986)).

Many of the traditional paradigms in social gaze research have been challenged for their reliance on static images or prerecorded videos as stimuli: In the light of the complex and dynamical character of nonverbal communication (Burgoon & Buller, 1989; Krämer, 2008; Vogeley & Bente, 2010), interactionist approaches call for higher ecological validity (Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012; U. Pfeiffer, Timmermans, Vogeley, Frith, & Schilbach, 2013; Schilbach et al., 2013; Risko, Richardson, & Kingstone, 2016). Of particular interest in contemporary social gaze research is the dynamic, reciprocal gaze behavior of two interacting agents during social interaction, as evident, for example during instances of joint attention (JA) (Moore, Dunham, & Dunham, 2014). JA is characterized by a shared attentional focus of two people on an object ((Emery, 2000), commonly also referred to as triadic joint attention). More recently, gaze-contingent paradigms have been developed which are suited to tackle the dynamic aspect of JA (e.g. (Wilms et al., 2010; U. J. Pfeiffer et al., 2014; Oberwelland et al., 2016, 2017), for a review see (U. J. Pfeiffer, Vogeley, & Schilbach, 2013)). However, even these gaze-contingent approaches are mostly restricted to explicitly instructed, simple units of interaction, e.g. single gaze shifts, and thus neglect the aspect of the highly reciprocal and dynamic character of real-life gaze encounters.

Such atomic units of interaction are governed by the actual higher-order mental state of the interacting agent(s) which can vary over time. Therefore, it is important to develop situation-specific taxonomies of such social interactive mental states to provide a conceptual basis for the description and investigation of observable behavior. Based on these considerations, we have recently proposed the concept of *Social Gaze States* (Jording, Hartz, Bente, Schulte-Rüther, & Vogeley, 2018) which provides a comprehensive description of the space of possible mental states during triadic encounters (see Tab. 1 for details). Starting from such a theoretical framework of behaviorally defined states, it is possible to construct prototypes of algorithmically controlled agents mimicking such states.

This paper presents TriPy, a multi-platform toolbox to create behavioral oscillations of a virtual agent during triadic encounters in a highly controlled, yet flexible, dynamic, and ecologically-valid approach. It allows for additional fine-grained, systematic, and interactive registration of other aspects of nonverbal behavior (i.e. facial action unit activity and emotional expressions via a connection to a commercial third party software). The aim of this paper is to exemplify the algorithmic implementation of Social Gaze States: It demonstrates (1) the parameterized implementation of gaze behavior during triadic JA with a virtual agent and (2) covers the empirical estimation of behavioral model parameters. This procedure allows for the creation of a virtual agent mimicking the typical gaze behavior of a real human interactant (HI) in an empirically informed fashion.

2 Method

2.1 Implementation

TriPy creates a shared virtual environment (VE) for a HI to interact with a virtual agent who is presented at the center of a computer screen surrounded by objects (Fig. 1). The agent has the ability to blink and change its gaze direction as well as facial expression, all potentially in response to the HI's non-verbal behavior expressed by gaze or facial expression.

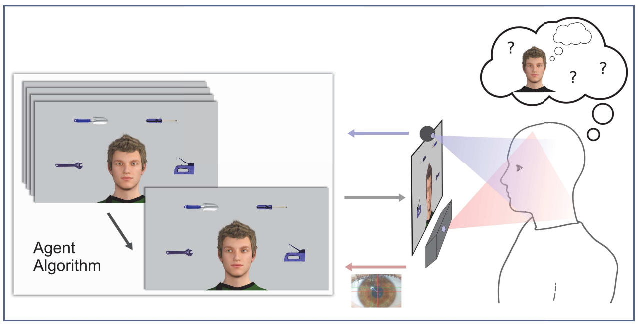


Fig. 1 Technological architecture of TriPy: Eye tracker and video camera for data acquisition, computer screen for stimulus presentation, and computer for gaze- and emotion-contingent agent algorithm, system integration, and data collection.

TriPy, as a platform for experiment design, is implemented in Python 2.7 (Python Software Foundation, https://www.python.or Compatibility with Python 3.X was ensured wherever possible and can be fully implemented, once all dependencies allow for a Python 3.X version. TriPy is based on the open-source package(s) PyGaze (Dalmaijer, Mathôt, & Van der Stigchel, 2014) acting as a wrapper for PsychoPy (Peirce, 2009) for stimulus presentation and the software development kits (SDK)s offered by the major eye tracker manufactures to keep the need for system-specific adaptions at a minimum.

To make integration of real-time emotion classification of facial expression available, a stand-alone Python client for the FaceReader API (Noldus Information Technology,

Table 1 Comprehensive description of possible behavioral (gaze) states of an agent (A) in triadic encounters from (Jording et al., 2018)

state	description
Partner-oriented (*PO*)	A has its focus of attention on the interaction partner, but not the object(s) in a shared environment (SE).
Object-oriented (*OO*)	A has its focus of attention on the object(s), but not on a possible interaction partner in a SE.
Introspective (*INT*)	A is neither focused on a possible interaction partner in a SE, nor on the objects, but in a self-referential state of mind.
Responding JA (*RJA*)	A tries to respond to JA bids on an object in a SE.
Initiating JA (*IJA*)	A tries to initiate JA on an object in a SE.

The Netherlands, tested for Versions 6.X and 7.X, available for Microsoft Windows only) is integrated into TriPy.

For automated, synchronized video recordings of the HI during the experiment, we implemented support for video capturing via μCap (Doyle & Schindler, 2015). This allows for offline analysis of video recordings of the experimental procedure, e.g. for in-depth offline facial action unit activity and emotion (Friesen & Ekman, 1978; Schulte-Rüther et al., 2017).

For interested readers, more details of hard- and software requirements, dependencies, toolbox layout, and technical reliability can be found in sections 6.3.2 to 6.3.4 of the supplement.

2.2 Agent behavior

2.2.1 *Definition of behavioral states*

TriPy contains algorithmic implementations of our concept of Social Gaze States as described in Tab. 1 and (Jording et al., 2018). At the level of the algorithmic implementation of a triadic gaze paradigm, we use the term *macro state M* to refer to these states.

Macro states We define a macro state as a cognitive socio-emotional gaze state of an agent and its associated expressed behavior. A macro state is a relatively persistent state which consists of a series of shorter Micro States (see below). The complete set of macro states $\{M_i\}$ in a given experiment spans the space of all possible cognitive-emotional states of the artificial agent in this setting and needs to be defined according to theoretical a-priori considerations[1].

Micro states The temporal dynamics (i.e. gaze shifts) of a macro state are governed by (gaze-contingent) transitions among well-defined *micro states*: Each macro state M is a set composed of n_M micro states $M = \{x_1^M, \ldots, x_{n_M}^M\}$. A single micro state x is defined by (1) its *visual appearance* a_x to the HI (i.e. gaze direction), (2) a micro state *duration* d_x described by a set of (a) random distribution(s) q_x^α with describing parameters $\Theta_x^\alpha = \{\mu_x^\alpha, \sigma_x^\alpha, \ldots\}$, (3) *transition probabilities* $\mathbf{p}_x^\alpha$ to all other micro states $x_j \in M$, and, in some cases (4), *sensitivity to socio-emotional signals* $\mathbf{S}_x$ emitted by the HI. While non-interactive micro state transitions (3) can be considered a Markovian process ($\alpha = \mathbf{p}$, (Gagniuc, 2017)), well-defined socio-emotional signals can evoke micro state transitions interactively ($\alpha = \mathbf{s}$) (4). Taken together, these rules create a (micro) state sequence which makes up the macro state and in emerging observable behavior (Fig. 2). For a detailed, formal definition, see section 6.3.5 of the supplement.

This conceptualization within TriPy allows for a broad range of possible implementations and alternations of agent behavior in triadic settings by simply creating, modifying, and (re)combining micro states.

2.2.2 *Implementation of macro states: Non Gaze-contingent*

Here, *OO*, *PO*, and *INT* macro states were defined as not gaze-contingent: All micro state transitions occur non-interactively via a Markovian process ($\mathbf{p}_x^\mathbf{p}$, Fig. 2**b**). These states differ in observable macro behavior by their difference in micro state composition: E.g. in the *OO* state, transition probabilities $p_x^i \in \mathbf{p}_x^\mathbf{p}$ for those micro states with gaze directed at an object are much higher than those in the *PO* state. This creates the impression of an agent either mostly paying attention to the objects (*OO*) or a potential interaction partner (*PO*). For the algorithmic state diagram, see Fig. S1, for formal micro state definitions, see Tab. S1.

2.2.3 *Implementation of macro states: Gaze-contingent*

Gaze-contingent macro states need to contain micro states that are responsive to gaze cues, one type of socio-emotional signal $\mathbf{S}_x$ that can be emitted by the HI. In TriPy, two types of gaze-contingent macro states are implemented:

Responsive joint attention macro state (RJA) *RJA* is characterized by the agent following the HIs gaze, i. e. when the HI fixates any object in the VE or looks straight at the HI when the HI gazes straight at the agent (thus establishing eye-contact), with probability p_f^{RJA} and a temporal delay of $d_f \sim f_\mathbf{s}(\tau_f, s_f^i)$. Then, it will keep fixating the AOI for a

[1] When trivial or generalizable, indices in the notation will be sometimes omitted to keep information to the reader as readable and compact as possible.

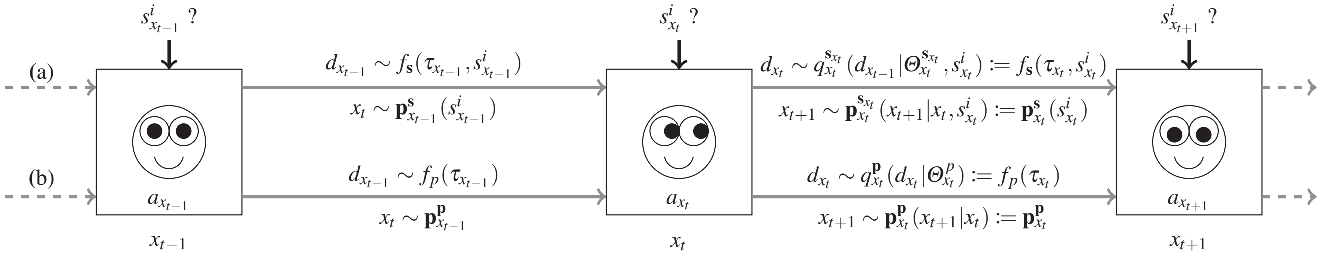


Fig. 2 Snapshot from algorithmically generated micro state chain of an agent changing its gaze direction. Micro state transition can occur either **(a)** interactively in response to a HI's socio-emotional signal $s^i_x \in \mathbf{S}_x$ to which micro state x is sensitive (e.g. fixation on AOI, target facial expression classification) or **(b)** non-interactively via a Markovian process. $d_x \sim q^{\alpha}_x(d_x|\Theta_x(,S^i_x))$ denotes that micro state duration d_x is sampled from a state-defining random distribution q_x with a corresponding set of parameters Θ_x, for interactive ($\alpha = \mathbf{s}$) and passive state transitions ($\alpha = \mathbf{p}$) respectively. The short notation $q^{\alpha}_x(d_x|\Theta_x(,s^i_x)) := f_{\alpha}(\tau_x(,s^i_x))$ denotes that the duration d_x is a function of the set $\tau_x = \{q_x, \Theta_x\}$ of the distribution q_x and corresponding parameters $\Theta_x = \{\mu_x, \sigma_x, \dots\}$, and in the interactive case, also of the detected socio-emotional signal s^i_x.

duration of $d_o \sim f_{\mathbf{p}}(\tau_o)$ before being ready to follow again. (State diagram: Fig. S2, micro state definitions: Tab. S2.)

Initiating joint attention macro state (IJA) *IJA* is characterized by the agent attempting to initiate JA with the HI on an object within the SE. First, the agent gazes straight waiting for eye contacts or a maximum duration of $d_s \sim f_{\mathbf{p}}(\tau_s)$ before shifting its gaze towards object i with probability p_i ($\sum p_i = 1$). Next, if the gaze shift was followed by the HI on object i, after a delay of $d^{s^i}_f \sim f_{\mathbf{s}}(\tau^{\mathbf{s}}_f, s^i_f)$, or, if no JA occurred, after a delay of $d^p_f \sim f_{\mathbf{p}}(\tau^{\mathbf{p}}_f)$, the agent gazes straight back to the HI trying to reestablish eye contact to commence another JA bid. (State diagram: S3, micro state definitions: S3.)

2.2.4 Cross-state behavior

Within TriPy, it is possible to implement aspects of behavior that are not dependent on the specific macro states, but consistently displayed during face-to-face interactions ("superimposed"). As a prototype, eye blink behavior is implemented: It is generated by a short transition to a micro state with closed eyes of duration $d_c = 100\,\text{ms}$, and transitioning back to the previous micro state. The time between two blinks (inter-blink interval (IBI)) sampled analog to the micro state duration above: $d_{ibi} \sim f_{\mathbf{p}}(\tau_{ibi})$. Please note, that these parameters can be implemented state dependent as well.

2.3 Empirically informed behavioral parameters

Probabilistic distributions $\{q_x\}$ and parameters $\{\Theta_x, \mathbf{p}_x\}$ of the micro states must be defined a-priori when creating an agent for experiments within TriPy. However, to ensure high ecological validity, these need to be determined empirically, i. e. based on gaze behavior of real HIs. We tackled this dilemma with an iterative approach: In a 0^{th} approximation, all distributions were defined as Gaussians and numerical parameter values were chosen based on apriori knowledge ((U. Pfeiffer et al., 2012; Oberwelland et al., 2016, 2017; Willemse, Marchesi, & Wykowska, 2018)) and refined using a face-validity strategy such that the desired behavior of the agents appeared "natural" based on intuitive judgments (Tab. S7 - S10). These values defined the agents that were used in our empirical investigation of HI behavior.

2.3.1 Paradigm

Thirty-seven (four excluded, see 6.3.7) adult participants were asked to interact with the algorithmically controlled agent in 60 blocks of ~30 s each. The experiment was divided into two parts with a short break in between to give HIs a chance to relax and prevent drifts in eye tracking measurements by recalibration. For details on the recruitment procedure, see section 6.3.6 of the supplement.

Before each block, HIs were instructed via screen messages (Tab. S4) to explicitly engage in one of the five gaze states (Tab. 1, (Jording et al., 2018)). Block order was assigned randomly, evenly balanced across the course of the experiment (Tab. S6).

When HIs were asked to engage in the non-interactive states $M \in \{PO, OO, INT\}$, the agent either gazed directly at the HI most of the time (similar to the *PO* state) or mostly averted eye contact (similar to the *INT* state) with all state transitions occurring only non-interactive. When the HI was instructed to follow the agents gaze (*RJA*) the agent was always in the *IJA* state. When the HI was asked to initiate interactions (*IJA*), the agent was put either in *RJA*1 state always following towards the object ($p^{RJA1}_f = 1.$), or *RJA*2 state, only following with probability $p^{RJA2}_f = .33$. For each block, four images of household items (adapted from (Bayliss, Paul, Cannon, & Tipper, 2006)) were randomly chosen , and displayed at given screen positions (Fig. 1 and S4).

2.3.2 Behavioral parameter estimation

This section focuses on the process of estimating the behavioral parameters described in section 2.2.1, details on data exclusion and preprocessing can be found in sections 6.3.7 and 6.3.8 of the supplement.

Temporal parameters To extract trial-by-trial values, we defined reaction times (RT) and dwell times (DT):

RTs: Time between the onset of an agent's micro state x (i.e. gaze shift on object i) and the response of the HI s_x^i (i.e. fixation on object i) in well-defined reciprocal behavioral sequences of the HI/agent interaction as depicted in Fig. 3. DTs: Time between the onset of the first and the end of the last of all consecutive fixations on an AOI by the HI to be robust against micro saccades within AOIs.

For each HI and respective temporal parameter, the respective parameters of probabilistic distributions (Normal (norm), Log-normal (lnorm) (Limpert, Stahel, & Abbt, 2001), exponentially modified Normal distribution (ExGauss) (Ratcliff, 1979)) were estimated[2]. We selected the best fitting distribution q via the Bayesian information criterion (BIC) and, for this distribution, averaged distribution parameters Θ across HIs.

Micro state transition probabilities Empirical micro state transition probabilities $\mathbf{p}_{x_i^M}$ away from micro state i for non-interactive macro states $\in M$ are defined as the fraction of the amount trials with saccades from AOI i to j over all saccades away from AOI i in macro state M. Values were calculated for each HI and subsequently averaged.

2.4 Behavioral measures

On-AOI ratio is defined as the ratio of the time of the HI dwelling on any of the defined AOIs in macro state M over to the total state duration as an indicator for the attention within the VE:

$$r_{aoi}^M = \frac{\sum DT^M}{duration\ state\ M} \leq 1 \qquad (1)$$

Face-object ratio is defined as the ratio of the time the HI spent dwelling on the agent compared to the objects per macro state M as an indicator for the distribution of attention within the VE:

$$r_{a/o}^M = \frac{\sum DT_a^M}{\sum DT_a^M + \sum DT_o^M} \leq 1 \qquad (2)$$

[2] As implemented in `fitdist()` in the R package fitdistrplus package v1.0.9 (Delignette-Muller et al., 2017) via maximum likelihood criterion using (Plummer, 2018) to obtain start values when necessary.

3 Results

3.1 Successful induction of Social Gaze States

State dependency of gaze behavior Fixation heat maps for the *OO* and *PO* states combined for all HIs demonstrate a qualitative difference in HI's gaze behavior (Fig. 4). A repeated measures One-way within subject ANOVA revealed that the fraction of time spent dwelling on the agent's face compared to the four objects s

HIs spent most of the time during the experiment either fixating one of the four objects or the agent $\langle r_{aoi} \rangle_{HI} = 0.84 \pm .07$ (SD) (Fig. S6), indicating that HIs focused their attention mostly within the VE created by TriPy.

The agent made attempts to initiate JA with a frequency of 18 ± 4 per minute, which is equivalent to 112 trials per HI on average. HIs responded with a gaze shift towards the same object on average in $85 \pm 20\%$ of the cases. With about the same frequency (15 ± 7 per minute), the HIs made attempts to initiate JA when asked to lead the gaze, which is equivalent to 96 trials per HI.

Overall, these results indicate that HIs were compliant in following the instructions and that social gaze states could reliably be induced and measured.

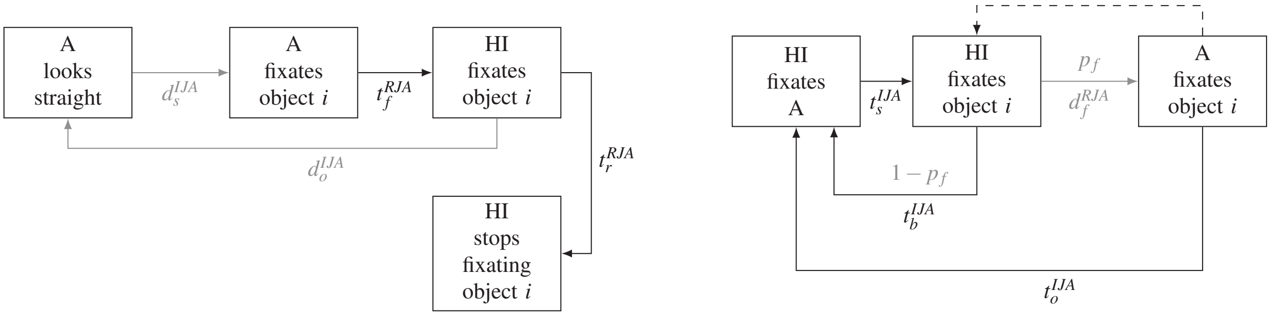
3.2 Empirical behavioral parameters

Temporal parameters Empirical mean values and standard deviation of all defined RTs and DTs, as well as their corresponding estimated temporal parameter sets (distribution and defining parameters) are shown in Tab. 2.

Probabilistic parameters Empirical transition matrices for exemplary defined AOIs (one for each object, one for the agent, see Fig. S4) for the *OO* and *PO* states are presented in Tab. 3 and 4. Higher transition probabilities towards objects AOIs compared to the facial AOI in the *OO* state are compatible with the higher proportion of object DT. Interestingly, object AOIs that were displayed in proximity have higher transition probabilities, suggesting mostly sequential exploring of adjacent objects. Please note, that for gaze-contingency and data analysis, arbitrary chosen AOIs (e.g. including finer AOIs within the face) can be defined.

4 Discussion

TriPy provides a new framework for a highly controlled fine-grained investigation of human social gaze behavior in a triadic setting. It provides the possibility to incorporate other socio-emotional behavioral modalities and is of particular interest for quantitative research related to non-verbal social



(a) HI: *RJA* A: *IJA* **(b)** HI: *IJA* A: *RJA*

Fig. 3 Sequences of behavioral events for temporal behavioral parameter estimation during the interaction between the HI and the algorithmically controlled agent for **(a)** HI in *RJA* and agent in *IJA* state and **(b)** vice versa. Times t_i denote RTs/DTs of a HI and are extracted on a single trial basis for parameter estimation. Note, that **(b)** allows for two possible variants of behavior after gazing at an object, namely, fixating the agent again before trying to initiate JA on new object (t_r^{IJA}), or directly fixating the next object without gazing back to the agent (dashed line). Algorithmically determined micro state durations are depicted by ds

Table 2 Estimated temporal behavioral parameter sets τ_i^M from RTs/DTs t_i for behavioral sequences in 3 using one AOI per object and one for the agent (S4). $\langle \cdot \rangle_{HI}$ denotes the mean across HIs, $E[\cdot]$ the mean/expectancy value and $sd[\cdot]$ the standard deviation. q_i denotes best fitting distribution via BIC criterion and $\Theta_i = \{\mu_i, \sigma_i(, \eta_i)\}$ the corresponding parameters .

M	set τ_i	description	$\langle E[d_i] \rangle_{HI}$	$\langle sd[d_i] \rangle_{HI}$	q_i	$\langle \mu_i \rangle_{HI}$	$\langle \sigma_i \rangle_{HI}$	$\langle \eta_i \rangle_{HI}$	$\langle E[q_i\|\Theta_i] \rangle_{HI}$	$\langle sd[q_x\|\Theta_i] \rangle_{HI}$
IJA	τ_s^{IJA}	DT on partner before trying to iniate JA	985	584	lnorm	6.72	0.46	-	979	580
	τ_o^{IJA}	RT after successful JA before gazing back at partner again	888	409	lnorm	6.50	0.58	-	917	566
	τ_b^{IJA}	DT on object when trying to initiate JA, but partner is not following	1590	665	lnorm	7.26	0.44	-	1540	702
RJA	τ_f^{RJA}	RT after which agent follows partner on AOI	463	133	exGAUS	350	34.20	124	474	130
	τ_o^{RJA}	DT on object after JA was successfully initated	995	290	lnorm	6.84	0.27	-	986	294
PO	τ_o^{PO}	DT on objects	435	280	lnorm	5.92	0.53	-	482	342
	τ_a^{PO}	DT on face	1810	1600	norm	2480	1670	-	1860	1640
OO	τ_o^{OO}	DT on objects	1430	1140	exGAUS	566	358	824	1390	968
	τ_a^{OO}	DT on face	660	612	lnorm	6.27	0.67	-	662	756
INT	τ_o^{INT}	DT on objects	687	628	lnorm	6.04	0.59	-	710	665
	τ_a^{INT}	DT on face	1870	1690	lnorm	7.27	1.04	-	2060	3030

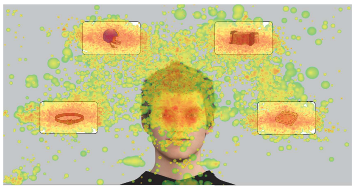
Table 3 Estimated transition probabilities among the objects (O_i) and the agent (A) for the OO state. Each row corresponds to a set $\mathbf{p}^{\mathbf{p}}_{x_i^{OO}}$ of transition probabilities away from micro state x_i as described in section 2.2.1.

aoi	O_1	O_2	O_3	O_4	A
O_1 (left)	-	0.65	0.08	0.09	0.15
O_2 (up left)	0.38	-	0.50	0.04	0.08
O_3 (right)	0.05	0.45	-	0.43	0.08
O_4 (up right)	0.15	0.10	0.55	-	0.23
A (straight)	0.28	0.31	0.19	0.21	-

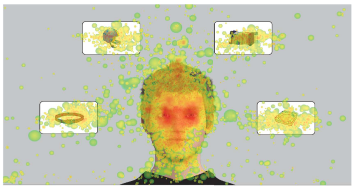
Table 4 Estimated transition probabilities for the PO state.

aoi	O_1	O_2	O_3	O_4	A
O_1 (left)	-	0.17	0.01	0.06	0.76
O_2 (up left)	0.27	-	0.42	0.02	0.29
O_3 (right)	0.00	0.33	-	0.21	0.46
O_4 (up right)	0.03	0.01	0.12	-	0.84
A (straight)	0.35	0.15	0.07	0.43	-

behavior in human face-to-face interactions. Promising applications include (1) social gaze behavior and its deviations in psychiatric conditions and (2) the facilitation of naturalistic human-agent (either virtual or robotic) interaction. In both cases, information on the exact timing in reciprocal social gaze behavior is crucial.



(a) *OO*



(b) *PO*

Fig. 4 Aggregated fixation heat maps for all HI in the **(a)** *OO* and **(b)** *PO* states. The higher fixation density on the objects AOIs in the *OO* is visible (for quantitative values see Fig. 5). The small cluster in the lower left corner in both states is the position where the stimtracker sensor was located on the screen.

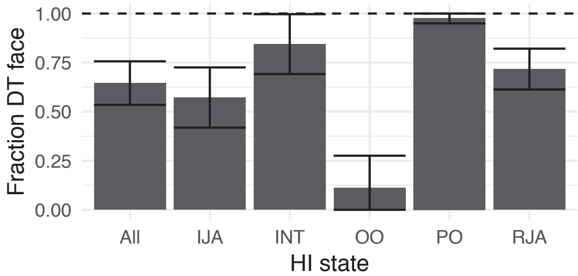


Fig. 5 Average face-object ratio $r^M_{a/o}$ across HIs for the social gaze states error bars indicating SD (capped at 1.).

4.1 Social Gaze State Parameters

To exemplify our approach, TriPy provides the algorithmic implementation of our taxonomy of Social Gaze States (Jording et al., 2018). We were able to extract probabilistic and temporal parameters of typical human gaze behavior to provide an empirical foundation for agent model parameters to be used in future studies using the concept of Social Gaze State Space and beyond.

To our knowledge, this is the first study providing a complete description of the behavioral parameter space for ongoing triadic interaction. The findings are in line with reports from the few previous studies investigating receptive and interactive gaze.

Compatible with our finding of a mean gaze following latency of 463 ms ($\langle\tau_f^{RJA}\rangle_{HI}$, see Tab. 2 and Fig. 3) when responding to a JA bid, (Caruana, McArthur, Woolgar, & Brock, 2017; Caruana et al., 2018) report median gaze following latencies of $\sim$ 430 and $\sim$ 465 ms, respectively, during a social situation (as compared to 300ms in response to non-social cues). Furthermore, a recent study of human-robot interaction (Willemse et al., 2018) demonstrated a similar latency (485ms) for responding to robot gaze cues (after sustained experiences of JA), albeit a button press was used as a proxy for a gaze cued reaction in this study. In this line, (U. Pfeiffer et al., 2012) found that a latency of 400-800 ms for gaze following was perceived as being most interactive.

For non-responded JA bids, our object DT ($\langle\tau_b^{IJA}\rangle_{HI}$ 1590ms) are shorter than those found by (Pfeiffer-Leßmann, Pfeiffer, & Wachsmuth, 2012) who reported empirical mean DTs of 1900ms. However, in this study the agent never respond to the JA bids and the HI was instructed to look at the object until he felt the agent should have responded. Since we measured this parameter in non-responsive trials during an ongoing interaction of otherwise often successful JA bids, our results likely reflect the natural behavior during continuous interaction. Accordingly, DT of 1200 and 1800 ms were most likely perceived as intentional by HIs in the study by Pfeiffer-Leßmann.

Interestingly, when initiating JA the time for a saccade back to the face after a joint fixation of an object was considerably lower in our study ($\langle\tau_f^{RJA}\rangle_{HI} = 888$ ms) than in other studies e.g. (Bayliss et al., 2013) and (Willemse et al., 2018). Bayliss et al. used an implicit gaze leading task with merely implied interaction and latencies were slightly shorter for conditions when a face followed the gaze of the participant. Similarly, (Willemse et al., 2018) found that saccade latency considerably decreased with the amount of experienced JA with a robotic agent (1500 vs. 1100 ms.). However, in contrast to our paradigm, both studies used fairly restricted experimental tasks, whereas our paradigm created a continuous interactive experience.

Taken together, these findings suggest that social referencing (i.e. refocusing attention on a social partner) (Feinman, Roberts, Hsieh, Sawyer, & Swanson, 1992; Bayliss et al., 2013; Willemse et al., 2018) is enhanced during continuous JA (i.e. shorter latencies for return to face saccades), lending credence to the immersiveness of the JA experience as evoked in our approach using TriPy. Accordingly, we also observed much longer latencies for back-to-face saccades if the agent did not respond with JA ($\langle\tau_b^{IJA}\rangle_{HI} > \langle\tau_o^{IJA}\rangle_{HI}$). A further interesting aspect is the considerable amount of fixations on the agent's face we observed during the state. Even in the absence of any attempt to interact and despite explicit instruction to focus on objects (see also (Bayliss et al., 2013)), the presence of an agent with direct gaze still captures much of the HIs attention (see e.g. (Senju & Hasegawa, 2005)).

Taken together, our results are well in accordance with previous studies. Furthermore, our approach of an algorithmic implementation within a theoretical model of social gaze allows for substantial extensions of previous findings: Using the toolbox, agents and their parameters can be defined and varied according to the needs of specific interactions settings and experimental contexts, allowing for systematic and fine-grained exploration of social gaze behavior in a triadic VE (e.g. (Jording, Hartz, Bente, Schulte-Rüther, & Vogeley, 2019), submitted).

4.2 Possible applications

4.2.1 Psychiatric conditions

Increasingly, many psychiatric conditions have been conceived as disorders of social cognition (e.g. (Crespi & Badcock, 2008; Vogeley & Newen, 2009; Moutoussis et al., 2014)). Interactive, and adaptive paradigms in social gaze research allow for a highly controlled investigation of dyadic social interaction and thus can grant insights into subtle deviations in behavior during on-going social interactions and their underlying cognitive mechanisms (Frischen, Bayliss, & Tipper, 2007). Numerous psychiatric conditions have been associated with deviations in gaze behavior during face-to-face situations, for example social anxiety (Weeks, Howell, Srivastav, & Goldin, 2019), depression (Grossheinrich et al., 2018), schizophrenia (Caruana, Seymour, Brock, & Langdon, 2019), and most prominently in autism spectrum conditions (ASC) (Frazier et al., 2017). Such approaches, however, are typically non-interactive and only rarely gaze-contingent. In ASC, difficulty in establishing JA is one of the earliest signs of a deficit in social communication in these individuals (Mundy, 2003; Mundy & Newell, 2007). Although individuals in the spectrum typically develop basic JA capabilities later in development, subtle alterations can still be detected in adolescence (Oberwelland et al., 2017). However, standardized behavioral assessment is available only for young children and a more fine-grained behavioral investigation of JA and its deviation during the whole course of development is urgently needed. Unrestricted settings with real human interactants (see e.g. (Birmingham, Johnston, & Iarocci, 2017) for a real-life setting in children and adolescents with ASC) In this study, children and adolescents with ASC were free to look at objects or follow someones gaze in a real-life setting. Interestingly, this study revealed differences in temporal patterns of gaze following, but lacked detail and rich data. More controlled settings including fine-grained eye-tracking recordings but also providing sufficient immersiveness, such as TriPy, would be ideal for further investigation: TriPy, can be used for human-agent interactions in a much more dyadic and interactive way than in previous computerized approaches, thus combining the advantages of a controlled setting and natural interaction. It offers full control over the parameters governing gaze behavior of the agent and allows for in-depth assessment of temporal gaze patterns of the HI to differentiate between specific states of attention and dynamic markers of on-going communication. This approach could be fruitful for a range of psychiatric conditions with respect to the characterization of gaze behavior and social deficits.

A further interesting application would be the implementation of agents displaying gaze pattern that resemble the behavior of persons under different psychopathological conditions (e.g. ASC) and how that may influence communicative behavior in a dyad with typical HIs or in dyads of individuals with specific psychiatric conditions. Furthermore, prototypical gaze parameters for psychiatric conditions and their deviation from typical reference parameters could be used as diagnostic markers and to define training targets for interventional studies that attempt to ameliorate social interaction and eye gaze in ASC.

4.2.2 Naturalistic human-agent interaction

Advances in technical possibilities, proposed algorithms, and computational power sparked the emergence of social agents for face-to-face interaction. Such agents are increasingly used in diverse contexts: As assistants for "customer relations" (e.g. (Kopp, Gesellensetter, Krämer, & Wachsmuth, 2005; Heaven, 2018)), in interactive teaching contexts (e.g. (Lee, Kanakogi, & Hiraki, 2015; Mabanza, 2016))), and basic scientific research (e.g. (von der Pütten, Krämer, Gratch, & Kang, 2010; Jording et al., 2019)(submitted)); for a general review on social robots and more examples see (Mavridis, 2015).

To design such artificial agents, the underlying cognitive architectures incorporating natural JA behavior for action coordination need to be better understood (Deák, Fasel, & Movellan, 2001). This includes the production of "natural" behavior which can be perceived as intentional by the HI, but at the same time also a real-time prediction of the HI's intentions based on his behavior. Both inference and display of intentional JA behavior can only be achieved with sufficient knowledge about the pattern and temporal fine-tuning of human reference behavior. The incorporation of such knowledge greatly increases acceptance of artificial agents. For example, (Huang & Thomaz, 2011) found that robots which exhibit joint attention behavior during interactive tasks were consistently judged as performing better and their behavior is perceived as more natural.

In this respect, TriPy could be a valuable tool to define respective JA situations, their affordances, and extract temporal gaze patterns to construct respective artificial agents and to fine-tune them empirically in an iterative process of parameter estimation from human-agent interaction and therefore has a scope beyond other systems and approaches for human-agent interaction (e.g. (Pfeiffer-Leßmann & Wachsmuth, 2009; Yu, Scheutz, & Schermerhorn, 2010; Yu, Schermerhorn, & Scheutz, 2012; Grynszpan et al., 2012; Stephenson, Edwards, Howard, & Bayliss, 2018; Willemse et al., 2018)).

5 Conclusion

TriPy aims at encouraging in-depth exploration of patterns in human social gaze behavior and their dynamics in triadic encounters offering an unprecedented ecologically validity: An virtual agent resembling the *Social Gaze States* (Jording et al., 2018) has been implemented and its behavioral parameters have been obtained empirically from human gaze data in an iterative process. TriPy's modular and extendable structure offers a flexible approach to create highly controlled experimental virtual environments for social gaze research with potential applications in the investigation of psychiatric conditions and naturalistic human-agent interaction. It demonstrates utmost flexibility for implementing behavioral states of virtual agents in instances of triadic JA for the creation of individual paradigms along with technical reliability and may spark further research and insights into the dynamics of social gaze interaction.

Acknowledgements AH and BG designed and implemented agent behavior and data acquisition. AH and MJ designed and implemented data analysis. MSR supervised software development. AH drafted the manuscript; BG, MJ, KV, and MSR revised it critically. The authors want to thank Sarah Schwöbel from TU Dresden for helpful feedback on the mathematical notation and Anna Grundmann from Uniklinik RWTH Aachen for supporting the development of the toolbox aspect and documentation of the published code.
This work is a part of the project EMOTISK from the *Bringing technology to the people* research program on human-machine interaction, funded by the Federal German Ministry of Education and Research (grant number for MSR: 16SV7242 / for KV: 16SV7244). MSR received funding by the Excellence Initiative of the German federal and state governments.
The authors have no conflict of interest to declare.

Open Practices Statement The toolbox will be available at https://github.com/msrresearch/TriPy and the data will be avaiable upon request. The experiment was not preregistered.

References

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

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

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006, Dec 01). Gaze cuing and affective judgments of objects: I like what you look at. Psychonomic Bulletin & Review, 13(6), 1061–1066. Retrieved from `https://doi.org/10.3758/BF03213926` doi: 10.3758/BF03213926

Birmingham, E., Johnston, K. H. S., & Iarocci, G. (2017). Spontaneous gaze selection and following during naturalistic social interactions in school-aged children and adolescents with autism spectrum disorder. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 71(3), 243.

Burgoon, J., & Buller, D. (1989). Woodall, wg (1989). nonverbal communication: The unspoken dialogue. New York: Harper and Row.

Caruana, N., Ham, H. S., Brock, J., Woolgar, A., Kloth, N., Palermo, R., & McArthur, G. (2018). Joint attention difficulties in autistic adults: An interactive eye-tracking study. Autism, 22(4), 502-512. Retrieved from `https://doi.org/10.1177/1362361316676204` (PMID: 28423919) doi: 10.1177/1362361316676204

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017). Detecting communicative intent in a computerised test of joint attention. PeerJ, 5, e2899.

Caruana, N., Seymour, K., Brock, J., & Langdon, R. (2019). Responding to joint attention bids in schizophrenia: An interactive eye-tracking study. Quarterly Journal of Experimental Psychology, 0(0), 1747021819829718. Retrieved from `https://doi.org/10.1177/1747021819829718` (PMID: 30672375) doi: 10.1177/1747021819829718

Crespi, B., & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. Behavioral and Brain Sciences, 31(3), 241261. doi: 10.1017/S0140525X08004214

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014, Dec 01). Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. Behavior Research Methods, 46(4), 913–921. Retrieved from `https://doi.org/10.3758/s13428-013-0422-2` doi: 10.3758/s13428-013-0422-2

Deák, G. O., Fasel, I., & Movellan, J. (2001). The emergence of shared attention: Using robots to test developmental theories. In Proceedings of the 1st international workshop on epigenetic robotics.. lund university cognitive studies (Vol. 85, pp. 95–104).

Delignette-Muller, M. L., Dutang, C., Pouillot, R., Denis, J.-B., Siberchicot, A., & Siberchicot, M. A. (2017). Package fitdistrplus.

Doyle, L., & Schindler, D. (2015, May). muCap: Connecting FaceReader to z-Tree (Discussion Papers in Economics No. 24809). University of Munich, Department of Economics. Retrieved from https://ideas.repec.org/p/lmu/muenec/24809.html

Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience & Biobehavioral Reviews, 24(6), 581 - 604. Retrieved from http://www.sciencedirect.com/science/article/pii/S0149763400000257 doi: https://doi.org/10.1016/S0149-7634(00)00025-7

Feinman, S., Roberts, D., Hsieh, K.-F., Sawyer, D., & Swanson, D. (1992). A critical review of social referencing in infancy. In S. Feinman (Ed.), Social referencing and the social construction of reality in infancy (pp. 15–54). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-1-4899-2462-9_2 doi: 10.1007/978-1-4899-2462-9_2

Fiske, S. T., & Taylor, S. E. (2013). Social cognition: From brains to culture. Sage.

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. Journal of the American Academy of Child & Adolescent Psychiatry, 56(7), 546 - 555. Retrieved from http://www.sciencedirect.com/science/article/pii/S0890856717302071 doi: https://doi.org/10.1016/j.jaac.2017.05.005

Friesen, E., & Ekman, P. (1978). Facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists, San Francisco.

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.

Gagniuc, P. A. (2017). Markov chains: From theory to implementation and experimentation. John Wiley & Sons.

Gibson, J. J., & Pick, A. D. (1963). Perception of another person's looking behavior. The American Journal of Psychology, 76(3), 386–394. Retrieved from http://www.jstor.org/stable/1419779

Grossheinrich, N., Firk, C., Schulte-Rüther, M., von Leupoldt, A., Konrad, K., & Huestegge, L. (2018). Looking while unhappy: a mood-congruent attention bias toward sad adult faces in children. Frontiers in psychology, 9.

Grossmann, T. (2017). The eyes as windows into other minds: An integrative perspective. Perspectives on Psychological Science, 12(1), 107-121. Retrieved from https://doi.org/10.1177/1745691616654457 (PMID: 28073330) doi: 10.1177/1745691616654457

Grynszpan, O., Nadel, J., Martin, J.-C., Simonin, J., Bailleul, P., Wang, Y., … Constant, J. (2012, Aug 01). Self-monitoring of gaze in high functioning autism. Journal of Autism and Developmental Disorders, 42(8), 1642–1650. Retrieved from https://doi.org/10.1007/s10803-011-1404-9 doi: 10.1007/s10803-011-1404-9

Heaven, D. (2018). Ai to interrogate travellers. New Scientist, 240(3202), 5. Retrieved from http://www.sciencedirect.com/science/article/pii/S0262407918320141 doi: https://doi.org/10.1016/S0262-4079(18)32014-1

Huang, C., & Thomaz, A. L. (2011, July). Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In 2011 ro-man (p. 65-71). doi: 10.1109/ROMAN.2011.6005230

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2018). The social gaze space: A taxonomy for gaze-based communication in triadic interactions. Frontiers in Psychology, 9, 226. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2018.00226 doi: 10.3389/fpsyg.2018.00226

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2019). Inferring interactivity from gaze patterns during triadic person-object-agent interactions. Frontiers in Psychology, X, XXX.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. Acta Psychologica, 26, 22 - 63. Retrieved from http://www.sciencedirect.com/science/article/pii/0001691867900054 doi: https://doi.org/10.1016/0001-6918(67)90005-4

Kleinke, C. L. (1986). Gaze and eye contact: a research review. Psychological bulletin, 100(1), 78.

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. Nature, 387(6635), 767.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. Journal of Human Evolution, 40(5), 419 - 435. Retrieved from http://www.sciencedirect.com/science/article/pii/S0047248401904683 doi: https://doi.org/10.1006/jhev.2001.0468

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide – design and evaluation of a real-world application. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), Intelligent virtual agents (pp. 329–343). Berlin, Heidelberg: Springer Berlin Heidelberg.

Krämer, N. C. (2008). Nonverbal communication. Human behavior in military contexts, 150–188.

Lee, H., Kanakogi, Y., & Hiraki, K. (2015). Building a responsive teacher: how temporal contingency of gaze interaction influences word learning with virtual tutors. Royal Society open science, 2(1), 140361.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clueson the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probabilitynormal or log-normal: That is the question. BioScience, 51(5), 341-352. Retrieved from http://dx.doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2 doi: 10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Mabanza, N. (2016). Determining the usability of pedagogical interface agents in the context of adult computer literacy training: a south african perspective (Unpublished doctoral dissertation). University of the Free State.

Mavridis, N. (2015). A review of verbal and non-verbal humanrobot interactive communication. Robotics and Autonomous Systems, 63, 22 - 35. Retrieved from http://www.sciencedirect.com/science/article/pii/S0921889014002164 doi: https://doi.org/10.1016/j.robot.2014.09.031

Moore, C., Dunham, P. J., & Dunham, P. (2014). Joint attention: Its origins and role in development. Psychology Press.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. Consciousness and Cognition, 25, 67 - 76. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053810014000105 doi: https://doi.org/10.1016/j.concog.2014.01.009

Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. Journal of Child Psychology and Psychiatry, 44(6), 793-809. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/1469-7610.00165 doi: 10.1111/1469-7610.00165

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. Current Directions in Psychological Science, 16(5), 269-274. Retrieved from https://doi.org/10.1111/j.1467-8721.2007.00518.x (PMID: 19343102) doi: 10.1111/j.1467-8721.2007.00518.x

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S., Vogeley, K., Fink, G., … Schulte-Rüther, M. (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fmri in a developmental sample. NeuroImage, 130(Supplement C), 248 - 260. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811916001336 doi: https://doi.org/10.1016/j.neuroimage.2016.02.026

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S., Vogeley, K., Fink, G., … Schulte-Rüther, M. (2017). Young adolescents with autism show abnormal joint attention network: A gaze contingent fmri study. NeuroImage: Clinical, 14(Supplement C), 112 - 121. Retrieved from http://www.sciencedirect.com/science/article/pii/S2213158217300062 doi: https://doi.org/10.1016/j.nicl.2017.01.006

Peirce, J. (2009). Generating stimuli for neuroscience using psychopy. Frontiers in Neuroinformatics, 2, 10. Retrieved from https://www.frontiersin.org/article/10.3389/neuro.11.010.2008 doi: 10.3389/neuro.11.010.2008

Pfeiffer, U., Schilbach, L., Timmermans, B., Jording, M., Bente, G., & Vogeley, K. (2012). Eyes on the mind: Investigating the influence of gaze dynamics on the perception of others in real-time social interaction. Frontiers in Psychology, 3, 537. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2012.00537 doi: 10.3389/fpsyg.2012.00537

Pfeiffer, U., Timmermans, B., Vogeley, K., Frith, C., & Schilbach, L. (2013). Towards a neuroscience of social interaction. Frontiers in Human Neuroscience, 7, 22. Retrieved from https://www.frontiersin.org/article/10.3389/fnhum.2013.00022 doi: 10.3389/fnhum.2013.00022

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. NeuroImage, 101(Supplement C), 124 - 137. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811914005382 doi: https://doi.org/10.1016/j.neuroimage.2014.06.061

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, Part 2), 2516 - 2528. Retrieved from http://www.sciencedirect.com/science/article/pii/S0149763413001887 doi: https://doi.org/10.1016/j.neubiorev.2013.07.017

Pfeiffer-Leßmann, N., Pfeiffer, T., & Wachsmuth, I. (2012).

An operational model of joint attention-timing of gaze patterns in interactions between humans and a virtual human. In Proceedings of the 34th annual conference of the cognitive science society.

Pfeiffer-Leßmann, N., & Wachsmuth, I. (2009). Formalizing joint attention in cooperative interaction with a virtual human. In B. Mertsching, M. Hund, & Z. Aziz (Eds.), Ki 2009: Advances in artificial intelligence (pp. 540–547). Berlin, Heidelberg: Springer Berlin Heidelberg.

Plummer, M. (2018). rjags: Bayesian graphical models using mcmc [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rjags (R package version 4-8)

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. Psychological bulletin, 86(3), 446.

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

Risko, E., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. Current Directions in Psychological Science, 25(1), 70-74. Retrieved from https://doi.org/10.1177/0963721415617806 doi: 10.1177/0963721415617806

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), 393414. doi: 10.1017/S0140525X12000660

Schulte-Rüther, M., Otte, E., Adigüzel, K., Firk, C., Herpertz-Dahlmann, B., Koch, I., & Konrad, K. (2017). Intact mirror mechanisms for automatic facial emotions in children and adolescents with autism spectrum disorder. Autism Research, 10(2), 298-310. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/aur.1654 doi: 10.1002/aur.1654

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

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

Stephenson, L. J., Edwards, S. G., Howard, E. E., & Bayliss, A. P. (2018). Eyes that bind us: Gaze leading induces an implicit sense of agency. Cognition, 172, 124 - 133. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010027717303311 doi: https://doi.org/10.1016/j.cognition.2017.12.011

Vogeley, K., & Bente, G. (2010). artificial humans: Psychology and neuroscience perspectives on embodiment and nonverbal communication. Neural Networks, 23(8), 1077 - 1090. Retrieved from http://www.sciencedirect.com/science/article/pii/S0893608010001164 (Social Cognition: From Babies to Robots) doi: https://doi.org/10.1016/j.neunet.2010.06.003

Vogeley, K., & Newen, A. (2009). Consciousness of oneself and others in relation to mental disorders. The neuropsychology of mental illness, 408–13.

von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010, November). "it doesn't matter what you are!" explaining social effects of agents and avatars. Comput. Hum. Behav., 26(6), 1641–1650. Retrieved from http://dx.doi.org/10.1016/j.chb.2010.06.012 doi: 10.1016/j.chb.2010.06.012

Weeks, J. W., Howell, A. N., Srivastav, A., & Goldin, P. R. (2019). fear guides the eyes of the beholder: Assessing gaze avoidance in social anxiety disorder via covert eye tracking of dynamic social stimuli. Journal of Anxiety Disorders, 65, 56 - 63. Retrieved from http://www.sciencedirect.com/science/article/pii/S0887618518304092 doi: https://doi.org/10.1016/j.janxdis.2019.05.005

Willemse, C., Marchesi, S., & Wykowska, A. (2018). Robot faces that follow gaze facilitate attentional engagement and increase their likeability. Frontiers in Psychology, 9, 70. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2018.00070 doi: 10.3389/fpsyg.2018.00070

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). Its in your eyes – using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. Social Cognitive and Affective Neuroscience, 5(1), 98-107. Retrieved from +http://dx.doi.org/10.1093/scan/nsq024 doi: 10.1093/scan/nsq024

Yarbus, A. (1967). Eye movements and vision. 1967. New York.

Yu, C., Schermerhorn, P., & Scheutz, M. (2012, January).

Adaptive eye gaze patterns in interactions with human and artificial agents. ACM Trans. Interact. Intell. Syst., 1(2), 13:1–13:25. Retrieved from http://doi.acm.org/10.1145/2070719.2070726 doi: 10.1145/2070719.2070726

Yu, C., Scheutz, M., & Schermerhorn, P. (2010). Investigating multimodal real-time patterns of joint attention in an hri word learning task. In Proceedings of the 5th acm/ieee international conference on human-robot interaction (pp. 309–316). Piscataway, NJ, USA: IEEE Press. Retrieved from http://dl.acm.org/citation.cfm?id=1734454.1734561

6 Supplement

6.1 Supplementary figures

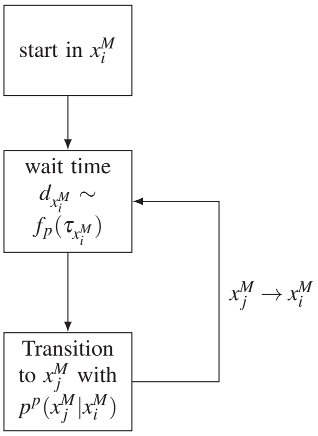


Fig. S1 State diagram: *Non-interactive Agents* ($M \in \{PO, OO, INT\}$). The agent starts in a micro state x_i^M and waits for a state duration $d_{x_i^M} \sim f_p(\tau_{x_i^M})$ before transitioning to a new micro state x_j^M with probability $p^p(x_j^M|x_i^M) \in \mathbf{p}_{x_i}$, and samples again new micro state duration. This resembles the type of state transitions depicted by the bottom arrows in Fig. 2.

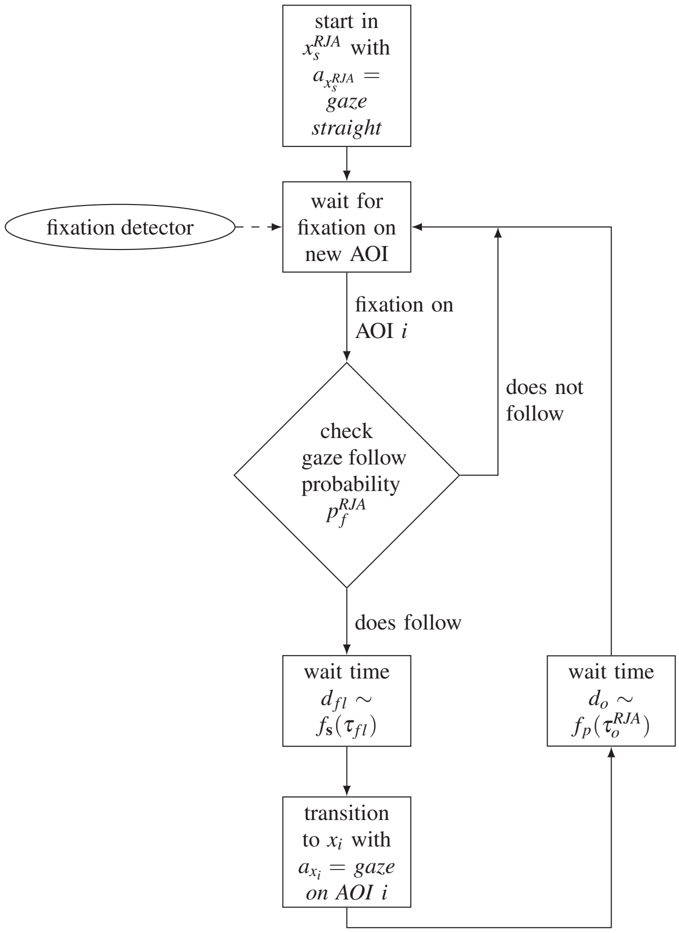


Fig. S2 State diagram: *RJA* Agent. The agent starts gazing straight at the HI waiting for a fixation on one of the predefined AOIs. With follow probability p_f, it changes to micro state x_i with a_x =gaze on AOI i. After a time $d_o \sim f_p(\tau_o^{RJA})$, the agent waits again for a fixation on a new AOI, ready to follow another JA bid by the HI.

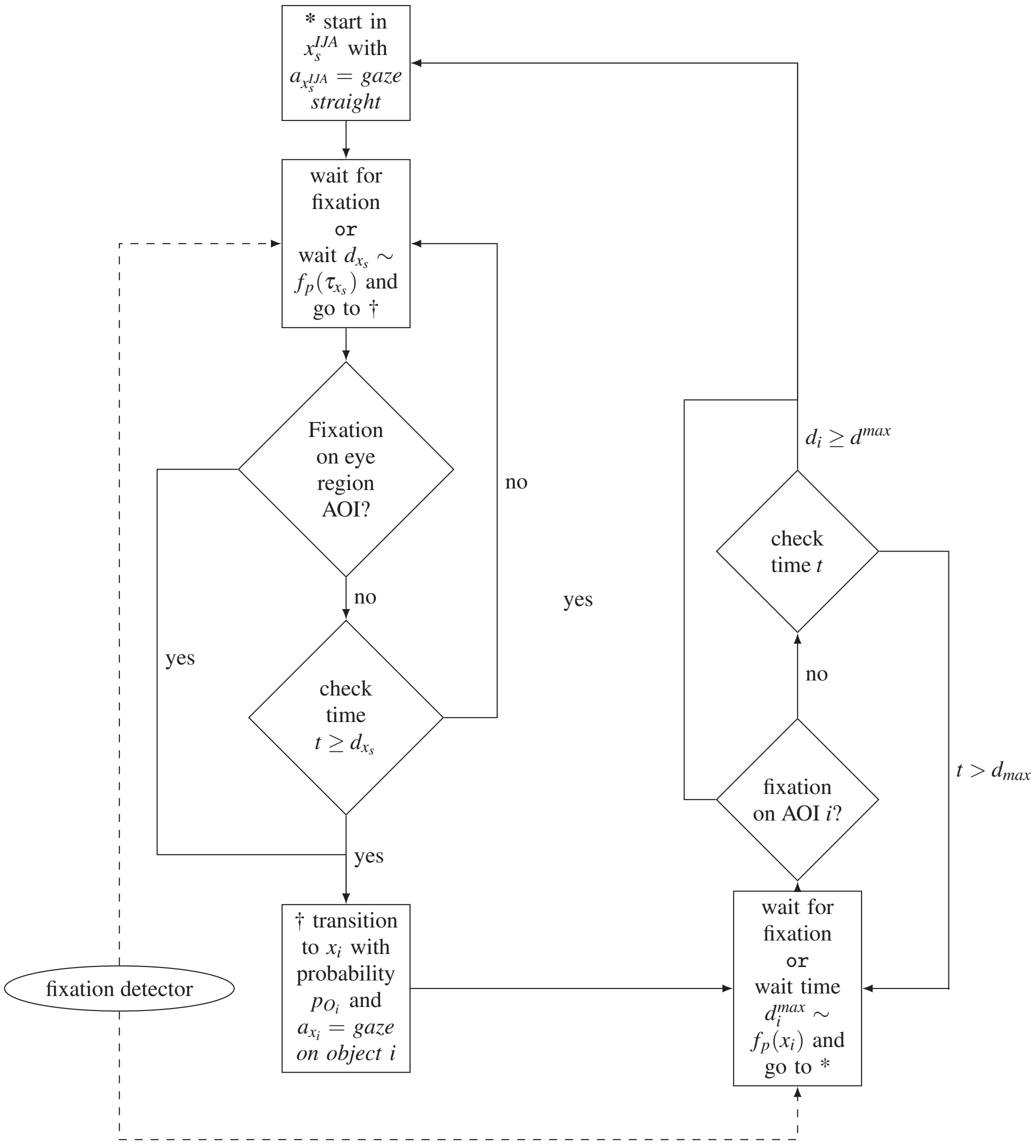


Fig. S3 State diagram: IJA macro state. The agent starts in micro state x_s^{IJA} gazing straight at the HI trying to establish eye contact. After eye contact is established, or after a waiting time t, the agent tries to initiate JA on object i, i.e. it transitions to micro state x_i, with probability p_{O_i} and gaze on object i. It then waits for a time t for a fixation by the HI on the same AOI i, before switching back to micro state x_s and starts a new JA bid. Note that in this case, micro state transitions can either be passive or interactive

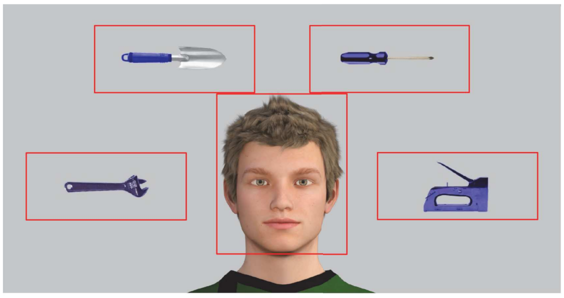


Fig. S4 AOIs (red rectangles) for offline data analysis to estimate behavioral parameters. Slightly larger-than-object-size AOIs were used for noise robustness (Hessels, 2017)

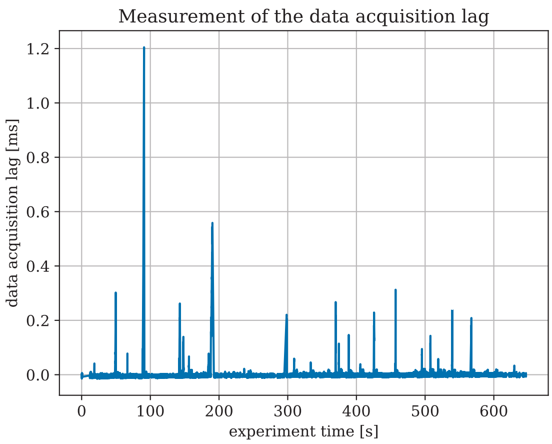


Fig. S5 Difference between the system clocks of the eye tracker (as provided by the Tobii SDK) and stimulus PC over the course of one experimental block. Spikes indicate a neglectable lag ($\sim$ 1 ms) in the acquisition of data from the eye tracker by the PC clock.

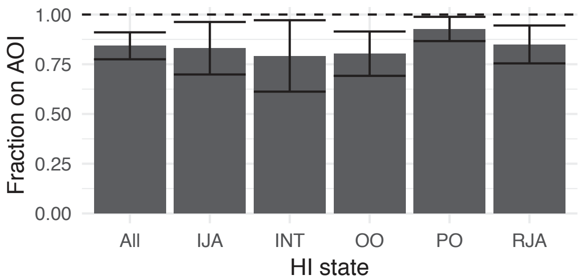


Fig. S6 On-Aoi ratio r_{aoi}^{M} across HIs.

6.2 Supplementary tables

Table S1 Non-interactive macro states: Each macro state $M \in \{PO, OO, INT\}$ is composed of n_M unique micro states $M = \{x_k^M\} = \{x_1^M, \dots, x_{n_M}^M\}$. The k^{th} state is described by an appearance $a_{x_k^M}$, set(s) $\tau^{\alpha}_{x_k^M}$ of a probabilistic distribution $q^{\alpha}_{x_k^M}$ and corresponding parameters $\Theta^{\alpha}_{x_k^M} = \{\mu^{\alpha}_{x_k^M}, \sigma^{\alpha}_{x_k^M}, \dots\}$ for drawing a state duration $d_{x_k^M} \sim q^{p}_{x_k^M}(d_{x_k^M}|\Theta_{x_k^M})$ as well as probabilities $\mathbf{p}_{x_k^M}$ for passive state transitions to the other micro states $\{x_{j\neq i}\}$. but their numerical values and distribution types may differ.

state	$a_{x_k^M}$	set	**S**	description
$\{x_k^M\}$	gaze at AOI k	$\tau^p_{x_k^M}$	-	duration of micro state i
		$\mathbf{p}^p_{x_k^M}$	-	probabilities of transitioning from micro state x_k^M to any other micro state from $\{x_k^M\}$

Table S2 Interactive macro state: RJA. Notation is the same is in Tab. S1.

state	a_{x_k}	set	description
$\{x_k^{RJA}\}$	gaze on AOI k	$\tau^i_{x_k^{RJA}}$	Time after which agent will follow with gaze on AOI for successful JA
		$p^{RJA}_{x_k}$	Probability of agent following on same AOI for successful JA
		$\tau^i_{x_k^{RJA}}$	Time agent keeps fixating object after successful JA and HI not fixating AOI anymore

Table S3 Interactive macro state: IJA. Notation is the same is in Tab. S1. Note that all states have both passive and interactive possibilities for state transitions.

state	a_x	set	**S**	description
x_s^{IJA}	gaze straight (on HI)	$\tau^i_{x_s^{IJA}}$	*gaze on AOI i*	Duration of the agent looking straight after detected fixation on agent's eyes AOI before fixating an object to try to initiate JA.
		$\tau^p_{x_s^{IJA}}$	-	Duration of the agent looking straight at the HI if no fixation on the eyes AOI was detected before choosing an AOI to gaze at to try to initiate JA.
		$\mathbf{p}^{p,i}_{x_s^{IJA}}$	-	Transition probablities to micro states $\{x^{IJA}\}$.
$x_{o_i}^{IJA}$	gaze at AOI i (object i)	$\tau^{i,IJA}_{x_i}$	*fixation loss*	Duration of the continuation of the agent looking at the AOI on which in tried to initiate JA and after a fixation of the HI has been detected on that AOI.
		$\tau^{p,IJA}_{x_i}$	-	Duration of the agent looking at the AOI on which it tried to initiate JA and if no fixation of the HI was detected on that AOI.
		$\mathbf{p}^{p,i}_{x_{o_i}^{IJA}}$	-	Transition probablities to micro states $\{x^{IJA}\}$. With $p(x_s|x_{o_i}) = 1$ it is ensured that the agent will always look straigt after a round of (tried) IJA.

Table S4 Instructions presented to HI to interact with the agent. For the sake of enhanced ecological validity the agent was named Paul in the instructions.

macro state	instruction (translated from German)
INT *introspective*	Please keep your eyes open and focus on your breath.
OO *object-oriented*	Please focus on the objects.
PO *partner-oriented*	Please focus on Paul.
RJA *responding JA*	Please interact with Paul and let him guide you.
IJA *initiating JA*	Please interact with Paul and try to guide him.

Table S5 PC hardware used to implement and test TriPy.

component	System 1	System 2
CPU	Intel Xenon E5-1620v2	Intel Xenon E5-1603v3
RAM	8 GB DDR3 1600 MHz	16 GB DDR4 3000 MHz
GPU	Nvidia Quadro K2000	Nvidia Quadro K420
OS	Windows 7	Windows 7

Table S6 Macro state combinations measured in 1^{st} paradigm: The numbers show how often a given macro state combination was used in the paradigm. The $6+6$ for the RJA-IJA combination denotes that the RJA agent was used with two different follow probabilities. ($p_f^1 = 100\%$ and $p_f^2 = 33\%$). To keep the experiment as short as possible, we only used macro state combinations of interactive states that can lead to successful JA.

HI \ A	INT	OO	PO	RJA	IJA
INT	6	-	6	-	-
OO	6	-	6	-	-
PO	6	-	6	-	-
RJA	-	-	-	-	12
IJA	-	-	-	$6+6$	-

Table S7 Transition probabilities in *INT* state in 1^{st} paradigm

aoi	down	down left	down right	straight
down	.3	.3	.3	.1
down left	.3	.3	.3	.1
down right	.3	.3	.3	.1
straight	.3	.3	.3	.1

Table S8 Transition probabilities in *PO* state in 1^{st} paradigm

aoi	down	down left	down right	straight
down	.033	.033	.033	.9
down left	.033	.033	.033	.9
down right	.033	.033	.033	.9
straight	.033	.033	.033	.9

Table S9 Temporal parameters in 1^{st} paradigm

State	Parameter name	mean in ms	SD in ms
all	τ_{IBI}	3500	500
	τ_{bd}	100	0
RJA	τ_{fl}^{RJA}	500	100
	τ_{rl}^{RJA}	500	150
IJA	τ_{s}^{IJA}	1500	300
	τ_{sa}^{IJA}	500	100
	τ_{bore}^{IJA}	6000	300
INT	τ_{d}^{INT}	2800	700
PO	τ_{d}^{PO}	2800	700

Table S10 Interactive probabilistic parameters in 1^{st} paradigm

State	set	value
IJA	$\{p_{O_i}\}$	.25
*RJA*1	p_f^1	1.0
*RJA*2	p_f^2	.33

6.3 Supplementary methods

6.3.1 Availability and documentation

TriPy is available via https://github.com/msrresearch/TriPy and comes with two example paradigms, a documentation generated with Sphinx (http://sphinx-doc.org/), and manually annotated code to give the interested user a first impression and the possibility to easily adapt and extend the code. Included features experimental features and behavioral parameters can be adapted via simple configuration files.

6.3.2 Hardware requirements

TriPy was developed using a Tobii TX300 eye tracker (Tobii Technology, Sweden), which allows for head movement within a tracking box of 30 * 37 * 17 cm^3 at a sampling rate of 300 Hz. This relatively unrestrained setting (i.e. no need for a chin rest) supports the impression of a natural interaction and increases ecological validity especially for social encounters. Other eye tracking systems could be added with minimal/reasonable effort if they are supported by PyGaze. Preliminary: We will test an EyeLink 1000.

A standard off-the-shelf webcam with reasonably good image quality can be used for synchronized video recording of the HI via μCap (Doyle & Schindler, 2015).

TriPy was developed and tested on contemporary mid range performance PC hardware (see Tab. S5 for a detailed lists of systems). Compatibility to Linux was only briefly tested on a slow machine (Lenovo ThinkPad x230) and no software issues occurred.

For offline analysis and more accurate stimulus presentation timings on the computer screen, support for Cedrus StimTracker (Cedrus Corporation, USA) connected to the TX300 is implemented as an optional feature.

6.3.3 Supplementary software information

The visual representation of the agent's behavior is generated by switching between pre-rendered images displaying changes in gaze direction and/or facial expression. They were created with DAZ Studio 4.9 (DAZ Productions, Inc., USA), a software package which is freely available and creates detailed and highly configurable facial expressions with sufficient objective realism while providing an easy to use interface. The virtual character (VC) representing the agent to the HI was selected from an online survey of multiple possible VC meeting the criteria of an average rating for dominance and trustworthiness ((Oosterhof & Todorov, 2008) (rated 5.2 ± 2.2 and 5.3 ± 1.6, respectively on a a nine-point likert scale ($0 =$ characteristic not shown, $9 =$ very strongly shown) in a sample of twenty-seven participants. Within TriPy, VCs can be easily changed to meet the requirements for experimental manipulation.

6.3.4 Technical reliability

Since our toolbox adds a rather big overhead to PyGaze, accurate monitoring and minimization of any possible latencies in the availability of eye tracking data was of utmost importance throughout the development. This is particularly important for gaze-contingent paradigms because inaccuracies are a source of noise in experimental data and might disrupt the smooth flow of the experiment and the experience of natural interaction. To ensure the accuracy of measurements within TriPy, we compared the timestamp of the last available gaze data package (from the eye tracker, as provided by the Tobii SDK) to the CPU time (on the computer running TriPy) at the moment of its acquisition. Due to the format of gaze data from Tobii eye trackers (time variable with an arbitrary starting point), it is not possible to give a reliable estimate of this lag during run time, only a jitter in acquisition. TriPy introduces uncertainties at least two magnitudes smaller than the effects of interest in social gaze research (1 to 100's ms) (Fig. S5).

6.3.5 Formal micro state definition

We defined a micro state $x = \{a_x, \tau_x^p, \mathbf{p}_x^p, \tau_x^i, \mathbf{p}_x^i\}$ by

1. a_x: nonverbal appearance of the agent conveying social information to the HI (e.g. gaze direction, facial expression and/or combinations thereof).
2. τ_x^p: a set of random distribution q_x^p and corresponding parameters Θ_x^p to draw a micro state duration $d_x^p = f(\tau_x^p) = q_x^p(\Theta_x^p)$. [3] The index p denotes that these parameters are for passive, i.e. non-interactive, micro state transitions.
3. $\mathbf{p}_x^p = (p^p(x_1|x), p^p(x_2|x), \ldots, p^p(x_n|x))$ with $\sum_i p^p(x_i|x) = 1$ and $p(x_i|x_i) = 0$: a vector containing probabilities for transitioning passively from micro sate x to another micro state x_k with $k \in (1 \ldots n)$ after state duration d_x^p.[4]
4. $\tau_x^{s^i}$: Same as τ_x^p, but for determining micro state duration $d_x^{s^i}$ after registration predefined socio-emotional signal s^i of the HI (e.g. fixations on AOIs, target facial expressions) to which micro state x is sensitive.
5. $\mathbf{p}_x^{s^i}$: Same as $\mathbf{p}_x^p$, but again for an socio-emotional signal s^i as in 4..

According to this definition, there are two different possibilities for micro state transition: Either (1) interactively in response to a behavioral event s_i emitted by the HI after a specified post-event time $d_x^{s_i} = q_x^{s_i}(\Theta_x^i)$ (top arrows in Fig. 2) or (2) passively after a micro state duration $d_x^p = q_x^p(\Theta_x^p)$ (bottom arrows in Fig. 2). In this latter case, agent behavior resembles a simple Markov chain as the next micro state

[3] An example would be being a Gaussian distribution and Θ being variance σ and mean μ.

[4] $p(\cdot|\cdot)$ reads as a conditional probability: $p(x_j|x_k)$ is the probability of transitioning to micro state x_j given that the agent is in state x_k.

x_{t+1} only depends on the current state x_t via transition probabilities $\mathbf{p}_{x_t}$.

To the HI only the obvious presentation of non-verbal appearance a_x and the actual micro state duration d_x, but not the dependencies and underlying distributions are directly observable.

6.3.6 Participants

37 adult volunteers (20 identifying themselves as female, 16 as male, 1 as queer; mean age 28.2 ± 9.2) with no self-reported record of neurological or psychiatric conditions took part in the experiment. The study was approved by the Research Ethics Committees of University Hospital Cologne and University Hospital RWTH Aachen. All volunteers were recruited via mailing lists and postings on the campus of the University of Cologne and gave written informed consent prior to participation. All data was acquired in Cologne.

6.3.7 Data (pre)processing and exclusion

Data analysis was performed using R v3.4.4 (R Core Team, 2018) on Ubuntu 18.04. Due to calibration problems, four HIs had to be excluded entirely from further analysis. Of a total of 2188 blocks, 38 had to be excluded due to a bug in instruction presentation had to be excluded. Due to a bug in agent blinking behavior 184 blocks had to be excluded. Both bugs are fixed in the release version of TriPy. Another 419 blocks were excluded, because less than 66.6 % of eye tracking data was available, due to calibration issues of single experimental part and head movement outside the tracking box of the Tobii system, and HIs wearing glasses incompatible with the Tobii system. This leaves a total of 1547 blocks included in data analysis (71%). Excluded blocks are not evenly distributed among HIs as most cluster on a single experimental part of single HIs.

6.3.8 Fixation/saccade classification

For offline gaze event detection (fixations and saccades) the raw eye tracking data was smoothed in a sliding window procedure to reduce the impact of noise before classifying fixations and saccades based on an algorithm proposed by (Engbert, Longtin, & Kliegl, 2002) as implemented in the analysis pipeline for a freely-moving head.

This choice was made to keep on- and offline results as similar as possible: An adapted version of this algorithm for real time saccade detection is implemented in PyGaze, which should be favored according to the author (Dalmaijer et al., 2014) over the one for fixation detection for gaze-contingency. Unfortunately, the PyGaze saccade detection algorithm is not robust against head movements and TriPy so far relies on its fixation detection routines which appeared to be robust.

AOI definitions Fixations were assigned to user-defined AOIs (Fig. S4). Generally, arbitrary AOI shapes are possible as they are defined via geometric shapes of the (Bivand, Pebesma, & Gomez-Rubio, 2013) package.

Heat maps For each induced gaze state, we constructed logarithmized heat maps for illustration by drawing a circle around each mean fixation position (x, y) with a radius of $max(SD(fix_x), SD(fix_y))$. This circle was convoluted with a quadratic density distribution around the center integrating to a value proportional to the duration of the fixation. The resulting map was logarithmized due to high fixation density in the eye region and overlayed on a screen shot of the paradigm.

Supplementary References

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). Applied spatial data analysis with R, second edition. Springer, NY. Retrieved from `http://www.asdar-book.org/`

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014, Dec 01). Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. Behavior Research Methods, 46(4), 913–921. Retrieved from `https://doi.org/10.3758/s13428-013-0422-2` doi: 10.3758/s13428-013-0422-2

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. Vision Research, 42(5), 621 - 636. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0042698901003017` doi: https://doi.org/10.1016/S0042-6989(01)00301-7

Hessels, R. S. (2017). Toward early markers for autism spectrum disorder using eye tracking. Journal of Cognitive Education and Psychology, 16(3), 310–312.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. Proceedings of the National Academy of Sciences, 105(32), 11087–11092. Retrieved from `http://www.pnas.org/content/105/32/11087` doi: 10.1073/pnas.0805664105

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 August 2019
doi: 10.3389/fpsyg.2019.01913

Inferring Interactivity From Gaze Patterns During Triadic Person-Object-Agent Interactions

Mathis Jording[1,2]*, *Arne Hartz*[3,4], *Gary Bente*[5], *Martin Schulte-Rüther*[1,3,4] and *Kai Vogeley*[1,2]

[1] Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Center Jülich, Jülich, Germany, [2] Department of Psychiatry, University Hospital Cologne, Cologne, Germany, [3] JARA-BRAIN, Aachen, Germany, [4] Translational Brain Research in Psychiatry and Neurology, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, RWTH Aachen University, Aachen, Germany, [5] Department of Communication, Michigan State University, East Lansing, MI, United States

OPEN ACCESS

Edited by:
*Andreas K. Engel,
University Medical Center
Hamburg-Eppendorf, Germany*

Reviewed by:
*Jari K. Hietanen,
University of Tampere, Finland
Jean Baratgin,
Université Paris 8, France*

***Correspondence:**
*Mathis Jording
m.jording@fz-juelich.de*

Specialty section:
*This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology*

Received: *23 April 2019*
Accepted: *05 August 2019*
Published: *21 August 2019*

Citation:
*Jording M, Hartz A, Bente G,
Schulte-Rüther M and Vogeley K
(2019) Inferring Interactivity From
Gaze Patterns During Triadic
Person-Object-Agent Interactions.
Front. Psychol. 10:1913.
doi: 10.3389/fpsyg.2019.01913*

Observing others' gaze informs us about relevant matters in the environment. Humans' sensitivity to gaze cues and our ability to use this information to focus our own attention is crucial to learning, social coordination, and survival. Gaze can also be a deliberate social signal which captures and directs the gaze of others toward an object of interest. In the current study, we investigated whether the intention to actively communicate one's own attentional focus can be inferred from the dynamics of gaze alone. We used a triadic gaze interaction paradigm based on the recently proposed classification of attentional states and respective gaze patterns in person-object-person interactions, the so-called "social gaze space (SGS)." Twenty-eight participants interacted with a computer controlled virtual agent while they assumed to interact with a real human. During the experiment, the virtual agent engaged in various gaze patterns which were determined by the agent's attentional communicative state, as described by the concept of SGS. After each interaction, participants were asked to judge whether the other person was trying to deliberately interact with them. Results show that participants were able to infer the communicative intention solely from the agent's gaze behavior. The results substantiate claims about the pivotal role of gaze in social coordination and relationship formation. Our results further reveal that social expectations are reflected in differential responses to the displayed gaze patterns and may be crucial for impression formation during gaze-based interaction. To the best of our knowledge, this is the first study to document the experience of interactivity in continuous and contingent triadic gaze interactions.

Keywords: social gaze, joint attention, eye contact, triadic interaction, non-verbal communication, social psychology, human-agent interaction

INTRODUCTION

During social interactions, we consistently focus on the eyes of our interaction partner because it is the fastest and easiest way to access the inner experience of another person (Yarbus, 1967; Baron-Cohen et al., 1997; Emery, 2000). From the eye region alone we are able to infer age, gender, and personality and even identify individual persons (George and Conty, 2008; Itier and Batty, 2009).

We also use gaze to ensure successful communication and smooth interactions by coordinating turn-taking (Argyle and Cook, 1976) and coordinating attention with others. This ability may constitute the phylogenetic and ontogenetic basis of cooperation (Tomasello et al., 2007; Grossmann, 2017). The most prevalent example of coordinated gaze is joint attention i.e., the joint focus of two persons gaze on an object, including gaze following and leading the gaze of others (Emery, 2000). The ability to follow someone else's gaze toward objects is acquired very early in life, possible starting at the age of 6 months (Senju and Csibra, 2008), it provides the basis for reinforcement learning (Vernetti et al., 2017), and the development of a theory of mind and language (Morales et al., 1998). It is therefore not surprising that the proficiency in gaze following predicts social competence, self-regulation abilities, and even the depth of information processing and IQ (Mundy and Newell, 2007).

During everyday encounters with other people, we do not know in advance whether the person we meet is trying to engage us in an interaction or is merely exploring the environment. In other words, we have to disambiguate the dual function of social gaze (Gobel et al., 2015; Jarick and Kingstone, 2015), or the simultaneous use of gaze for visual perception and for communicating with others. That is, we take the communicative states of others into account and adjust our gaze behavior for social adequacy accordingly (Risko and Kingstone, 2011; Wu et al., 2013). Conversely, this also implies that by observation alone we cannot be sure of whether gaze behavior of others is a communicative signal toward us or merely serves perceptual means. One powerful communicative signal is mutual eye contact (Senju and Johnson, 2009) which increases emotional empathy and modulates attention (Farroni et al., 2002; Senju and Hasegawa, 2005; Dalmaso et al., 2017). Thus, eye contact likely fosters the experience of a connection with another person. Furthermore, attempts to establish joint attention can be considered as prototypical gaze-based interaction. However, as of yet it is unclear, which cues are most informative in disambiguating the dual function of social gaze and inferring social communicative intent based on observed gaze alone.

Here we investigate the human ability to recognize communicative attempts from gaze. Using gaze-contingent paradigms with virtual characters (VC) it is possible to investigate ongoing interactions while retaining full experimental control (Vogeley and Bente, 2010; Wilms et al., 2010; Pfeiffer et al., 2013b; Georgescu et al., 2014; Oberwelland et al., 2016, 2017). However, these paradigms suffer from two major limitations: (1) gaze communication is implemented as a series of short, discrete and isolated events and not as an ongoing flux of interaction; (2) the respective paradigms mostly relied on explicit instructions or repetitive, monotonic, and predictable agent behavior. Resolving these limitations required both a theoretical foundation and technological advancements. Theoretically, we developed a new holistic taxonomy of social gaze, the "social gaze space (SGS)" (Jording et al., 2018). The SGS covers all possible categorical states of attention and interaction during gaze-based triadic interactions (constituted by two interactants and at least one object in a shared environment). The different gaze states include: "partner-oriented (PO)," during which the attention is directed solely on the interaction partner; "object-oriented (OO)," attention directed solely on the object(s) in the environment; "introspective (INT)," attention disengaged from the outside world and directed toward inner (e.g., bodily) experiences; "responding joint attention (RJA)," a state of actively following the partner's gaze toward objects of his choice; and "initiating joint attention (IJA)," a state in which the partner's gaze is led toward the objects of one's own choice. The two joint attention states (RJA and IJA) are interactive states in which the agents' behavior depends on the interaction partner, whereas the other three describe states of passive observation. Note, that these five states individually describe the behavior of one of the interaction partners. The interaction between both can be characterized as the combination of both individual states toward a "dual state" (Jording et al., 2018).

Technically, we implemented all five different gaze states of the SGS in the gaze-contingent agent-platform "TriPy" (Hartz et al., submitted). Unlike previous agent-systems, it can generate all SGS states including their responsive properties in real-time. The agent allows for mutual interactions in a continuous and immersive, hence, ecologically valid fashion. The agent's behavior is governed by sets of probabilistic parameters and timing parameters, based on empirical observations during continuous gaze-based interactions (Hartz et al., submitted).

We used this setup to address the question whether and how humans identify communicative intentions from gaze alone. To this end, we asked participants to interact with an algorithmically controlled VC while believing that a real human controlled the VC. Participants had to rate, whether their interaction partner was trying to interact with them or not. We analyzed the participants' decisions and response times (RT) as well as their gaze behavior and the occurrence of eye contact and instances of joint attention. We were interested whether participants would experience differences in the degree of interactivity of the different gaze states as implied by the SGS. We assumed that from the non-interactive states, PO would be rated the most interactive because here the agent focused on the perceiver proportionately more, increasing the probability of eye contact. With respect to interactive states, we hypothesized that the IJA state might be experienced less frequently as interactive compared to RJA. While in IJA participants need to actively follow the agent in order to learn, whether this would move the agent to "show" them the next object, in RJA the agent would strictly follow the participant which we assumed to be easily noticeable. After the experiment, we let participants rate the difficulty of the task and compared it to their performance in identifying interactive situations as an indicator of the conscious accessibility of the underlying cognitive processes.

MATERIALS AND METHODS

Participants

A total of 28 participants without any record of psychiatric or neurological illnesses were recruited via mailing lists, gave their written consent and were compensated for their participation (10€ per hour). Three participants were excluded due to technical failure ($n = 1$) and lack of conviction to interact with a real person ($n = 2$). Data from 25 participants (aged 19 – 57; mean = 31.08, $SD = 11.21$; 16 identifying as female, 9 as male) were further analyzed. This study was approved by the ethics committee of the Medical Faculty of the University of Cologne, Germany, and strictly adhered to the Declaration of Helsinki and the Principles of Good Scientific Practice.

Procedure and Tasks

Before the experiment, participants were briefly introduced to a confederate of the same sex but were brought to another room where they received the detailed written experimental instructions that were repeated orally. Participants were told that both communication partners would be represented by the same standard male VC serving as avatar and that both could only communicate via gaze behavior. They were further told that they would be seated in front of a monitor that displayed the avatar of their partner representing the partner's eye movements on the basis of data provided by two identical eye-tracking systems and updating the respective gaze direction of the avatars in real-time (**Figure 1A**). In fact, participants always solely communicated with an agent controlled by a computer algorithm (Hartz et al., submitted). Participants would further see four trial-wise changing objects, at fixed positions and obviously visible for the partner's avatar (**Figure 1B**). Neither the VC nor the objects were shown to the participant before the start of the experiment.

Participants were further instructed to take two different roles: (1) The Observation-Role (ObR), and (2) the Action-Role (AcR). For the ObR condition, there were no trial specific instructions apart from the task to ascertain whether their partner was trying to "interact" or not (German "austauschen" or "interagieren"), "interacting" was defined as an encounter in which both partners respond to the gaze behavior of the partner in a mutual and reciprocal fashion. Participants were asked to answer only as soon as they felt "quite sure" but were reminded that each trial ended at the latest after 30 s and they therefore would have to hurry. The time between beginning of the trial and button press was logged as RT. When participants had not pressed a button within 30 s, they were asked to decide more quickly in the next trial. After each trial, the participant's choice was displayed on the screen until participants indicated their readiness to continue via button press. Afterward, a message was displayed, asking the participants to wait until their partner was ready for the next trial. This delay was introduced in order to support the participants believe in the confederate based coverstory. The next trial would then begin after a random (uniformly distributed) duration of 1 – 5 s with the appearing of the agents face on the screen.

During the AcR condition, participants were explicitly instructed to engage in one of the states of the SGS (Jording et al., 2018) with the following instructions: "Please concentrate on your partner" (German: "Bitte konzentrieren Sie sich auf Ihren Partner"; PO); "Please attend to the objects" (German: "Bitte achten Sie auf die Objekte"; OO); "Please keep your eyes open and concentrate on your breath" (German: "Bitte lassen Sie Ihre Augen geöffnet und konzentrieren Sie sich auf Ihren Atem"; INT); "Please interact with your partner and let his gaze guide you" (German: "Bitte versuchen Sie sich mit Ihrem Partner auszutauschen und lassen Sie sich von seinem Blick leiten"; RJA), or "Please interact with your partner and use your gaze to guide him" (German: "Bitte versuchen Sie sich mit Ihrem Partner auszutauschen und nutzen Sie Ihren Blick um ihn zu leiten"; IJA). No further instructions were given and participants were told that there was no correct or wrong behavior and they should behave according to their intuitive understanding of these instructions. Trials stopped after 30 s and were followed by a short break of 2 – 6 s.

Whereas ObR was the target condition allowing measuring the experience of interactivity, the AcR condition was included to support the cover story, as participants believed to be interacting with some other real participants and thus would expect a balanced study design with the same tasks for both participants. Both roles were presented alternatingly in three blocks each, with 16 trials per block during ObR and 10 trials per block for AcR. The order of blocks and state instructions within blocks was randomized across participants. After two blocks participants were given a short break of up to 3 min to prevent fatigue and to allow for recalibration of the eyetracker to avoid drifting artifacts.

Setup, Agent-Platform, and Pilot Study

The setup consisted of an eye-tracker with a sampling rate of 120 Hz and an accuracy of 0.5° (Tobii TX300; Tobii Technology, Stockholm, Sweden). A 23" monitor with a screen resolution of 1920*1080 pixels mounted on top of the eye-tracker was used as display (**Figure 1A**). Participants were seated at a distance between 50 – 70 cm to the monitor. A PC-keyboard with the marked buttons "J" and "N" was used for participant responses during ObR. A light sensor based system (StimTracker, Cedrus Corporation, San Pedro, CA, United States) ensured that timing of presented stimuli by the algorithm and actual graphical output were in sync.

The agent's behavior and graphical output was controlled by the agent-platform "TriPy" (Hartz et al., submitted), implemented in Python 2.7 (Python Software Foundation[1]) using PyGaze (Dalmaijer et al., 2014) and PsychoPy (Peirce, 2008). TriPy is based on a gaze-contingent algorithm that adapts the behavior of a VC to the behavior of the participant in real-time (Wilms et al., 2010). In contrast to previous setups, TriPy does not require a prior determination of the exact course and timing of the agents' behavior. Instead, behavior in the non-interactive states is implemented on a probabilistic basis in which the agent displays different micro states (e.g., a moment of looking at one of the objects) with different probabilities (**Figure 2**). In the RJA state the agent follows the participants gaze toward the objects and looks back at the eyes of the participant, when being looked at himself, with a randomly drawn offset between 311.06 – 589.93 ms (lognorm distributed, range 6.06 $\pm$ 0.32). In the IJA state the agent looks at the participant and as soon as eye contact is established or after a randomly drawn waiting period of 772.78 – 2321.57 ms (lognorm distributed, range 7.2 $\pm$ 0.55) looks at one of the objects at random. As soon as the participant follows or after a randomly drawn waiting period of 780.55 – 2440.60 ms (lognorm distributed, range

[1] https://www.python.org

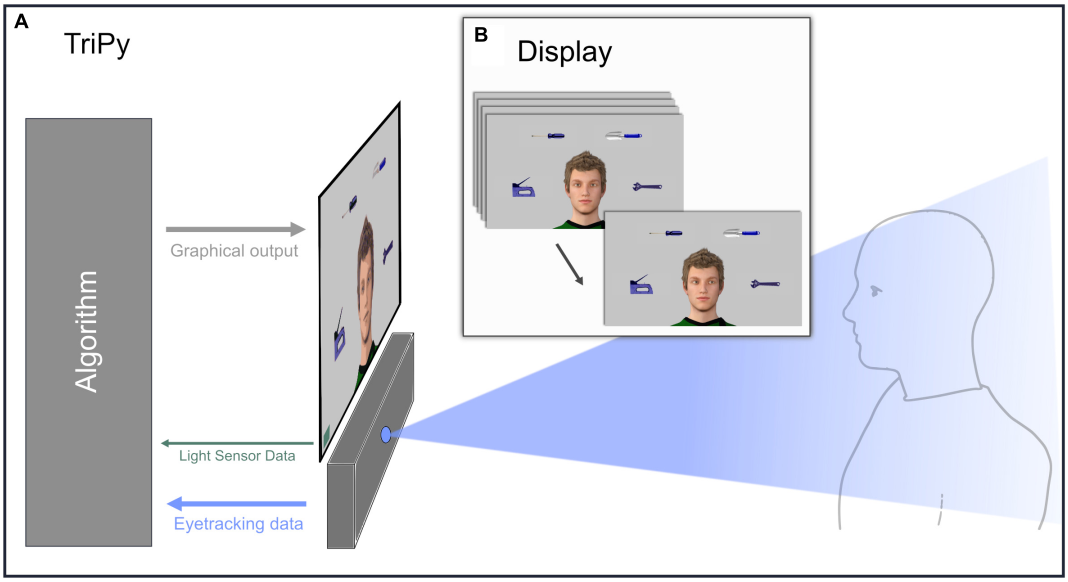


FIGURE 1 | Illustration of the technical setup and the participants' perspective during the experiment. **(A)** Illustration of a participant interacting with the agent controlled by the platform TriPy. **(B)** The behavior of the agent created by TriPy as seen from the perspective of the participant **(B)**.

7.23 ± 0.57), the agent starts anew with trying to establish eye contact and subsequently choosing a new object at random (video examples of the agents behavior in all states can be found in the **Supplementary Material**). These microstates, their durations and transition probabilities, as well as temporal parameters of the interactive agents' states were empirically informed by a pilot study (Hartz et al., submitted). The anthropomorphic VC was created with the modeling software Daz Studio 3.1 (DAZ Productions, Inc., United States).

During the ObR condition, the agent equally often displayed either any of the interactive (25% for each of the interactive states RJA and IJA) or any of the non-interactive states (16.67% for each of the non-interactive states PO, OO, and INT). This partitioning ensured that participants encountered interactive and non-interactive states equally often and thus could not exceed a 50% correctness rate by guessing. During AcR – which was established only to let participants continuously believe that they were interacting with the interaction partner to whom they had been introduced before the experiment – the agents' states corresponded to the states of the participant the agent displayed non-interactive states (PO, OO, or INT) when the participant herself was in a non-interactive state with all combinations of agent and participant states appearing equally often. Each interactive-state of the participant was answered by the agent with the complementary interactive-state (RJA with IJA; IJA with RJA).

Questionnaires and Post-experimental Inquiry and Information

After the experiment participants filled out a post-experimental questionnaire asking on visual analog scales (ranging from 1 to 6): (A1) how difficult they had experienced the ObR tasks, (A2) how difficult the AcR tasks, (A3) how natural they had experienced the interaction, and (A4) how they rated the quality of the technical realization of the VC's eye movements. In addition, participants were given the chance to respond in open texts relating to: (B1) their assumptions as to the purpose of the study, (B2) anything that bothered them during tasks of both types ObR and AcR, (B3) any strategies they had employed in their attempt to communicate with the other person, (B4) how the naturalness of the interaction could be improved, (B5) whether there was anything else to the experiment which bothered them. The participants' belief in the cover story was further tested in an interview by the experimenter. Participants were asked how well the communication with the partner had worked, whether they had considered what their partner was thinking and whether they had tried to empathize with their partner and whether they had applied specific strategies in their communication with the partner. In addition to the post-experimental questionnaire, participants, either before or after the experiment, also answered a demographic questionnaire and the German version of the autism-spectrum-quotient (AQ; Baron-Cohen et al., 2001). However, for none of the participants AQ results pointed toward autistic symptomatology (cut-off > 32; Baron-Cohen et al., 2001). After the experiment, interview, and questionnaires participants were informed about the nature of the cover story and explained its necessity. Now, participants were asked directly, whether they have had any suspicions as to the nature of the experiment or their partner.

Data Preprocessing and Statistical Analysis

From a total 1200 trials in the ObR condition (25 participants with 48 trials each), 39 trials were excluded due to missing

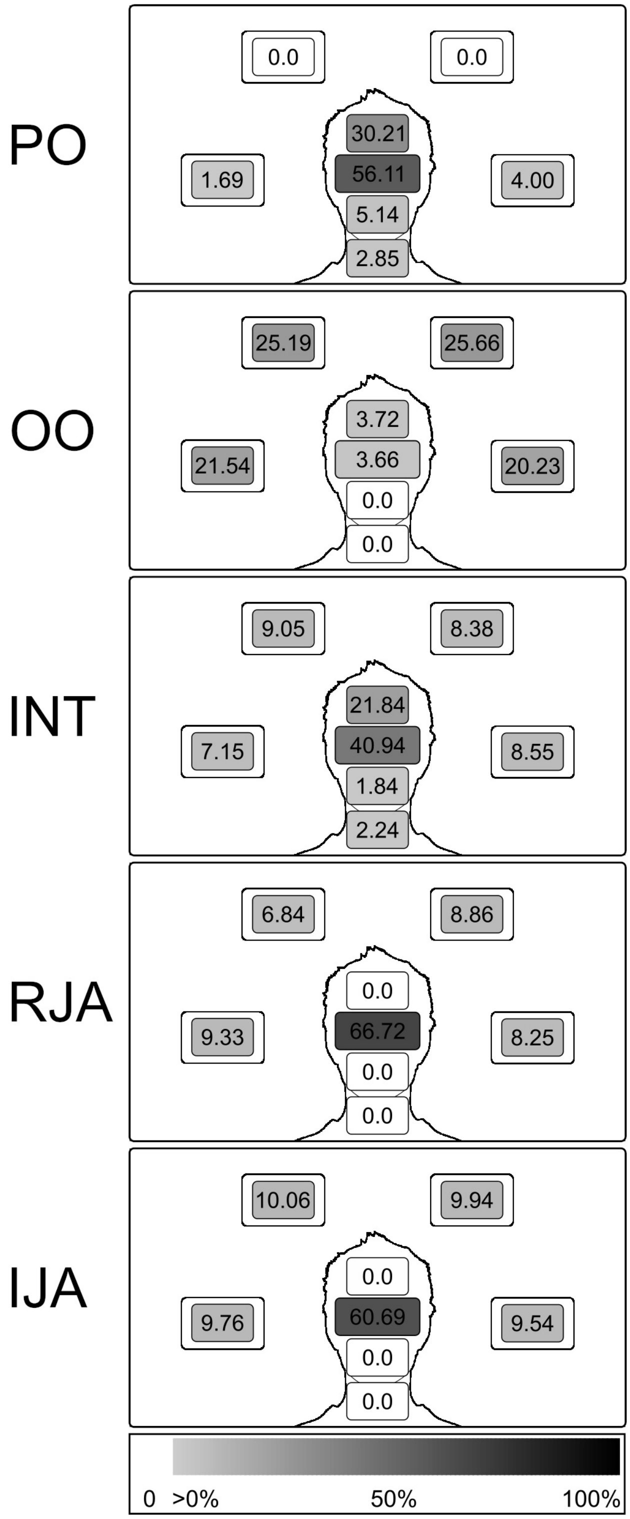


FIGURE 2 | Illustration of the distribution of the agent's visual attention separately cumulated for the different gaze states. Numbers express the rate in percent with which the agent looked at the AoIs in total in the specific state as portion of all fixations, color schemes coding serve as additional illustration (white, AoIs not being targeted; light gray, low rate; black, high rate; see color bar legend at the bottom).

responses or RT exceeding 30 s, another 201 trials were excluded because more than 20% of gaze data were missing due to technical problems, 960 trials remained for analysis. Response, eye-tracking, and questionnaire data were preprocessed and statistically analyzed with R (R Development Core Team, 2008) and RStudio (RStudio Team, 2016). Response and eye-tracking data were analyzed with (generalized) linear mixed effects models, as recommended for data from repeated measures designs (Pinheiro and Bates, 2009), using the lmer() and glmer() function from the lme4 package (Bates et al., 2015). The general influence of predictors was assessed in likelihood ratio tests, comparing how well models including different predictors fit a given data set while taking into account (i.e., penalizing) the models' complexity. The significance of the effect of each predictor was tested by comparing a model comprising the predictor with the same model without the predictor against a significance level of 0.05. Where likelihood ratio tests revealed significant effects of factors, we conducted Tukey *post hoc* tests for the comparison between all individual factor levels (correcting for multiple comparisons) with the glht() function from the multcomp package (Hothorn et al., 2008).

For the analysis of gaze data we computed "relative fixation durations" as the portion of cumulative fixation durations spent on the AoIs "eyes", "face" (not including the eyes), or "objects" (the four objects taken together). Instances of eye contact and joint attention were defined as situations in which the participant and the agent both looked at the eyes of the partner (eye contact) or simultaneously at the same object (joint attention). Two consecutive eye contact or joint attention events on the same object were treated as a single continuous event when they were less than 100 ms apart in order to prevent artificial inflation of events due to eye blinks. Only eye contact and joint attention events with a minimum duration of 50 ms were included in the analysis.

Data from the visual analog scales in the post-experimental questionnaire were summarized as group means. In addition, Spearman correlations between participants' post-experimental self-reports and their task performance were computed. The effect of the participants' age and gender on their responses were analyzed in linear models. Open text responses and statements from the interview were checked for any indications of mistrust in the cover story (e.g., statements indicating lack of conviction to interact with a real person).

RESULTS

Interactivity Ratings

In order to test whether participants were able to correctly identify interactive situations we first compared within ObR the ratings between the non-interactive states (PO, OO, and INT) and the interactive states (RJA and IJA) as a logistic regression with random intercepts for participants. The analysis revealed a highly significant effect on the model fit [$\chi^2(1) = 222.59$, $p < 0.001$]. The chance of being rated as interactive was 27.07% for the non-interactive states and 73.32% for the interactive states, corresponding to a difference in the predicted odds ratio by the factor of 8.45 (M = 2.13, SD = 0.16).

In a next step we looked at the difference between the individual states (**Figure 3A**), again analyzed as logistic regression with random intercepts for participants. A model comprising the agent state as fixed effects fitted the data significantly better than the null model including only the intercept [$\chi^2(4) = 266.70$, $p < 0.001$]. *Post hoc* tests revealed significantly lower ratings for PO vs. INT ($M = -0.86$, $SD = 0.26$, $z = -3.30$, $p = 0.009$), INT vs. RJA ($M = -1.06$, $SD = 0.22$, $z = -4.79$, $p < 0.001$), and RJA vs. IJA ($M = -1.13$, $SD = 0.23$, $z = -4.92$, $p < 0.001$), but not between OO and PO ($M = -0.17$, $SD = 0.28$, $z = -0.60$, $p = 0.975$). Note that for the sake of simplicity we only report comparisons between neighboring ranks when sorted by mean estimates. All other comparisons between states yielded highly significant differences (all $p < 0.001$).

RTs (**Figure 3B**), were logarithmized and again analyzed in a linear mixed effects model with random intercepts for subjects. A group-wise comparison between the interactive and the non-interactive states as fixed effects had no significant effect on the model fit [$\chi^2(1) = 0.36$, $p < 0.55$]. However, including the individual agent states in the model as fixed effects proofed to fit the data significantly better than the null model [$\chi^2(4) = 82.55$, $p < 0.001$]. Corresponding to the results from the interactivity ratings, *post hoc* tests revealed significant differences between OO & PO ($M = -0.18$, $SD = 0.04$, $z = -4.49$, $p < 0.001$), PO & INT ($M = -0.12$, $SD = 0.04$, $z = -2.85$, $p = 0.035$), INT & RJA ($M = 0.22$, $SD = 0.04$, $z = 5.83$, $p < 0.001$), and RJA & IJA ($M = -1.84$, $SD = 0.03$, $z = -5.55$, $p < 0.001$). Note that the differences between OO & INT ($M = -0.30$, $SD = 0.04$, $z = -7.33$, $p < 0.001$), PO & RJA ($M = 0.10$, $SD = 0.04$, $z = 2.748$, $p = 0.048$), and OO and IJA ($M = -0.26$, $SD = 0.04$, $z = -7.17$, $p < 0.001$) also reached significance. In order to investigate whether the quality of the participants' ratings would increase with longer decision time we computed mean correctness scores (RC; correct = "non-interactive" for PO, OO, and INT or "interactive" for RJA and IJA) for each participant. We found a significant relationship between the participants' mean RC and mean RT (r = 0.45, $p < 0.05$). In addition, we analyzed, whether the participants' age or gender had an influence on their decisions. However, neither age nor gender had any significant effect on the mean RCs [age: $\chi^2(1) = 2.21$, $p < 0.151$; gender: $\chi^2(1) = 2.12$, $p < 0.159$] or mean RTs [age: $\chi^2(1) = 0.518$, $p < 0.479$; gender: $\chi^2(1) = 1.43$, $p < 0.245$].

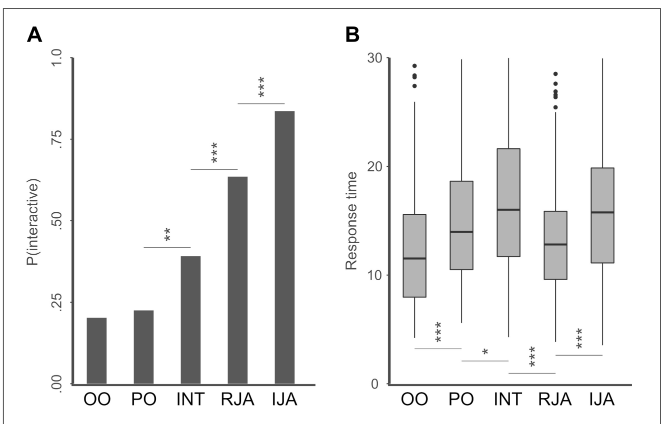


FIGURE 3 | Plots of mean interactivity ratings and mean response times separately for the different gaze states. **(A)** Mean interactivity ratings for different agent states. Asterisks indicate significant differences between neighboring states (when ranked in ascending order) in *post hoc* tests (* <0.05; ** <0.01; and *** <0.001). **(B)** Mean RTs in ms for different agent states. Asterisks indicate significant differences between neighboring states (when ranked in ascending order of mean interactivity ratings) in *post hoc* tests (* <0.05; ** <0.01; and *** <0.001).

Gaze Behavior

For the participants' gaze behavior during ObR, we analyzed the effect of non-interactive vs. interactive states, of the AoIs Eyes, Face and Object and the interaction between states and AoIs on relative durations (proportion of cumulative fixation durations from 0 to 1, **Figure 4A**). Tests did not reveal significant improvements in model fit for including states [$\chi^2(1) = 0.00$, $p = 0.994$] but for AoI [$\chi^2(2) = 948.37$, $p < 0.001$], and the interaction of state*AoI [$\chi^2(2) = 12.40$, $p = 0.002$]. A *post hoc* test between factor combinations was conducted in order to identify effects potentially driving the interaction. However, corrected for multiple testing, the comparisons between non-interactive and interactive states did not reveal any significant differences for the AoIs Eyes ($M = -0.03$, $SD = 0.02$, $z = -1.80$, $p = 0.467$), Face ($M = -0.03$, $SD = 0.02$, $z = -1.64$, $p = 0.565$), or Objects ($M = -0.04$, $SD = 0.07$, $z = -2.58$, $p = 0.102$).

The effect of a non-interactive vs. interactive agent on the number of instances of eye contact (**Figure 4B**) and joint attention (**Figure 4C**) per trial was analyzed in generalized mixed effects models for Poisson distributed data. Including the interactivity of the agent significantly increased model fits for the prediction of the amount of eye contact [$\chi^2(1) = 68.19$, $p < 0.001$] as well as the amount of joint attention instances [$\chi^2(1) = 72.75$, $p < 0.001$]. When the agent behaved interactively, the occurrence of eye contact instances increased by a factor of 1.31 ($M = 0.27$, $SD = 0.03$) and the occurrence of joint attention instances increased by a factor of 1.52 ($M = 0.42$, $SD = 0.05$).

We then analyzed whether the occurrence of instances of eye contact (**Figure 4D**) or joint attention (**Figure 4E**) had a predictive value for the participants' subsequent interactivity rating and whether the prediction would differ depending on the agent behaving either non-interactively or interactively. To this end, we compared linear mixed effects models including the agents' interactivity, the number of instances of eye contact or joint attention, respectively, as well as the interaction between both. All three, the inclusion of the agents' interactivity [$\chi^2(1) = 222.57$, $p < 0.001$], the inclusion of the number of eye contact instances [$\chi^2(1) = 14.86$, $p < 0.001$], as well as the interaction between both [$\chi^2(1) = 9.52$, $p = 0.002$], and significantly improved model fits. The predicted probability of the agents′ behavior being rated as interactive increased with the number of eye contact instances ($M = 0.05$, $SD = 0.03$), but this effect was especially strong when the agent actually behaved interactively ($M = 0.15$, $SD = 0.05$). For the analysis of the effect

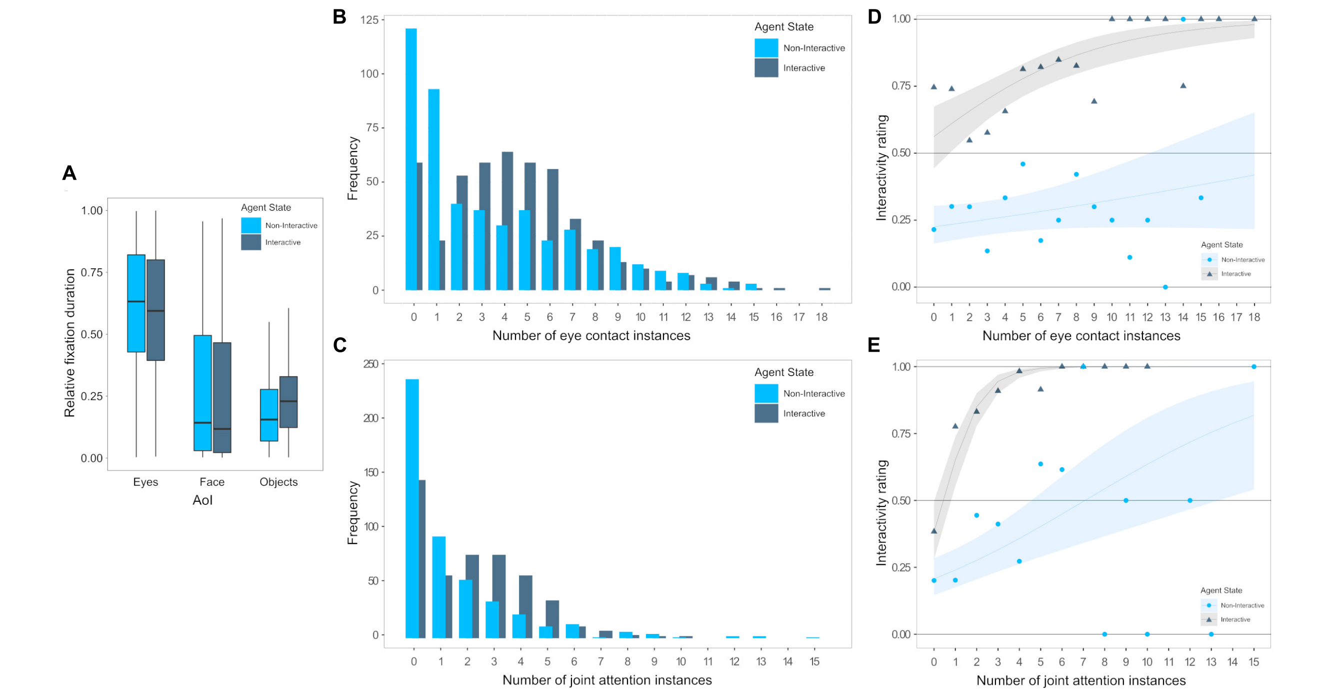


FIGURE 4 | Illustration of the participants gaze behavior and instances of eye contact and joint attention between participant and agent in connection to the participant's rating of the agents interactivity, separately for an agent behaving non-interactively (light blue) vs. interactively (dark blue). **(A)** Boxplots of relative fixation durations as the portion of time spent on the AoIs Eyes, face, and objects per trial. **(B)** Frequencies of eye contact instances per trial. **(C)** Mean rates (circles and triangles) and model predictions with 95% confidence intervals (lines and ribbons) of interactivity ratings for differing numbers of eye contact instances per trial. **(D)** Frequencies of joint attention instances per trial. **(E)** Mean rates (circles and triangles) and model predictions with 95% confidence intervals (lines and ribbons) of interactivity ratings for differing numbers of joint attention instances per trial.

of joint attention, again, the inclusion of the agents' interactivity [$\chi^2(1) = 222.59$, $p < 0.001$], the inclusion of the number of joint attention instances [$\chi^2(1) = 96.54$, $p < 0.001$], as well as the interaction between both [$\chi^2(1) = 73.16$, $p < 0.001$], significantly improved model fits. Accordingly, the predicted probability of the agents′ behavior being rated interactive increased with the number of joint attention instances ($M = 0.19$, $SD = 0.05$) with an even stronger effect when the agent actually behaved interactively ($M = 0.92$, $SD = 0.12$).

Questionnaires and Post-experimental Inquiry

In the post experimental inquiry participants reported on the perceived difficulty of the ObR task ($M = 2.80$, $SD = 1.38$) and the AcR task ($M = 1.76$, $SD = 0.72$), the quality of the technical implementation of the agents′ eye movements ($M = 3.21$, $SD = 0.88$), and the naturalness of the interaction ($M = 2.96$, $SD = 1.30$). We compared ratings of the task difficulty to the participants' mean tendency to experience the agent as interactive, their mean performance (response correctness) as well as mean RTs. Difficulty ratings neither correlated significantly with the participants' tendency to rate the agent's behavior as interactive ($rs = -0.07$, $p > 0.05$) nor with their response correctness ($rs = 0.02$, $p > 0.05$) nor with RTs ($rs = -0.24$, $p > 0.05$).

In order to assess effects of autistic traits we compared models comprising and not comprising the AQ scores as predictor. Neither including the quotient as main effect [$\chi^2(1) = 0.98$, $p < 0.323$] nor as interaction with interactive vs. non-interactive states [$\chi^2(1) = 0.27$, $p < 0.607$] significantly improved model fits for mean interactivity ratings. Similarly, for mean RTs, neither including the quotient as main effect [$\chi^2(1) = 0.45$, $p < 0.50$] nor as interaction with interactive vs. non-interactive states [$\chi^2(1) = 0.01$, $p < 0.908$] significantly improved model fits.

None of the answers to the written open text questions indicated any suspicions about the cover story or any awareness of deceit. In the interview, two participants indicated that during the experiment they developed the suspicion or had asked themselves whether they actually had interacted with the partner they previously had met (both participants were excluded from further analysis, see above).

DISCUSSION

This study focuses on the question whether and how humans are able to recognize interactivity in triadic interactions. To this extent, we gave our participants two tasks, one in which participants had to observe and recognize gaze states (ObR) and one in which they had to engage in different gaze states (AcR). While the former condition was the actual target condition and basis for the analysis, the latter was necessary to maintain the semblance of a balanced study design suggested by the cover story. As our main result, we can show for the first time that human participants are perfectly able to use gaze cues to judge interactivity by spotting the contingencies between their own and the agents' behavior without any explicit instructions how to do that. In the analysis of the interactivity ratings, we found that participants consistently and successfully discriminated between interactive and non-interactive states. These findings empirically substantiate the hypothesis of gaze communication being a precursor of human cooperation (Moll and Tomasello, 2007; Tomasello et al., 2007). Findings from phylogenetic and ontogenetic studies support this notion by showing that attending to eyes and communicating via gaze are pivotal steps toward higher levels of social cognition (Tomasello and Carpenter, 2005; Tomasello et al., 2007; Grossmann, 2017). So far, however, these proposals have been hypothetical, i.e., based on phylogenetic and evolutionary considerations. Here, we can explicitly show that gaze is sufficient for humans to establish the experience of mutual interaction as a prerequisite for building social relationships.

We also found differences in the interactivity ratings within interactive-states and within non-interactive states suggesting considerable sensitivity to variations in the tempo-spatial parameters of perceived gaze behavior. Our expectation that a gaze following agent would more easily elicit the experience of interactivity was not confirmed. This hypothesis was based on the assumption that actively following an initiating agent would be more demanding than being followed by a responding agent. Earlier studies had shown that humans innately expect gaze following (Pfeiffer et al., 2011) and perceive the initiation of joint attention as rewarding (Schilbach et al., 2010; Pfeiffer et al., 2014; Oberwelland et al., 2016). However, the present data suggest that agents who initiate joint attention are significantly more readily experienced as interactive than a merely gaze following agent. This might be explained by the fact that responding to joint attention bids might be considerably easier than to actively initiate joint attention. This interpretation is in accordance with phylogenetic and ontogenetic findings suggesting that IJA requires more complex cognition as compared to RJA. For example, chimpanzees are able to follow someone's gaze but do not initiate joint attention themselves (Tomasello and Carpenter, 2005). Human children acquire the basis of RJA from the early age of 6 month in comparison to the initiation of attention which does not occur before the second year of life (Mundy and Newell, 2007; Mundy et al., 2007).

The non-interactive states OO and PO were significantly more often identified correctly as non-interactive than the INT state. During OO the agent was mainly focused on the objects and looked at the participant only to a lesser extent. Humans are typically very sensitive to how other persons explore and behave in a shared environment. Our perception and processing of objects seem to be fundamentally altered when we observe other person attending to them (Becchio et al., 2008). Objects subsequently appear more familiar (Reid et al., 2004; Reid and Striano, 2005) and likeable (Bayliss and Tipper, 2006; Bayliss et al., 2006). Our results suggest that despite such effects, we are still able to discern that the behavior we observe is not related to us or at least not aimed at us. The same might be true for the PO state. Contrary to our prior hypothesis, participants did not report the PO agent as more interactive than OO, notwithstanding the higher chances of eye contact in these situations due to the agent more frequently looking at the participant. The instructions defined an interaction in terms of mutual and reciprocal responses between both partners. Low interactivity ratings for PO might therefore be just a sign for the participants' adherence to the instructions instead of disclosing their intuitive, subjective definition of an interaction. Despite that, participants were able to differentiate between an active, reciprocal interaction and person-focused but passive visual attention. This is in line with findings showing that humans are very sensitive to differences in the interactional affordance in the context of more pronounced contrast between encountering real persons as compared to facing static pictures (Hietanen et al., 2008; Pönkänen et al., 2011).

In our experimental setup, INT appears to be the most ambiguous of all states, receiving almost as many interactive as non-interactive ratings. The inward directed attention and thus absence of any obvious attentional focus in the environment probably made it impossible to attribute intentions of interaction. In other words, gaze alone is no longer informative as soon as the interaction partner is in a state of introspection or mind-wandering (see section "Limitations").

In order to better understand the emergence of the experience of interaction, we analyzed the relationship between the gaze behavior of the participants and the agent's behavior. We did not find any effect of the agents' intended interactivity of the encounter on the distribution of the participants visual attention between objects and agent. However, when looking at the synchronization with the agent's behavior, we found an increase in the number of eye contact instances and joint attention instances in interactive as compared to non-interactive states. Thus, one of the participants' strategies to judge upon interactivity might have been based on the frequency of eye contact and joint attention instances. The analysis of the effect of the number of eye contact and joint attention instances on the participants' decisions revealed significant differences between non-interactive and interactive encounters. Importantly, during interactive encounters, the emergence of eye contact and joint attention had much higher effects on the subsequent interactivity ratings. One plausible interpretation could be that participants "tested" the agents' reciprocity by attempting to establish eye contact and joint attention and subsequently assessing whether the timing of resulting joint contingencies could be attributed to an interacting agent that takes into account the gaze behavior of the participant. Considering the importance of fine-grained timing during such gaze-based interactions it is plausible that the emergence of interactivity is deeply embedded in the temporal enfolding of gaze-based encounters and can only be experienced over time. This is in line with the understanding that non-verbal communication is a dynamic and continuous process (Burgoon et al., 1989) that cannot be fully comprehended through the passive observation of discrete events, uncoupled from the flow of communication.

With respect to the differences in the duration of the decision between the different conditions, we found a correlation between the mean RT of participants and mean correctness scores, suggesting that participants who invested more time were able to make better informed decisions. When comparing RT between states on a single trial level, RTs in non-interactive states showed a pattern roughly corresponding to that of the correctness scores. i.e., RTs reflected the ambiguity and associated difficulty to judge the interactivity. When comparing the participants' reactions to RJA vs. IJA agents we found longer RTs for the more unequivocal IJA state (as reflected in higher interactivity ratings). One explanation might be that participants needed more time to identify this maximal complex state.

Previous studies about social gaze, even those employing gaze-contingent interactive paradigms, were mostly based on a trial structure that sharply restricted the interaction to a few seconds (Wilms et al., 2010; Pfeiffer et al., 2011, 2012; Oberwelland et al., 2016, 2017). Our findings suggest that such short time intervals are probably not sufficient to establish the full experience of interaction during a spontaneous encounter. Earlier studies circumvented this problem by focusing on "atomic" elements of interaction using an exactly predefined time course of specific behavioral elements and explicitly instructing participants. However, this restriction is not compatible with the implicit and dynamical character of social interactions and thus threatens ecological validity (Risko et al., 2012, 2016; Pfeiffer et al., 2013a; Schilbach et al., 2013).

Overcoming this problem required both theoretically and methodologically new approaches. From a theoretical perspective the SGS provides the holistic framework that is able to encompass and describe the entire span of possible interactive states (Jording et al., 2018). Methodologically this study profits from the development of the new agent-platform TriPy that implements the states of the SGS and allows for a degree of interactional freedom not available with previous setups (Hartz et al., submitted). In combination, these developments allowed us for the first time to investigate the unfolding of a purely gaze based interaction.

Limitations

Several limitations with respect to the study design need to be considered when interpreting the results. First, we deliberately focused on gaze and restricted all communication to this particular important non-verbal communication channel. The availability of additional channels would certainly have facilitated the establishment of interactions in this study, resulting in more decisive, and faster interactivity ratings. However, the goal of this study was to test explicitly the potential of gaze communication to establish interactions in a way that results can inform studies about non-verbal multi-channel communication. Furthermore, we aimed at studying the individual characteristics of predefined states of gaze interactions and therefore chose a design where the agents displayed only one state at a time. Based on these results it would now be interesting to investigate how transitions between these states might take place (Jording et al., 2018). Therefore, sampling experiences of participants at random time points in an interaction with an agent who dynamically transitions from one state to another might constitute a promising approach.

We did not aim for the systematic investigation of effects of inter-individual differences during the establishment of gaze interactions and while we included a broad age range, we did not balance our sample with regard to gender. In addition, we only used one VC with a male, middleaged appearance and did not systematically match age and gender between participants and agent. Although we did not find any significant effects of age or gender on the quality or timing of the participants' ratings, we cannot rule out the possibility of any influence. Further investigations controlling for the participants' age and gender distribution and a systematic matching between participants and agents are required to elucidate this question.

Conclusion

Results indicate that humans are able to establish gaze interaction without any instructions or additional communication channels, supporting theoretical assumptions of the fundamental role of gaze communication in the development of human social behavior. Our data suggest that human participants are able to identify interactivity not only based on passive observation but potentially by actively studying the agents' responsiveness based on successfully established mutual eye contact and joint attention. However, participants were not only able to distinguish interactive and non-interactive situations, but behavioral differences between the non-interactive states elicited differential experiences of the interaction. Interestingly, the participants' performance did not predict their post-experimental assessment of the tasks difficulty. This suggests that decisions were based on intuition or at least partly beyond conscious processing, which corresponds to the presumably implicit and automatic character of non-verbal communication (Choi et al., 2005). An intriguing next step would now be to integrate additional non-verbal communication channels, potentially in a more immersive environment (e.g., a virtual reality), or to investigate the establishment of interactions in cases of impaired communication abilities as in autism spectrum conditions.

DATA AVAILABILITY

The datasets for this manuscript are not publicly available because the supervising ethics committee has not yet approved the publication of the raw data. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Principles of Good Scientific Practice, ethics committee of the Medical Faculty of the University of Cologne, Germany, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee of the Medical Faculty of the University of Cologne, Germany.

AUTHOR CONTRIBUTIONS

All authors substantially contributed to the conception of the work. AH, MJ, KV, and MS-R designed the study protocol. AH implemented the paradigm code. MJ conducted the pilot study and the main experiment. MJ and AH analyzed the data. MJ drafted the manuscript. AH, GB, KV, and MS-R critically revised the manuscript.

FUNDING

This study was received funding from the Federal German Ministry of Education and Research (Grant Numbers: 16SV7242 to MS-R and 16SV7244 to KV), the Excellence Initiative of the German federal and state governments (MS-R), and the EC, Horizon 2020 Framework Program, FET Proactive (Grant Agreement Id: 824128 to KV).

ACKNOWLEDGMENTS

The authors want to express their gratitude to Björn Guth for his contributions to the development of the experimental code.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01913/full#supplementary-material

REFERENCES

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

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a "Language of the Eyes"? evidence from normal adults, and adults with autism or asperger syndrome. *Vis. Cogn.* 4, 311–331. doi: 10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bayliss, A. P., Paul, M. A., Cannon, P. R., and Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: i like what you look at. *Psychon. Bull. Rev.* 13, 1061–1066. doi: 10.3758/BF03213926

Bayliss, A. P., and Tipper, S. P. (2006). Predictive gaze cues and personality judgments: should eye trust you? *Psychol. Sci.* 17, 514–520. doi: 10.1111/j.1467-9280.2006.01737.x

Becchio, C., Bertone, C., and Castiello, U. (2008). How the gaze of others influences object processing. *Trends Cogn. Sci.* 12, 254–258. doi: 10.1016/j.tics.2008.04.005

Burgoon, J. K., Buller, D. B., and Woodall, W. G. (1989). *Nonverbal Communication: The Unspoken Dialogue*. New York, NY: Harper & Row.

Choi, Y. S., Gray, H. M., and Ambady, N. (2005). "The glimpsed world: unintended communication and unintended perception," in *The New Unconscious*, eds R. R. Hassin, J. S. Uleman, and J. A. Bargh (New York, NY: Oxford University Press), 309–333.

Dalmaijer, E. S., Mathôt, S., and Van der Stigchel, S. (2014). PyGaze: an open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behav. Res. Methods* 46, 913–921. doi: 10.3758/s13428-013-0422-422

Dalmaso, M., Castelli, L., Scatturin, P., and Galfano, G. (2017). Trajectories of social vision: eye contact increases saccadic curvature. *Vis. Cogn.* 25, 1–8. doi: 10.1080/13506285.2017.1290727

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-27

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9602–9605. doi: 10.1073/pnas.152159999

George, N., and Conty, L. (2008). Facing the gaze of others. *Neurophysiol. Clin. Neurophysiol.* 38, 197–207. doi: 10.1016/j.neucli.2008.03.001

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., and Vogeley, K. (2014). The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Front. Hum. Neurosci.* 8:807. doi: 10.3389/fnhum.2014.00807

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

Grossmann, T. (2017). The eyes as windows into other minds: an integrative perspective. *Perspect. Psychol. Sci.* 12, 107–121. doi: 10.1177/1745691616654457

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

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Jarick, M., and Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Front. Psychol.* 6:1423. doi: 10.3389/fpsyg.2015.01423

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2018). The "Social Gaze Space": a taxonomy for gaze-based communication in triadic interactions. *Front. Psychol.* 9:226. doi: 10.3389/fpsyg.2018.00226

Moll, H., and Tomasello, M. (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 639–648. doi: 10.1098/rstb.2006.2000

Morales, M., Mundy, P., and Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behav. Dev.* 21, 373–377. doi: 10.1016/S0163-6383(98)90014-90015

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., and Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Dev.* 78, 938–954. doi: 10.1111/j.1467-8624.2007.01042.x

Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. *Curr. Dir. Psychol. Sci.* 16, 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., et al. (2016). Look into my eyes: investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *Neuroimage* 130, 248–260. doi: 10.1016/j.neuroimage.2016.02.026

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., et al. (2017). Young adolescents with autism show abnormal joint attention network: a gaze contingent fMRI study. *Neuroimage Clin.* 14, 112–121. doi: 10.1016/j.nicl.2017.01.006

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinformatics* 2:10. doi: 10.3389/neuro.11.010.2008

Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., and Vogeley, K. (2012). Eyes on the mind: investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Front. Psychol.* 3:537. doi: 10.3389/fpsyg.2012.00537

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., et al. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *Neuroimage* 101, 124–137. doi: 10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., and Schilbach, L. (2011). A Non-Verbal turing test: differentiating mind from machine in gaze-based social interaction. *PLoS One* 6:e27591. doi: 10.1371/journal.pone.0027591

Pfeiffer, U. J., Timmermans, B., Vogeley, K., Frith, C. D., and Schilbach, L. (2013a). Towards a neuroscience of social interaction. *Front. Hum. Neurosci.* 7:22. doi: 10.3389/fnhum.2013.00022

Pfeiffer, U. J., Vogeley, K., and Schilbach, L. (2013b). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev.* 37, 2516–2528. doi: 10.1016/j.neubiorev.2013.07.017

Pinheiro, J. C., and Bates, D. M. (2009). *Mixed-Effects Models in S And S-PLUS*, Reprinted paperback ed. of the 2000 Edn. New York, NY: Springer.

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., and Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Soc. Cogn. Affect. Neurosci.* 6, 486–494. doi: 10.1093/scan/nsq068

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reid, V. M., and Striano, T. (2005). Adult gaze influences infant attention and object processing: implications for cognitive neuroscience. *Eur. J. Neurosci.* 21, 1763–1766. doi: 10.1111/j.1460-9568.2005.03986.x

Reid, V. M., Striano, T., Kaufman, J., and Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport* 15, 2553–2555. doi: 10.1097/00001756-200411150-200411125

Risko, E. F., and Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Atten. Percept. Psychophys.* 73, 291–296. doi: 10.3758/s13414-010-0042-41

Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., and Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Front. Hum. Neurosci.* 6:143. doi: 10.3389/fnhum.2012.00143

Risko, E. F., Richardson, D. C., and Kingstone, A. (2016). Breaking the fourth wall of cognitive science: real-world social attention and the dual function of gaze. *Curr. Dir. Psychol. Sci.* 25, 70–74. doi: 10.1177/0963721415617806

RStudio Team (2016). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22, 2702–2715. doi: 10.1162/jocn.2009.21401

Senju, A., and Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Curr. Biol.* 18, 668–671. doi: 10.1016/j.cub.2008.03.059

Senju, A., and Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Vis. Cogn.* 12, 127–144. doi: 10.1080/13506280444000157

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13, 127–134. doi: 10.1016/j.tics.2008.11.009

Tomasello, M., and Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees. *Monogr. Soc. Res. Child Dev.* 70, 107–122. doi: 10.1111/j.1540-5834.2005.00324.x

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

Vernetti, A., Smith, T. J., and Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proc. R. Soc. B Biol. Sci.* 284:20162747. doi: 10.1098/rspb.2016.2747

Vogeley, K., and Bente, G. (2010). "Artificial humans": psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Netw.* 23, 1077–1090. doi: 10.1016/j.neunet.2010.06.003

Wilms, M., Schilbach, L., Pfeiffer, U. J., Bente, G., Fink, G. R., and Vogeley, K. (2010). It's in your eyes–using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Soc. Cogn. Affect. Neurosci.* 5, 98–107. doi: 10.1093/scan/nsq024

Wu, D. W.-L., Bischof, W. F., and Kingstone, A. (2013). Looking while eating: the importance of social context to social attention. *Sci. Rep.* 3:2356. doi: 10.1038/srep02356

Yarbus, A. L. (1967). *Eye Movements and Vision*. Boston, MA: Springer, doi: 10.1007/978-1-4899-5379-7

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

Reduced experience of interactivity during gaze-based interactions in participants with autism spectrum disorder

Mathis Jording[1,2*], Arne Hartz[3,4], David H. V. Vogel[1,2], Martin Schulte-Rüther[1,3,4,5], Kai Vogeley[1,2]

[1]Cognitive Neuroscience (INM-3), Institute of Neuroscience and Medicine, Research Center Juelich, Germany
[2]Department of Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany
[3]JARA-BRAIN, Aachen, Germany
[4]Translational Brain Research in Psychiatry and Neurology, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, University Hospital RWTH Aachen, Germany
[5]INM-11, Institute of Neuroscience and Medicine, Research Center Juelich, Germany

Keywords: Autism spectrum disorder, Social gaze, Joint attention, Eye contact, Triadic interaction, Nonverbal communication, Social cognition and social behavior, Human-agent interaction

Abstract

The impairment to adequately perceive social cues and infer mental states is a key feature of autism spectrum disorder (ASD) that can prevent afflicted persons from successfully navigating the social world. Making use of a recently developed human-agent interaction setup, we investigated the ability to infer mental states in ongoing gaze-based encounters. Results show that participants with ASD were impaired in ascertaining whether their partner was trying to interact with them or not as compared to participants without ASD. However, the distribution of visual attention did not differ between the two groups and both reliably established eye contact and joint attention. This suggests that the mere detection of social cues is preserved in ASD while their evaluation is impaired.

Introduction

One of the core symptoms of Autism Spectrum Disorder (ASD) are disturbances in interaction and communication abilities (American Psychiatric Association, 2013; World Health Organization, 1993). Persons with ASD have difficulties taking the perspective of others or understanding their mental states (Frith, 1996), they seem to be less interested in social situations and less motivated to engage in social interactions (Chevallier et al., 2012). Nonverbal communication appears to be especially impaired, with respect to both production and perception of nonverbal cues. In gaze communication, anomalies were found at different stages during the establishment of an interaction in ASD. However, reports are partially inconsistent and the great variety of potentially affected areas has made it difficult to outline the problem and assess the consequences for the (social) life of affected persons.

A strong indicator of general and pervasive alterations in gaze communication in ASD are reports of impairments in the ability to infer mental states from the eye region of others (Baron-Cohen et al., 2001a, 1997). Being able to infer e.g. the intentions of others is a key element in understanding their behavior and responding to them appropriately. However, previous studies on the ascription of mental states were based on static images and did not involve any ongoing eye movements (Baron-Cohen et al., 2001a, 1997). Many other studies investigating social gaze in persons with ASD focused on visual attention or gaze behavior but did not explicitly test whether participants were able to infer more complex mental states on the basis of gaze signals. Thus, the extent of impairments in inferring mental states in ASD and whether they can be also observed in interactive situations which are closer to everyday life is as of yet unclear. A follow-up question would be whether impairments in inferring mental states in ASD also translate to differences in their gaze behavior and their interaction with their partner.

Traditionally, individuals with ASD have been reported to focus less on socially relevant stimuli or, in case of looking at human faces, focus more on the mouth region than the eyes (Dalton et al., 2005; Klin et al., 2003, 2002b). Consequently, learning opportunities in socio-communicative domains would be decreased and the inference of mental states from gaze would be impeded (Chevallier et al., 2013, 2012). However, more recently, the notion of generally reduced visual attention for socially relevant stimuli in ASD has been challenged (Chita-Tegmark, 2016; Falck-Ytter & von Hofsten, 2011; Frazier et al., 2017; Guillon et al., 2014). It seems as if differences between affected and non-affected persons were most prominent in or even restricted to ongoing dynamic, socially enriched encounters. The attention to social stimuli such as faces is substantially more reduced when confronted with dynamic rather than static stimuli (Speer et al., 2007) and for faces depicted in interaction rather than isolated faces (Klin et al., 2002a; Riby & Hancock, 2008). While listening to a conversation partner, adults with ASD looked less at the partner when the partner herself looked back at them (Freeth & Bugembe, 2019). Interestingly in children with ASD even the social content of a conversation might be relevant to their gaze behavior with seemingly more pronounced avoidance of eye contact when talking about 'how people feel' compared to 'what people do' (Hutchins & Brien, 2016). Conversely, several studies, especially those using static or non-interactive stimuli, were not able to replicate earlier findings about reduced visual attention to socially relevant aspects (cf.: Chita-Tegmark, 2016; Falck-Ytter & von Hofsten, 2011; Frazier et al., 2017; Guillon et al., 2014). In summary, the extent and the circumstances of a reduction of visual attention for social stimuli in ASD cannot yet be fully assessed. Thus, the potential impact on the ability to infer mental states is unknown but is expected to be more prominent for less restricted and more interactive situations.

Aside from visual attention, another possible origin of difficulties in inferring mental states might lie in active participation in interactions. In typical development, a close ontogenetic link exists between gaze following skills and the development of ToM, i.e. mentalizing (Morales et al., 1998). In ASD, reflexive gaze following seems to be comparably intact (Chawarska et al., 2003; Kylliainen & Hietanen, 2004), but the initiation of joint attention has been reported to be impaired in children with ASD (MacDonald et al., 2006; Mundy, 2003; Mundy et al., 1994; Whalen & Schreibman, 2003). Therefore, it is worth investigating whether potential difficulties in the inference of mental states in ASD are reflected in the dynamics that unfold between interactants in a preceding gaze encounter.

On this background, this study aims at investigating the relationship between the dynamic behavior of two interactants and the inference of mental states. To this end, several technical and methodological challenges need to be overcome. Ongoing interaction processes are characterized by highly complex dynamics of nonverbal communication (Burgoon et al., 1989; Krämer, 2008) which need to be taken into account to ensure ecological validity (Pfeiffer et al., 2013; Risko et al., 2012; Schilbach et al., 2013). With a focus on gaze behavior, the recently developed taxonomy of the "Social Gaze Space" (SGS; Jording et al., 2018), serves as the theoretical foundation for this study. It covers all possible states of attention in triadic interactions (constituted by two interactants and one or more objects in a shared environment): 'partner-oriented' (attention directed towards the partner); 'object-oriented' (attention directed towards objects; 'introspective' (attention directed towards inner (bodily) experiences and disengaged from the outside world); 'responding joint attention' (active following of the partner's gaze); 'initiating joint attention' (proactive attempts to lead the partner's gaze towards objects of one's own choice). Only the latter two states are truly interactive in the sense that the behavior of one partner is contingent upon the behavior of the other. Based on this conceptual framework a new experimental approach was developed that allows for modelling all five different gaze states: the agent-platform "TriPy" (described in detail in: Hartz et al., submitted; Jording et al., 2019). TriPy processes eye-tracking data and controls an anthropomorphic agent in real-time to create a gaze-contingent interaction partner.

With this platform, we studied the inference of mental states in persons with ASD as compared to control persons during gaze-contingent social encounters. We focused on the most important research question: Do persons with ASD recognize the attempt of the partner to interact on the basis of gaze behavior? Participants with and without ASD were instructed to judge, whether a partner, displayed as a virtual character (VC), was trying to interact with them or not. The VC was a fully algorithm-controlled agent and behaved either in an interactive or in a non-interactive manner. In interactive states, the agent's behavior was contingent upon the participant, i.e. the agent was either responsive and followed the participant's gaze or tried to lead the interaction and make the participant follow the gaze of the agent. In non-interactive states the agent freely observed the environment and the participant, but its behavior was not influenced by the participant. Participants believed that the agent would represent another participant (i.e. confederate). We analyzed the participants' decisions, the distribution of visual attention, as well as the number of eye contacts and joint attention instantiations. Our hypotheses were that participants with ASD i) would be impaired in their ability to recognize the partner's intention to establish interaction(s), ii) would pay less visual attention to socially relevant elements such as the agent's eyes, iii) would be less responsive to the agent as compared to control participants, resulting in fewer eye contacts and joint attention events. In addition to this hypothesis-driven approach we analyzed the relationship between the occurrence of eye contact and joint attention during an encounter and the participant's subsequent decision in an explorative fashion.

Methods

All methods and procedures summarized below are described in full detail in Jording et al. (2019).

Participants

26 subjects with ASD, diagnosed in the Autism Outpatient Clinic at the […], were recruited via the Outpatient Clinic. The diagnostic procedure started by a screening with the Autism-Spectrum-Quotient (Baron-Cohen et al., 2001b) and was only continued when patients exceeded the cut-off value (>32). The diagnosis then had to be confirmed in two independent and extensive clinical interviews by two separate professional clinicians according to ICD-10 criteria (World Health Organization, 1993). After the exclusion of 5 participants (due to missing data or mistrust of the cover story) the remaining 21 subjects (6 identifying as female, 15 as male; aged 22 to 54, mean = 40.86, sd = 10.36) were compared to a group of control subjects (with an overlap to the population reported in Jording et al., 2019), without any record of psychiatric or neurological illnesses (10 identifying as female, 14 as male; aged 23 to 58, mean = 39.00, sd = 12.76). Demographic data and the Autism-Spectrum-Quotient (AQ; Baron-Cohen et al., 2001b) were obtained. None of the participants from the control group exceeded the commonly preferred cut-off of > 32, while all participants from the ASD group did, indicating a clear difference in the expression of autistic symptoms between both groups. Both groups of participants had comparable educational backgrounds in terms of years of education (ASD: mean = 17.77, sd = 5.93; Controls: mean = 16.27, sd = 4.28), impairments in intelligence were ruled out based on clinical interviews and educational background. All participants gave their written consent prior to participation and received a monetary compensation (10€ per hour). This study was presented to and approved by the ethics committee of […] and strictly adhered to the declaration of Helsinki and the principles of good scientific practice.

Procedure and tasks

To make the subjects believe that they were participating in an ongoing social encounter, they were introduced to a confederate of the same sex in a briefing room prior to the start of the experiment. Participants were informed that they would communicate with their interaction partner via a computer, with their partner being seated in a different room. After this introduction, participants were separated from the confederate and brought to the testing room, where they received detailed written and oral experimental instructions. Participants were informed that both partners were to be represented on their computer screen by an identical generic male VC serving as an avatar of the partner and that during the interaction they would see their partner's avatar instead of her/his real face. Furthermore, they would only be allowed to communicate with their partner via gaze, while all other communication channels (e.g. speech, gestures, facial expressions) would not be transmitted or displayed. Importantly, the avatar displayed on the participants' screen was always being controlled by the computer algorithm (Figure 1a; Jording et al., 2019). In addition to the avatar, four trial-wise changing objects were displayed on the screen at fixed locations in the avatar´s field of view (Figure 1b).

In full accordance with Jording et al. (2019), participants had two tasks or roles, the Observation-Role (ObR) and the Action-Role (AcR), grouped into alternating blocks. In total, three blocks were presented per role with 16 (ObR) or 10 (AcR) trials in randomized order per block. Trials were separated by short breaks of 2 – 6 s, blocks were separated by breaks of approx. 3 min allowing a short phase of rest for the participant and re-calibration of the eye-tracker. In the ObR condition participants had to ascertain by button-press whether their partner was trying to interact with them or not. The ObR was the primary target condition of our analysis with the main focus on the participants' responses ("interactive" or "not interactive") and the respective preceding gaze behavior. Trials lasted until participants' response but maximally 30 s. During AcR, for each trial one out of (equating to the five states of the SGS; Jording et al., 2018, 2019) was pseudo randomly chosen, and participants were asked to 1: concentrate on their partner, 2: concentrate on the objects; 3: concentrate on their breathing; 4: to let the partners gaze guide them; or 5: to guide the partner with their gaze. Each trial lasted 30 s. The main purpose of the AcR condition was solely to support the cover story by suggesting a balanced study design with the same tasks for both participants. Data of AcR were not analysed.

The experiment was followed by a post-experimental questionnaire asking participants on 6-point scales about the difficulty of the two tasks, as well as the naturalness of interaction and the quality of technical realization of the VC's eye movements. Additional open text items asked for the participants' experience during the experiment as well as their assumptions about the purpose of the study. An additional interview by the experimenter inquired whether participants believed the cover story. Answers to the written open text questions and during the post-experimental interview were carefully screened for indications of mistrust in the cover story (e.g. indicating a lack of conviction to having interacted with a real person). While none of the written answers indicated any such suspicions, in the interview two participants indicated that they had at least questioned the announcement to have interacted with a real human. Both participants were excluded from further analysis (see above).

Setup, agent-platform and pilot study

The agent's behavior and graphical output was controlled by the newly developed agent-platform 'TriPy' already applied before (Jording et al., 2019). TriPy adapts the behavior of a VC to the behavior of the participant in real-time ("gaze-contingent"; Schilbach, 2010; Wilms et al., 2010). While earlier studies relied on pre-determined behavior, in TriPy the agents' behavior in the non-interactive states is implemented on a probabilistic basis. For the purpose of this study, we implemented all five states of the SGS (with behavioral parameters being empirically informed by a pilot study; Jording et al., 2019) . In the non-interactive states, the agent was not responsive to the participant (although occasionally looking at the participant with incidental eye contact). In the interactive states, the agent either followed or "responded" to the participants gaze or tried to lead or "initiate" the participants gaze[1]. During the ObR condition, the number of interactive and non-interactive states were balanced with an occurrence of 50% each. Whenever the participants were in interactive-states during the AcR mode, the virtual agent reacted with the corresponding behavior (RJA with IJA; IJA with RJA) and during non-interactive states with another non-interactive state with all combinations of agent and participant states appearing equally often. The eye-tracker ran at a sampling rate of 120 Hz and an accuracy of 0.5° (Tobii TX300; Tobii Technology, Stockholm, Sweden). A 23'' monitor (screen resolution: 1920*1080 pixels) mounted on top of the eye-tracker was used as

[1] For video examples of all states see Supplementary of Jording et al. 2019: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01913/full#supplementary-material

the display (Figure 1a). Participants were seated at a distance ranging between 50 – 70 cm from the monitor and gave their responses (during ObR) via a keyboard with the marked buttons “J” (for “yes”) and “N” (for “no”).

Data preprocessing and statistical analysis

From 2160 trials total in the ObR condition (45 participants with 48 trials each), 92 trials were excluded due to missing responses or response times exceeding 30 seconds, another 432 trials were excluded because of more than 20% of missing gaze data. After trial exclusion, 1636 valid trials remained for statistical analysis. Response, eye-tracking data and questionnaire data were pre-processed and statistically analyzed using the software R (version 3.6.2; R Development Core Team, 2008). Analysis followed the procedure described previously in Jording et al. (2019). Response and eye-tracking data were analyzed in likelihood ratio tests of (generalized) linear mixed effects models (“lme4” package; Bates et al., 2015), followed by Tukey post-hoc tests correcting for multiple comparisons (“multcomp” package; Hothorn et al., 2008). We analyzed the effect of the diagnostic groups (factor “*group*”), the conditions grouped into interactive agent states vs. non-interactive states (factor “*interactivity*”) and the participants' gender. In order to analyze gaze data we computed ‘relative fixation durations’ of fixations on the AoIs (*aois*) ‘eyes’, ‘face’ (excluding the eyes) or ‘objects’ (all four objects). As a measure for the participants' responsiveness to the agent, we identified all situations in which the participant and the agent both looked at each other's eyes (eye contact) or both looked at the same object (joint attention). For analytical purposes, we will refer to these situations as *‘shared focus’*, irrespective of the state in which they occurred.

Results

Interactivity ratings

We analyzed differences in participants' experiences of interactivity with respect to agent states between both diagnostic groups (Figure 2a). Likelihood ratio tests of logistic regression models with random intercepts for participants revealed significant effects of the agents *interactivity* ($X^2(1) = 211.41$, $p < .001$) and a significant interaction effect for *interactivity***group* ($X^2(1) = 5.17$, $p = .023$), but no significant main effect of *group* ($X^2(1) = 2.67$, $p = .103$; Supplementary Table S1.1). The predicted odds ratio for identifying an agents' interactive state as interactive was decreased for the ASD subjects compared to the controls by a factor of 0.61 (Cl = 0.39 - 0.93, Supplementary Table S1.2). In a Tukey post-hoc test participants with ASD and control participants differed significantly in recognizing interactive ($M = 0.57$, $SE = 0.22$, $z = -2.56$, $p = .048$) but not in recognizing non-interactive agent states (Supplementary Table S1.3). For the logarithmized response times (Figure 2b), likelihood ratio tests did not reveal any significant effects of *interactivity* ($X^2(1) = 1.84$, $p = .175$), *group* ($X^2(1) = 0.22$, $p = .640$), or *interactivity***group* ($X^2(1) = 1.48$, $p = .115$; Supplementary Table S2.1). The inclusion of *gender* of participants as a factor did not significantly improve model fits for the mean interactivity ratings (ASD: $F(1, 19) = 0.82$, $p = .377$; controls: $F(1, 22) = 2.41$, $p = .135$; Supplementary Table S2.1 & S2.2) or mean response times (ASD: $F(1, 19) = 0.37$, $p = .553$; controls: $F(1, 22) = 0.47$, $p = .501$; Supplementary Table S2.3 & S2.4) in either group.

Gaze behavior

In order to better understand the participants´ ratings, we analyzed their gaze behavior during ObR (Figure 3), specifically with regard to the distribution of the visual attention between the AoIs (Eyes, Face, Object). Model fits for relative durations (as the portion of the total time that was spent on the specific AoI, ranging from 0 to 1) were significantly improved by including the factor *aoi* ($X^2(2) = 2144.33$, $p < .001$). The factor *group* did not significantly improve the model fit, neither directly ($X^2(1) = 1.39$, $p < .239$), nor as part of the interaction *aoi***group* ($X^2(2) = 5.83$, $p = .054$; Supplementary Table S4.1 $ S4.2). Tukey post-hoc tests comparing between the different combinations of *aoi* and *group* (Supplementary Table S4.3) did not reveal any significant differences between groups for any of the *aoi*s.

Next, we analyzed the establishment of instances of shared focus (eye contact or joint attention). In a first step, we analyzed whether the frequency of the establishment of shared focus differed between groups and interactive vs. non-interactive agents (Figure 4a). The fit for the prediction of the number of shared focus instances was significantly improved by *interactivity* ($X^2(1) = 230.91$, $p < .001$), but not by *group* ($X^2(1) = 1.68$, $p = .195$) or the interaction *interactivity*group* ($X^2(1) = 2.68$, $p = .101$; Supplementary Table S5.1). Averaged over groups, the predicted incidence rate of shared foci was increased by a factor of 1.33 (CL = 1.26 – 1.41, Supplementary Table S5.2) when the agent was interactive compared to non-interactive.

In a second step, we investigated, to what extent the number of shared focus instances (Figure 4B) could predict the interactivity ratings depending on whether the agent was actually trying to interact or not. We started from the model with best fit for the prediction of the participants' response, including the predictors *interactivity*, *group* and the interaction *interactivity*group* (see section 'Interactivity ratings'). We tested, whether additionally including the number of *shared focus* instances would, by themselves or via interaction effects, improve the model fit. The prediction of the participants response was significantly improved when including *shared focus* ($X^2(1) = 118.72$, $p < .001$), the interaction *interactivity*shared focus* ($X^2(1) = 39.47$, $p < .001$) and the interaction *group*shared focus* ($X^2(1) = 12.95$, $p < .001$). The interaction *interactivity*group*shared focus* instances did not significantly improve the fit ($X^2(1) = 0.69$, $p = .407$; Supplementary Table S6.1). The occurrence of shared foci on average increases the likelihood of interactive ratings. However, this effect is mediated by interactions and is stronger for interactive compared to non-interactive agents and for control compared to ASD subjects, averaged over the respective other conditions (Supplementary Table S6.2).

Discussion

The present study focused on the capacity of persons with ASD to infer mental states during gaze-based encounters. In accordance with our first hypothesis, it shows that persons with ASD have indeed substantially more difficulties in recognizing attempts to establish an interaction by another person in gaze communication as compared to control persons. Solving the task required several skills: the attention to social information, the detection of social cues, as well as adequately responding to these cues and evaluating the interaction as a whole. Thus, detailed observations of the participants' behavior allowed us to test also our second and third hypotheses, whether visual attention to socially relevant stimuli would be reduced in ASD; and whether responsiveness to the agent would be reduced in ASD. In the following we will discuss these hypotheses based on our results and will develop an educated guess as to the origin of impairments in inferring mental states in ASD.

Participants with ASD were significantly impaired in their ability to recognize the agent's intentions to interact compared to controls. This is in accordance with earlier studies reporting an impaired ability to infer mental states from the eye region of others (Baron-Cohen et al., 2001a, 1997). To the best of our knowledge, our study is the first demonstration that this impairment translates to the inference of mental states from gaze behavior in interactive gaze encounters. Potential implications for the everyday life for persons with autism are severe, especially given the importance of gaze communication in the establishment of social relationships and cooperation (Tomasello et al., 2007). For further investigations of this effect as well as its consideration in clinical contexts (diagnostics and therapy), it is, however, important to learn more about its origins. To this end, we additionally analyzed the distribution of participants' visual attention, and the occurrence of instances of shared focus, i.e. eye contact and joint attention.

The analysis of the participants' visual attention as the proportion of fixations on different AoIs did not reveal significant differences between the two groups. This is contrary to results about diminished attention to socially relevant stimuli and especially human eyes in ASD (Dalton et al., 2005; Klin et al., 2003, 2002b). Considering that in the past interactive or socially enriched settings were found to more clearly distinguish visual attention in ASD (Chita-Tegmark, 2016; Falck-Ytter & von Hofsten, 2011; Frazier et al., 2017; Guillon et al., 2014), we had expected more pronounced differences. However, our data show that the task explicitly draws the attention of participants to the eye region by instructing to guide or be guided via gaze.

Similarly, previous studies have demonstrated such explicit attentional mechanisms attenuating behavioral differences for facial processing (Schulte-Rüther et al., 2013, 2017) and gaze processing (Oberwelland et al., 2017) in ASD. In summary, we have no reasons to consider different degrees of attention allocation being responsible for the mentalizing impairments in ASD found in this study. Furthermore, we did not find any significant differences in response times, suggesting that both groups of participants were similarly involved in solving the task.

We further investigated the participants' responsiveness by analyzing the occurrence of joint/shared foci. Both diagnostic groups did not differ significantly with regard to the frequency of shared focus events (eyes, objects). In both groups, more shared focus events occurred in interactive as compared to non-interactive trials, but this effect was not modified by group affiliation. Again, these results do not support the intuitive hypothesis of a diminished occurrence of eye contact and joint attention in ASD, e.g. due to a reduced attention for social signals. A detailed analysis of the literature reveals a less clear picture regarding gaze processing and joint attention in ASD than commonly expected. Gaze direction processing, i.e. the ability to estimate the gaze angle and line of sight of another person, has been found to be impaired in ASD (Ashwin et al., 2009; Howard et al., 2000). However, in one study, children with ASD were able to trace the line of sight of another person – be it by less conventional strategies (Leekam et al., 1997). Furthermore, no differences were found in the ability to detect changes in gaze direction between adolescent and adult ASD and control participants (Fletcher-Watson et al., 2008). Previous results on gaze cueing in ASD, i.e. the effect that onlookers automatically shift their (covert) attention in accordance with the gaze direction of observed eyes, were heterogeneous as well (Frischen et al., 2007; Vlamings et al., 2005). Regarding the ability to follow the gaze of another person and to establish joint attention, again, it seems that no general impairments are identifiable in ASD (Chawarska et al., 2003; Kylliainen & Hietanen, 2004). However, children with ASD seem to be less eager to initiate joint attention (MacDonald et al., 2006; Mundy, 2003; Mundy et al., 1994; Whalen & Schreibman, 2003). Some authors have argued that while persons with ASD in principle may be able to "mechanically" follow the gaze of another person, the voluntary application of the skill in "mentalistic" contexts (i.e. to understand their mental states) is reduced (Driver et al., 1999; Vlamings et al., 2005).

The exploratory analysis of the relationship between shared focus instances and interactivity ratings yielded some interesting insights into the differential evaluation of gaze by ASD participants as compared to control participants. It seems that the probability of an 'interactive' rating generally increases with the number of shared focus events, although this effect is mediated by interactions with the agents' interactivity and the diagnostic group. This would suggest that both eye contact and joint attention were interpreted as a signal for an interactive situation. Furthermore, it would corroborate earlier findings demonstrating that shared attention, i.e. mutual awareness of the joint effort to coordinate attention, is established by alternating between eye contact and joint attention (Pfeiffer et al., 2012). Interestingly, the effect of eye contact and joint attention on the experience of interactivity was stronger when the agent was interactive, i.e. was reacting to the participant in a contingent fashion (statistical interaction: *eye contact / joint attention x Interactivity*). Thus, it seems that participants did notice the contingencies between them and the agent and took them as a hint for the agent's interactive intentions. This in turn prompts the question, whether participants were aware of that. In previous studies, healthy participants were able to detect and react to an agent following their gaze without becoming aware of the dependencies (Courgeon et al., 2014; Grynszpan et al., 2017). With regards to differences between diagnostic groups, we found that the occurrence of eye contact and joint attention predicted interactivity ratings more reliably for control participants than for ASD participants. This is in concordance with the generally reduced sensitivity to gaze cues reported for ASD (Dratsch et al., 2013; Freeth et al., 2010; Georgescu et al., 2013). However, other studies did not find impairments in the detection of social contingencies in ASD (Zapata-Fonseca et al., 2018) and future studies should focus on the question whether it is the detection or the evaluation of social contingencies that is responsible for the reduced impression of interactive intentions in ASD.

Taken together, our results suggest that the differences in the experience of interactivity are unlikely to be driven by attentional processes. Traditionally, impairments in social skills in ASD have often been attributed to a lack of motivation and subsequently attention to social stimuli (Chevallier et al., 2013, 2012). However, our data suggest that instead of basic attentional deficits, the higher-order evaluation of other people's gaze behavior was altered in ASD. It is not primarily a deficit in visual attention for communicative cues of their partner, or in the capacity to respond appropriately to them being impaired in ASD. Instead, it seems that the subsequent processing and/or interpretation of the arising social contingencies differ in persons with ASD. This conclusion relates well to approaches in social psychology bringing forth deficits in the evaluation of social stimuli, e.g. in the domain of person perception (Georgescu et al. 2013) or animacy experience (Kuzmanovic et al. 2014), but not in the mere detection. It is also in line with fMRI studies in which control participants and participants with ASD did not necessarily react differently to observing joint attention but still showed different activation patterns in areas related to social-cognitive processing (Greene et al., 2011; Redcay et al., 2013).

Similarly, altered activation patterns in areas of the "social brain" despite comparable behavioral performance were observed in ASD adolescents, suggesting less elaborated processing of gaze cues in social contexts (Oberwelland et al, 2017).

So far, studies about social gaze in ASD in large parts focused on isolated aspects of gaze behavior and examined these under highly controlled experimental conditions. These studies have given us a very detailed picture of some of the elements and building blocks of gaze communication in ASD and constitute a foundation for further advances in the field. However, the knowledge acquired from these reductionist approaches is fragmentary and very specific to the context and it is not clear how individual results relate to each other. A good example is the case of visual attention for social stimuli in ASD where results seem to be very contradictory and their comparability is unclear. Consequently, we can only speculate about the potential impact on more complex cognitive processes, e.g. the role of attentional deficits in ASD for the ascription of mental states. In this study, we followed a new, holistic approach in which we observed the unfolding encounter while participants engaged in gaze-based interaction. This allowed us to systematically differentiate behaviors related to different parts of the task and compare them between groups. However, it has to be noted that the design does not allow us to test a causal relationship or rule out additional differences between groups that might affect the performance in mentalizing.

Limitations

It is important to take into account some limitations of the design (see also Jording et al., 2019). First, we deliberately focused on gaze communication and restricted the communication to this channel. An unrestricted environment may allow for faster and more accurate evaluations of the interaction. Second, the agent did not fluently transition between attentional states as would be expected in a natural encounter (Jording et al., 2018), instead, the encounter was divided into separate trials. With regards to the differences between control and ASD subjects, it should be noted that both groups differed in gender distribution. We did not systematically manipulate agents' gender. However, we did not find any significant effects for gender on mean interactivity ratings or mean response times for either of the two groups. We also did not systematically control for the IQ of participants and while we can rule out cases of mental retardation in our sample, the possibility of an effect of IQ remains.

Conclusion

ASD participants did not show any perceptual differences in visual attention or mere "detection" during gaze encounters as compared to controls, nor did the emerging interactions differ in the establishment of eye contact or joint attention. Nonetheless, ASD participants evaluated the perceived cues differently when compared to control participants and experienced less interactivity. This potentially has highly relevant implications for the (neuroscientific) investigation of interaction disturbances in ASD as well as for the development of diagnostic and therapeutic instruments. Instead of a simple passive observation and quantification of patients' behaviors, a holistic and socially contextualized consideration of patients' inner experience during interaction is necessary. The new human-agent interaction platform TriPy, with its implementation of the SGS as a holistic taxonomy of triadic gaze interactions, has proven to be a reliable tool for this kind of investigation. Furthermore, it constitutes a promising basis for a future diagnostic or therapeutic instrument in clinical contexts.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: Author.

Ashwin, C., Ricciardelli, P., & Baron-Cohen, S. (2009). Positive and negative gaze perception in autism spectrum conditions. *Social Neuroscience*, *4*(2), 153–164. https://doi.org/10.1080/17470910802337902

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001a). The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001b). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Burgoon, J. K., Buller, D. B., & Woodall, W. G. (1989). *Nonverbal communication: The unspoken dialogue*. New York: Harper & Row.

Chawarska, K., Klin, A., & Volkmar, F. (2003). Automatic Attention Cueing Through Eye Movement in 2-Year-Old Children With Autism. *Child Development*, *74*(4), 1108–1122. https://doi.org/10.1111/1467-8624.00595

Chevallier, C., Huguet, P., Happé, F., George, N., & Conty, L. (2013). Salient Social Cues are Prioritized in Autism Spectrum Disorders Despite Overall Decrease in Social Attention. *Journal of Autism and Developmental Disorders*, *43*(7), 1642–1651. https://doi.org/10.1007/s10803-012-1710-x

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

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, *48*, 79–93. https://doi.org/10/f76qfz

Courgeon, M., Rautureau, G., Martin, J.-C., & Grynszpan, O. (2014). Joint Attention Simulation Using Eye-Tracking and Virtual Humans. *IEEE Transactions on Affective Computing*, *5*(3), 238–250. https://doi.org/10/ggv7m6

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., … Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*. https://doi.org/10.1038/nn1421

Dratsch, T., Schwartz, C., Yanev, K., Schilbach, L., Vogeley, K., & Bente, G. (2013). Getting a Grip on Social Gaze: Control over Others' Gaze Helps Gaze Detection in High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *43*(2), 286–300. https://doi.org/10.1007/s10803-012-1569-x

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

Falck-Ytter, T., & von Hofsten, C. (2011). How special is social looking in ASD: a review. *Progress in Brain Research*, *189*, 209–222. https://doi.org/10.1016/B978-0-444-53884-0.00026-9

Fletcher-Watson, S., Leekam, S. R., Findlay, J. M., & Stanton, E. C. (2008). Brief Report: Young Adults with Autism Spectrum Disorder Show Normal Attention to Eye-Gaze Information—Evidence from a New Change Blindness Paradigm. *Journal of Autism and Developmental Disorders*, *38*(9), 1785–1790. https://doi.org/10.1007/s10803-008-0548-8

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A Meta-Analysis of Gaze Differences to Social and Nonsocial Information Between Individuals With and Without Autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(7), 546–555. https://doi.org/10/gbncjs

Freeth, M., & Bugembe, P. (2019). Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults? *Autism*, *23*(2), 503–513. https://doi.org/10/ggwj5p

Freeth, M., Chapman, P., Ropar, D., & Mitchell, P. (2010). Do Gaze Cues in Complex Scenes Capture and Direct the Attention of High Functioning Adolescents with ASD?

Evidence from Eye-tracking. *Journal of Autism and Developmental Disorders*, *40*(5), 534–547. https://doi.org/10.1007/s10803-009-0893-2

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, U. (1996). Social communication and its disorder in autism and Asperger syndrome. *Journal of Psychopharmacology*, *10*(1), 48–53. https://doi.org/10/bvgdmd

Georgescu, A. L., Kuzmanovic, B., Schilbach, L., Tepest, R., Kulbida, R., Bente, G., & Vogeley, K. (2013). Neural correlates of “social gaze” processing in high-functioning autism under systematic variation of gaze duration. *NeuroImage: Clinical*, *3*, 340–351. https://doi.org/10.1016/j.nicl.2013.08.014

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

Grynszpan, O., Nadel, J., Martin, J.-C., & Fossati, P. (2017). The awareness of joint attention: A study using gaze contingent avatars. *Interaction Studies*, *18*(2), 234–253. https://doi.org/10/gcr292

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

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Howard, M. A., Cowell, P. E., Boucher, J., Broks, P., Mayes, A., Farrant, A., & Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism: *NeuroReport*, *11*(13), 2931–2935. https://doi.org/10/fm4qwn

Hutchins, T. L., & Brien, A. (2016). Conversational topic moderates social attention in autism spectrum disorder: Talking about emotions is like driving in a snowstorm. *Research in Autism Spectrum Disorders*, *26*, 99–110. https://doi.org/10.1016/j.rasd.2016.03.006

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2018). The “Social Gaze Space”: A Taxonomy for Gaze-Based Communication in Triadic Interactions. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00226

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2019). Inferring Interactivity From Gaze Patterns During Triadic Person-Object-Agent Interactions. *Frontiers in Psychology*, *10*, 1913. https://doi.org/10/ggdkgw

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 345–360. https://doi.org/10.1098/rstb.2002.1202

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002a). Defining and Quantifying the Social Phenotype in Autism. *American Journal of Psychiatry*, *159*(6), 895–908. https://doi.org/10.1176/appi.ajp.159.6.895

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002b). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809

Krämer, N. C. (2008). Nonverbal communication. In J. Blascovich & C. R. Hartel (Eds.), *Human Behavior in Military Contexts* (pp. 150–188). Washington, DC: The National Academies Press.

Kylliainen, A., & Hietanen, J. K. (2004). Attention orienting by another's gaze direction in children with autism. *Journal of Child Psychology and Psychiatry*, *45*(3), 435–444. https://doi.org/10.1111/j.1469-7610.2004.00235.x

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, *15*(1), 77–95. https://doi.org/10/bmprjn

MacDonald, R., Anderson, J., Dube, W. V., Geckeler, A., Green, G., Holcomb, W., … Sanchez, J. (2006). Behavioral assessment of joint attention: A methodological report. *Research in Developmental Disabilities*, *27*(2), 138–150. https://doi.org/10.1016/j.ridd.2004.09.006

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

Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry*, *44*(6), 793–809. https://doi.org/10.1111/1469-7610.00165

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology*, *6*(3), 389–401. https://doi.org/10.1017/S0954579400006003

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., … Schulte-Rüther, M. (2017). Young adolescents with autism show abnormal joint attention network: A gaze contingent fMRI study. *NeuroImage: Clinical*, *14*, 112–121. https://doi.org/10.1016/j.nicl.2017.01.006

Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., & Vogeley, K. (2012). Eyes on the Mind: Investigating the Influence of Gaze Dynamics on the Perception of Others in Real-Time Social Interaction. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00537

Pfeiffer, U. J., Timmermans, B., Vogeley, K., Frith, C. D., & Schilbach, L. (2013). Towards a neuroscience of social interaction. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00022

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, M., Pearrow, M. J., Triantafyllou, C., … Saxe, R. (2013). Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder: Neural Correlates of Joint Attention in ASD. *Human Brain Mapping*, *34*(10), 2511–2523. https://doi.org/10/gc6zcr

Riby, D. M., & Hancock, P. J. B. (2008). Viewing it differently: Social scene perception in Williams syndrome and Autism. *Neuropsychologia*, *46*(11), 2855–2860. https://doi.org/10.1016/j.neuropsychologia.2008.05.003

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

Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, *11*(6), 449–449. https://doi.org/10.1038/nrn2805-c1

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*(04), 393–414. https://doi.org/10.1017/S0140525X12000660

Schulte-Rüther, M., Greimel, E., Piefke, M., Kamp-Becker, I., Remschmidt, H., Fink, G. R., … Konrad, K. (2013). Age-dependent changes in the neural substrates of empathy in autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *9*(8), 1118–1126. https://doi.org/10.1093/scan/nst088

Schulte-Rüther, M., Otte, E., Adigüzel, K., Firk, C., Herpertz-Dahlmann, B., Koch, I., & Konrad, K. (2017). Intact mirror mechanisms for automatic facial emotions in children and adolescents with autism spectrum disorder: Intact mirror mechanisms in Austism. *Autism Research*, *10*(2), 298–310. https://doi.org/10/f9vkwd

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

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

Vlamings, P. H. J. M., Stauder, J. E. A., van Son, I. A. M., & Mottron, and L. (2005). Atypical Visual Orienting to Gaze- and Arrow-Cues in Adults with High Functioning Autism. *Journal of Autism and Developmental Disorders*, *35*(3), 267–277. https://doi.org/10/dt73n9

Whalen, C., & Schreibman, L. (2003). *Joint attention training for children with autism using behavior modi¢cation procedures*. 13.

Wilms, M., Schilbach, L., Pfeiffer, U. J., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes—Using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(1), 98–107. https://doi.org/10.1093/scan/nsq024

World Health Organization (Ed.). (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva: World Health Organization.

Zapata-Fonseca, L., Froese, T., Schilbach, L., Vogeley, K., & Timmermans, B. (2018). Sensitivity to Social Contingency in Adults with High-Functioning Autism during Computer-Mediated Embodied Interaction. *Behavioral Sciences*, *8*(2), 22. https://doi.org/10.3390/bs8020022

Figure 1. Illustration of the technical setup and the participants' perspective during the experiment (adapted from Jording et al., 2019). A: Illustration of a participant interacting with the agent controlled by the platform TriPy. B: The behavior of the agent created by TriPy as seen from the perspective of the participant (B).

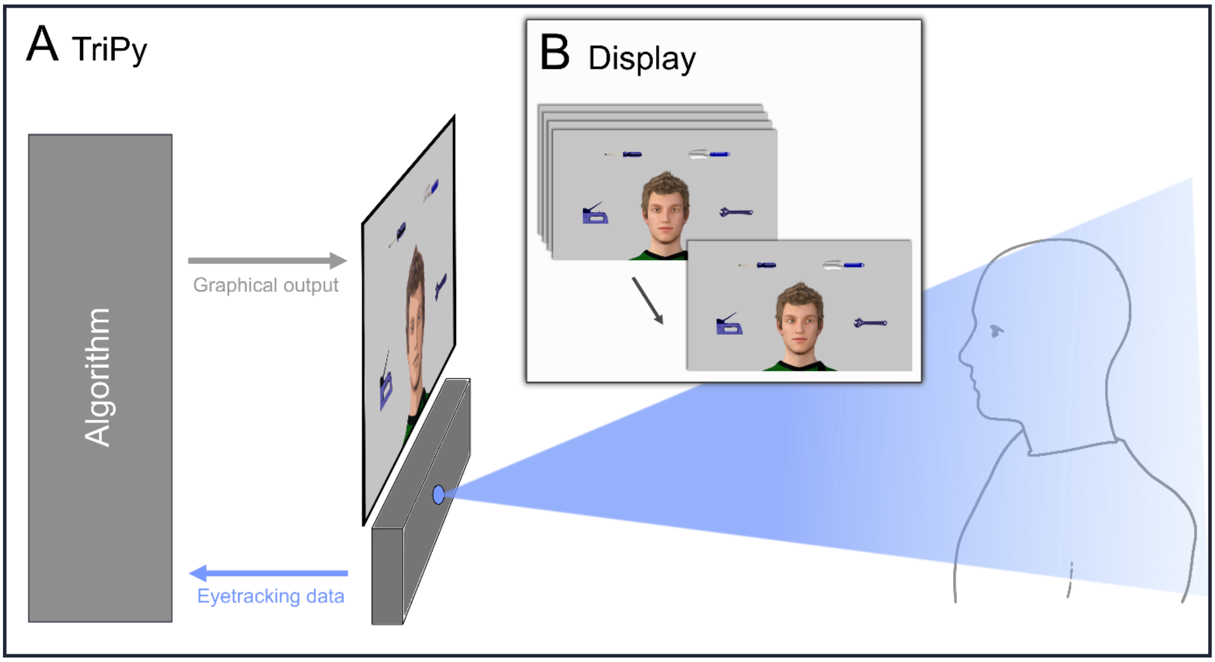


Figure 2. Plots of mean interactivity ratings (A) and mean response times (in ms, B) for diagnostic groups (blue: control persons / orange: ASD participants) and non-interactive (left) vs. interactive (right) agent states. A: Asterisks indicate significant post-hoc comparisons (* <.05; ** <.01, * <.001), dashed line indicates the 50% guessing rate). ASD participants have significantly smaller detection rates for interactive states compared to controls participants. Diagnostic groups do not differ with regard to non-interactive states or response times.**

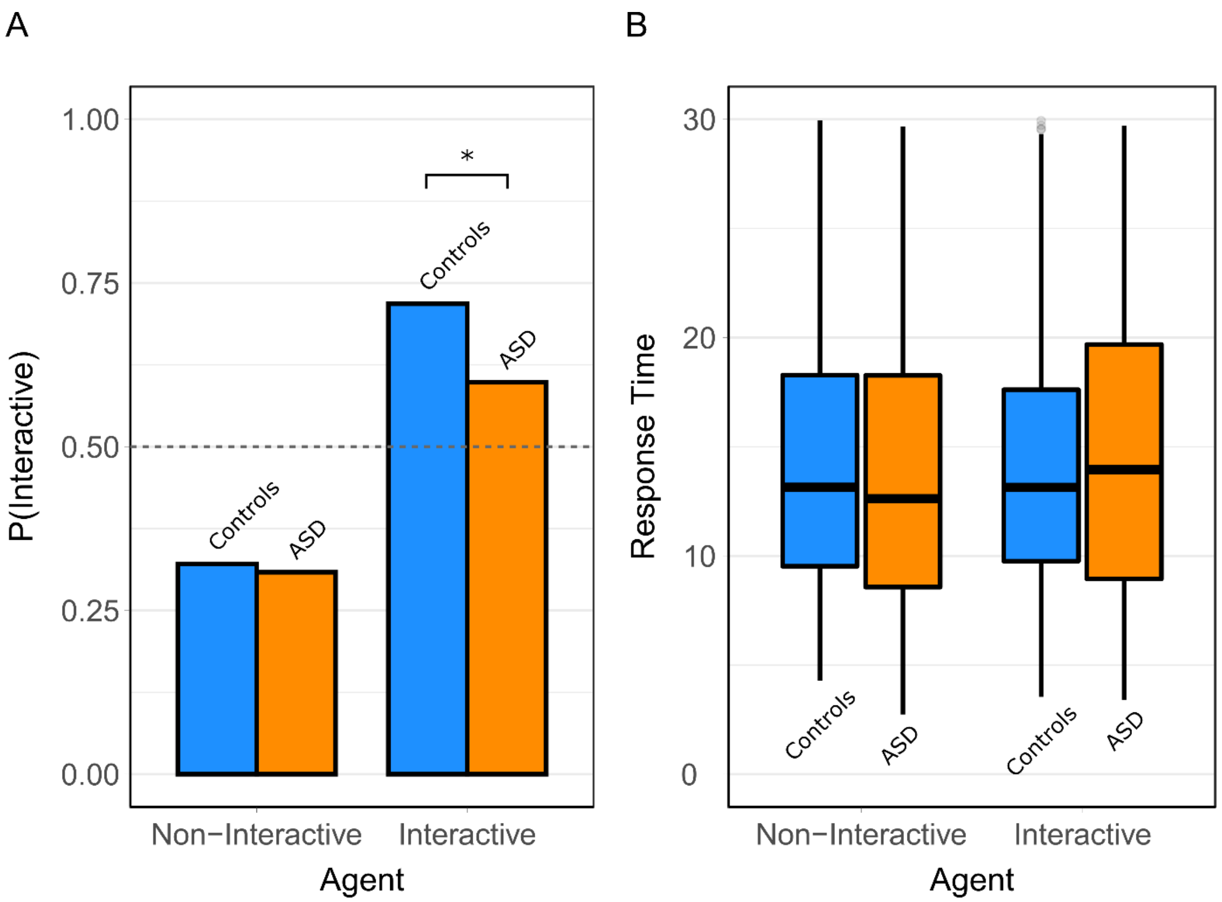


Figure 3. Boxplot of the distribution of the participant´s visual attention measured as relative fixation durations, i.e. the portion of time spent on the different AoIs (eyes, face, objects) per trial for both diagnostic groups (blue: control participants / orange: ASD participants). Diagnostic groups do not differ significantly in their distribution of visual attention.

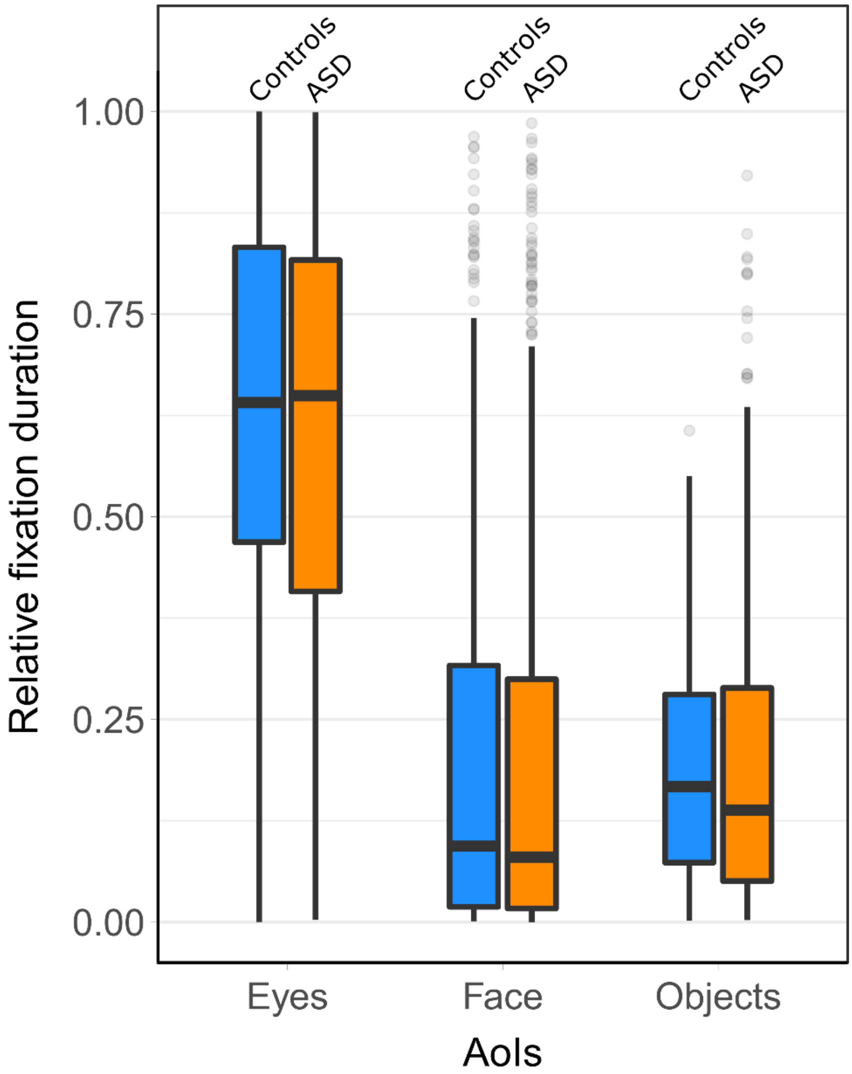
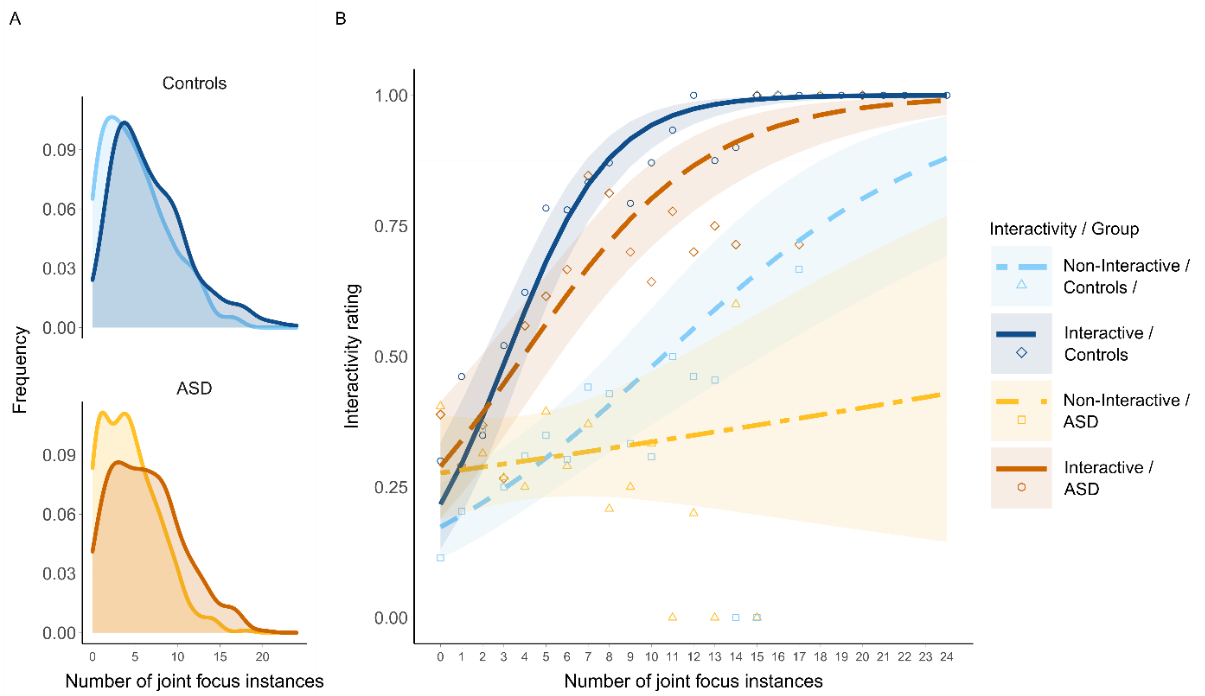


Figure 4. Illustration of the distribution of instances of shared foci (eye contact or joint attention) between participant and agent, separately for a non-interactive agent (light colors) vs. an interactive agent (dark colors) in both diagnostic groups (blue: control participants / orange: ASD participants). A: Shared foci frequencies per trial for control participants (top) and ASD participants (bottom); B: Mean rates (triangles, diamonds, squares, and circles) and model predictions (lines) with 95% confidence intervals (ribbons) of interactivity ratings for differing numbers of shared foci instances per trial. The establishment of shared foci on average predicts the participants' interactivity ratings. This effect is especially strong for interactive compared to non-interactive agents and for control participants compared to ASD participants.



Towards Computer Aided Diagnosis of Autism Spectrum Disorder Using Virtual Environments

1st Daniel Roth
Technical University of Munich
Munich, Germany
daniel.roth@tum.de

2nd Mathis Jording
Forschungszentrum Jülich
Jülich, Germany
m.jording@fz-juelich.de

3rd Tobias Schmee
University of Würzburg
Würzburg, Germany
tobias.schmee@stud-mail.uni-wuerzburg.de

4th Peter Kullmann
University of Würzburg
Würzburg, Germany
peter.kullmann@uni-wuerzburg.de

5th Nassir Navab
Technical University of Munich
Munich, Germany
nassir.navab@tum.de

6th Kai Vogeley
Forschungszentrum Jülich,
University Hospital Cologne
Cologne, Germany
kai.vogeley@uk-koeln.de

***Abstract*—Autism Spectrum Disorders (ASD) are neurodevelopmental disorders that are associated with characteristic difficulties to express and interpret nonverbal behavior, such as social gaze behavior. The state of the art in diagnosis is the clinical interview that is time intensive for the clinicians and does not take into account any objective measures of behavior. We herewith propose an empirical approach that can potentially support diagnosis based on the assessment of nonverbal behavior in avatar-mediated interactions in virtual environments. In a first study, ASD individuals and a typically developed control group were interacting in dyads. Head motion, and eye gaze of both interlocutors were recorded, replicated to the avatars and displayed to the partner through a distributed virtual environment. The nonverbal behavior of both interaction partners was recorded, and resulting preprocessed data was classified with up to 92.9% classification accuracy, with the amount of eye area focus and the average horizontal gaze change being the most relevant features. We expect that such systems could improve the diagnostic assessment on the basis of objective measures of nonverbal behavior.**

***Index Terms*—Computer Aided Diagnosis, Autism, Avatars, Virtual Environments, Nonverbal Behavior.**

I. Introduction

A. Nonverbal communication

Humans in social interactions communicate up to 95% of information through nonverbal channels [1]. Nonverbal communication serves in learning from and adjusting to others, as substitute for spoken language, in the coordination of the interaction itself and as way to communicate emotional states and attitudes (cf. [2]). Importantly, a majority of the nonverbal signals are produced and perceived automatically and implicitly [3] and their exchange is a complex and dynamic process [4]. This has to be taken into account when trying to investigate nonverbal communication in an ecologically valid fashion [5]. Experimental designs based on virtual characters (utilized for avatars, agents, or even hybrid concepts) and virtual environments have been proposed [6] and developed [7]–[9] more recently since they can combine both highly controlled presentation of behavior of the interaction partners and ecological validity [2]. These designs thus have been explicitly suggested as a valuable methodological tool to assess and train nonverbal communication in ASD patients [10].

This research was supported by the EC, Horizon 2020 Framework Program, FET Proactive (Grant Agreement Id: 824128 to KV), and the XR Hub Bavaria funding line (Bavarian State Ministry for Digitalization).

B. Nonverbal behavior in ASD

ASD are neurodevelopmental conditions that are characterized "by persistent deficits in the ability to initiate and to sustain reciprocal social interaction and social communication, and by a range of restricted, repetitive, and inflexible patterns of behaviour and interests." [11]. Especially affected are nonverbal skills with well documented impairments in the recognition (e.g., [12]) and interpretation of nonverbal cues (e.g., [13]). Teenagers and adolescents with ASD focus look less on socially relevant features like faces [14] and have been reported to avoid eye contact and instead focus more often on the mouth area [15], [16], resulting in deviating scanpaths compared to typically developed persons [17].

C. Prevalence and diagnosis of ASD

While the prevalence rates of ASD are constantly rising over the last 20 years and are currently estimated up to 1 in 59 [18], the average time to diagnosis of ASD has been estimated to 13 months [19]. Consequently, many autistic persons remain unrecognized for years or dare not recognized until adulthood. In children this is especially critical, as early interventions can lead to greater gains in later years [20]. In adults, recognizing ASD is vital in order to adapt interventions and offer the right support. Adults with ASD are much more likely to suffer from depression and anxiety disorders compared to healthy persons [21]. Timely diagnosis of ASD can elevate the clinicans' alertness for potential comorbidities in patients and facilitate therapeutic interventions. In addition, persons with ASD are also much more often affected by unemployment than healthy persons [22]. It seems that these rates can be reduced drastically when cases are identified correctly and receive support [23].

Fig. 1. **Apparatus**. a) Exemplary setup. Two remote users interact with each other via avatars. b) The avatars. Humanoid but abstracted avatars were chosen to avoid any social information bias and avatars were matched to the participants gender.

Increasing the capacities for diagnosing ASD seems to be a crucial first step in order to increase the availability of clinical and social services for affected persons. However, the diagnostic of ASD is very time consuming since it has to be based on clinical interviews according to international standards [11]. Therefore, previous work suggested alternative tests to support the screening process of patients and potentially free up diagnostic capacities [24].

Yet, the most commonly used instruments for screening or support of the diagnostic process like the questionnaire AQ [25] or semi-structured assessment ADOS2 [26] have been found to be quite unreliable [27]. Thus, additional objective and more reliable instruments may benefit the diagnostic process.

The motivation for the present work was to investigate whether the use of virtual environments and technologies in truly interactive virtual paradigms, i.e., with a replication of nonverbal behavior during ongoing social interaction in real time, may be used to assist the diagnosis of ASD.

II. Related Work

A. Computer-Aided Investigation and Assessment

Previous work showed good results for classification approaches in detecting ASD on the basis of static images (e.g., [28], [29]). Motor behavior in terms of a grasping task [30] lead to a mean classification accuracy of 96.7 % [30]. Further, interpersonal synchronicity [31] has been investigated and classified with real-world motion data using motion energy analysis with a classification accuracy of 75.9 %. Drimalla and colleagues [32] showed that a classification of facial behavior recorded from a video-based simulated dialogue study led to 73% accuracy in detecting ASD. However, their approach was based on a simulated interaction with a pre-recorded video, which does not account for the dynamics of social interactions and may not capture the deficits in ASD to their full extent.

It seems that many of the ASD specific characteristics of nonverbal behavior mentioned earlier are expressed especially in socially complex situations [33] or become only visible when confronted with dynamic stimuli [14]. Additionally, it seems that instead of a general attenuation of nonverbal behavior it is the temporal coordination between nonverbal channels that is impaired in ASD [34].

In this paper, we demonstrate a fully interactive dyadic system, that combines different communication channels and allows for extended and less restricted encounters. It does not rely on artificial tasks requiring explicit instructions or substantially constrain the interaction in any other way but is based on everyday conversations. Our objectives were to 1) develop an apparatus to record and transmit nonverbal behavior with high ecological validity during ongoing virtual social interactions, 2) validate the approach by confirming earlier findings regarding differences in the nonverbal behavior of persons with and without ASD, and 3) investigate a potential classification of ASD through the acquired nonverbal behavior data. Our work contributes by providing a first approach to assess and measure nonverbal behavior in virtual paradigms in a controlled fashion, and by suggesting that computer-aided diagnosis may be highly valuable for the assessment of ASD.

B. Hypothesis

Considering that a vast amount of literature found differences in the behavior of ASD patients, we hypothesized that *H1: the nonverbal behavior is significantly different between autistic persons and typically developed persons*. Specifically, we hypothesized that *H2: gaze patterns of ASD patients differ significantly from those of typically developed persons*. Our main research question was *R1: Can we classify ASD on the basis of the expressed nonverbal behavior (gaze, voice, head motion) acquired through an avatar-mediated communication system?*.

III. Method

We constructed a one factor between-subjects experiment in which we compared virtual social interactions of diagnosed ASD participants to virtual social interactions of a control group, recording the resulting nonverbal behavior.

A. Apparatus

The distributed virtual environment was constructed similar to previous approaches of desktop-based avatar mediated systems [35]–[37]. Two participant clients using Tobii 4C eye trackers and Asus Swift PG279Q (27”) displays were networked via a server for experimental control and data logging. Each participant was equipped with a headset and

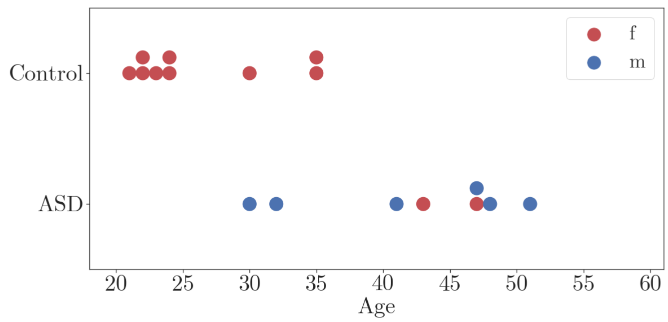


Fig. 2. Age and gender distribution of the 17 leftover subjects within the filtered dataset. Blue dots refer to male subjects; red dots refer to female subjects. The control group solely consisted of younger female participants, whereas subjects of the ASD group were mostly older males. Only two subjects of the ASD group were female. An anonymized, aggregated data set and the code will be made available with the publication.

calibrated for eye- and headtracking. Using a Unity3D simulation, we replicated gaze and eye tracking data to avatars (see Fig. 1b), that were displayed to the communication partner (see Fig. 1a). Voice was transmitted and mouth movements were animated via voice-to animation. Both, physical and virtual user positions and orientations were calibrated to match physical and virtual spaces. Thus, in case participants leaned forward, the avatars were displayed 'larger' to the communication partner since they moved 'closer' in 3D space. While participants had a certain movement ratio, the chair positions were fixed in order to ensure the best possible tracking quality and robustness. Participants and experimenter were located in three different rooms, each application was networked via a dedicated LAN (1Gbit). An end-to-end latency measure resulted in $M = 226$ ms ($SD = 35.79$) for the motion-to-photon video latency, and $M = 410.26$ ms ($SD = 19.69$) for the audio latency.

B. Participants

For the ASD group, 18 Participants (4 identifying as female, 14 as male; age $M = 43.44$, $SD = 9.55$) with the diagnosis Asperger syndrome (F84.5) were recruited via a outpatient clinic (University Hospital of Cologne). This group was compared to a group of 20 subjects without any record of psychiatric or neurological illnesses (18 identifying as female, 2 as male; age $M = 25.95$, $SD = 3.55$). The distribution of participants is depicted in Fig. 2.

C. Procedure

After giving their written informed consent, participants were introduced to each other, before they were then seated in remote rooms and calibrated for the experiment. Participants were instructed to discuss three different topics in 10-minute sequences each: 1. Introduction and getting to know each other; 2. Agree upon five items to take to a desert island; 3. Plan a five-course menu solely from ingredients they both dislike. The study was approved by the ethical committee of University Hospital of Cologne.

D. Data Acquisition

During the experiment, the behavior data was logged with 60 Hz by the server application resulting in 36000 data points per variable per task. The behavior data included the position (x,y,z) of the head of each participant as well as its orientation (x,y,z), the current gaze focus point on the screen, the gaze data validity, and whether or not the dynamic spatial areas of interest (AOIs, see Fig. 4 a) were in focus or the background was in focus.

E. Data Processing

Collected eye tracking data is usually affected by noise and invalid data [38], and was therefore preprocessed. We favored data quality over mass. Four dyads contained continuously corrupted data, and were excluded. Two dyads were removed because participants did not have the same gender. Fig. 3 illustrates our data processing pipeline. First, every conversation was split into splits of 2.5 minutes. Each split was treated separately. Second, problematic eye gaze data was removed. Sequences >100ms in duration that were tagged invalid by the system were excluded, sequences <100ms were interpolated. Third, invalid (i.e., without minimal change in head movement) head tracking sequences >300ms in duration were excluded, sequences <300ms were interpolated. Excluded sequences were excluded from both participating participants' data. To ensure a relative validity of the resulting data, a split is only kept if more than 3,000 frames (50 seconds) of tracking data are left. 54 splits (17 ASD, 37 control) fulfilled this requirement after filtering and are kept, which makes up 23.6%. Every split contains the information of both participants, meaning that 108 examples (34 ASD, 74 control) were extracted from the 54 splits. To create a balanced dataset 34 examples have randomly been selected from the control group so that the final balanced dataset consisted of 68 examples (34 ASD, 34 control). It is to note that the majority of problematic cases resulted from the ASD group, where behaviors may have included looking down or away from the screen more often. By nature, this results in a greater data loss. However, through this processing, we aimed at removing invalid data in a conservative fashion to hinder artificial bias (i.e., from tracking errors), and thus processed both groups equally on the basis of error corrections suggested by the literature (e.g., [39]).

From the resulting data, 14 features have been extracted. These comprise the average dwell times on each AOI, the average movements of the head position and rotation in all three directions, as well as the average horizontal and vertical shift of the gaze point on the screen per frame, see Table I. In the case of segment splits (data exclusions), border data was excluded to avoid distortions.

F. Classification

Three different classifier (logistic regression, SVM, neural network) have been evaluated. All were regularized using L_2-regularization and parameters have been chosen with grid

TABLE I

EXTRACTED FEATURES FROM THE RESULTING DATA.

Inner head	Relative dwell time on the inner head AOI.
Small Padding	Relative dwell time on the small padded AOI.
Big Padding	Relative dwell time on the big padded AOI.
Eyes	Relative dwell time on the eyes AOI.
Mouth	Relative dwell time on the mouth AOI.
Background	Relative dwell time on the background AOI.
Screen x	Average horizontal change of the gaze focus.
Screen y	Average vertical change of the gaze focus.
Head pos. x	Average horizontal change of the head position.
Head pos. y	Average vertical change of the head position.
Head pos. z	Average change of the z-axis distance to the screen.
Head rot. x	Average rotation of the head around the horizontal axis.
Head rot. y	Average rotation of the head around the vertical axis.
Head rot. z	Average rotation of the head around the z (depth) axis.

Note. The relative dwell time describes the relative percentage of gaze focus on an AOI.

search. The implementation was done with Python 3.7 and scikit-learn as machine learning library.

IV. RESULTS

A. Gaze Focus Distribution

Individuals with ASD show atypical gaze patterns and use alternative scanpaths. These deviations can also be observed in the recorded data. Mann-Whitney U tests of all 14 features (Bonferroni-corrected for multiple comparisons) between the 34 ASD and 34 control examples of the final dataset revealed significant more fixations on the eyes for the control group ($M = 0.51$, $SD = 0.15$, $W = 124$, $p < .001$) compared to the ASD group ($M = 0.23$, $SD = 0.16$). Whereas ASD subjects showed significant more fixations on big padded head ($M = 0.04$, $SD = 0.02$, $W = 926$, $p < .001$) and the background ($M = 0.47$, $SD = 0.19$, $W = 926$, $p < .001$) compared to controls (big padded head: $M = 0.02$, $SD = 0.01$; background: $M = 0.24$, $SD = 0.18$). Furthermore, ASD subjects made significantly more horizontal gaze shifts ($M = 15.56$, $SD = 4.68$, $W = 960$, $p < .001$) then controls did ($M = 10.18$, $SD = 3.22$). Mann-Whitney U tests of the remaining 10 features did not reveal any significant differences between the two groups. Fig. 4b+c show the dwell times on every AOI for individuals with and without ASD. It also reveals that gaze at the eyes and background makes up the largest difference between the groups.

B. Classification

Three approaches (logistic regression, SVM, and MLP) were chosen and their performance was evaluated. For classifier training the extracted variables were standardized by subtracting the mean of the training samples μ from the value x and subsequently dividing it by the standard deviation θ of the training samples.

$$\frac{x - \mu}{\theta} \tag{1}$$

This ensures that the features used in training are on the same scale. The classification approaches were evaluated in a 5-fold cross-validation process. To prevent overfitting, examples

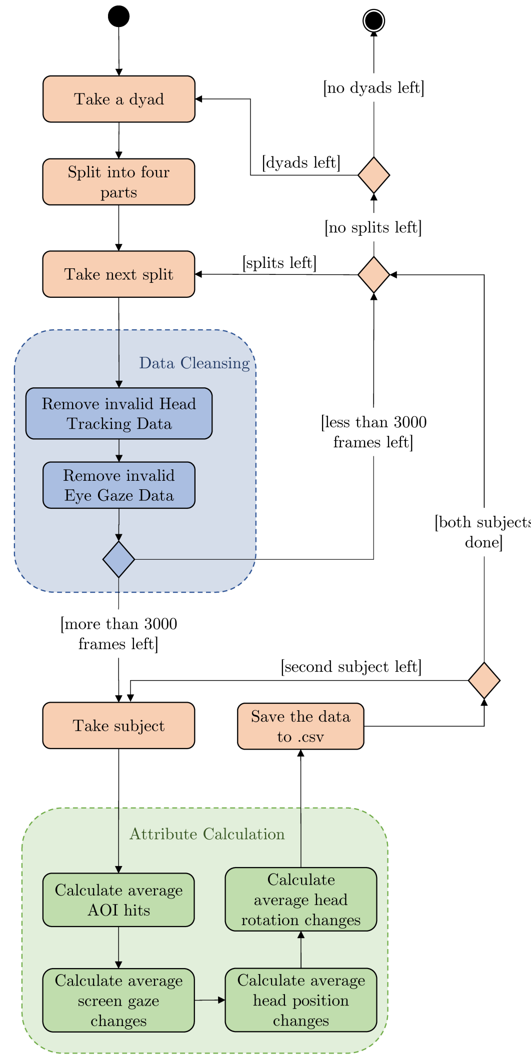


Fig. 3. **Overview of the dataset creation procedure, which subsumes the process of data filtering, as well as feature extraction.**

from the same subjects were grouped together and always in the same fold so that examples from one subject never appeared in training and test set at the same time. A 5-fold cross-validation reveals that a logistic regression model with $C = 0.2$ (max epochs = 5000) achieved an accuracy of 88.5% ($SD = 0.035$), a sensitivity of 88.8% ($SD = 0.089$) and a specificity of 87.0% ($SD = 0.125$). The classification accuracy of an SVM with radial basis function kernel reached an accuracy of 88.6% ($SD = 0.116$), a sensitivity of 99.3% ($SD = 0.116$) and a specificity of 89.0% ($SD = 0.054$). The regularization parameter $C = 2.9$ and the kernel parameter $\gamma = 0.02$ have been chosen with grid search. A neural network (Rectified Linear Unit as activation function, L-BFGS as solver, $\alpha = 1e^{-5}$, max epochs = 3000) with 2 hidden layers (2 and 4 neurons) achieved an accuracy of 92.9% ($SD = 0.160$), a sensitivity of 92.0% ($SD = 0.160$), and specificity of 98.0% ($SD = 0.040$).

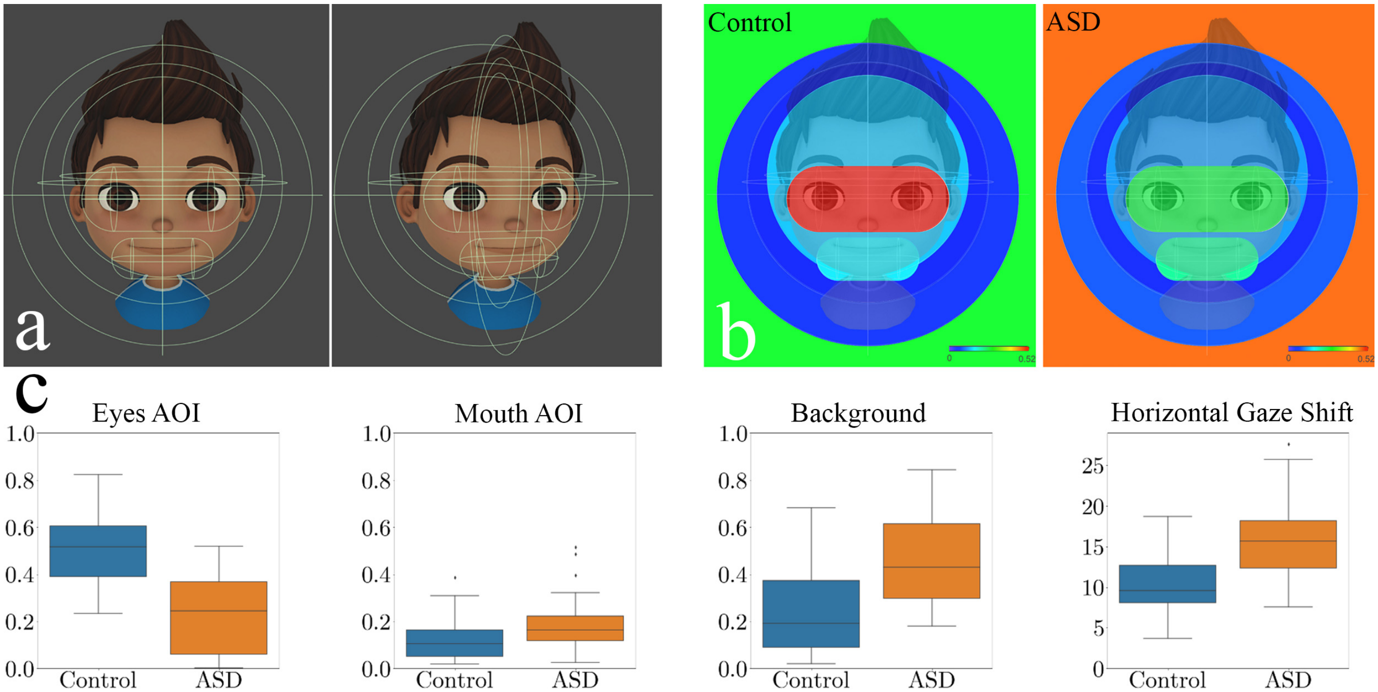


Fig. 4. **AOIs and focus distribution**. a) Dynamic spatial AOIs included an eyes AOI, a mouth AOI, as well as large, medium, and small padded head AOIs. b) The heat map illustrates that the control group had a much longer focus (i.e., longer dwell time) on the eyes AOI. c) From left to right: Whisker plots for the focus distributions of important AOIs according to the literature: Eyes AOI, Mouth AOI, and Background AOI. In addition, focus distribution of an important feature we found in our analysis: Horizontal Gaze Shift.

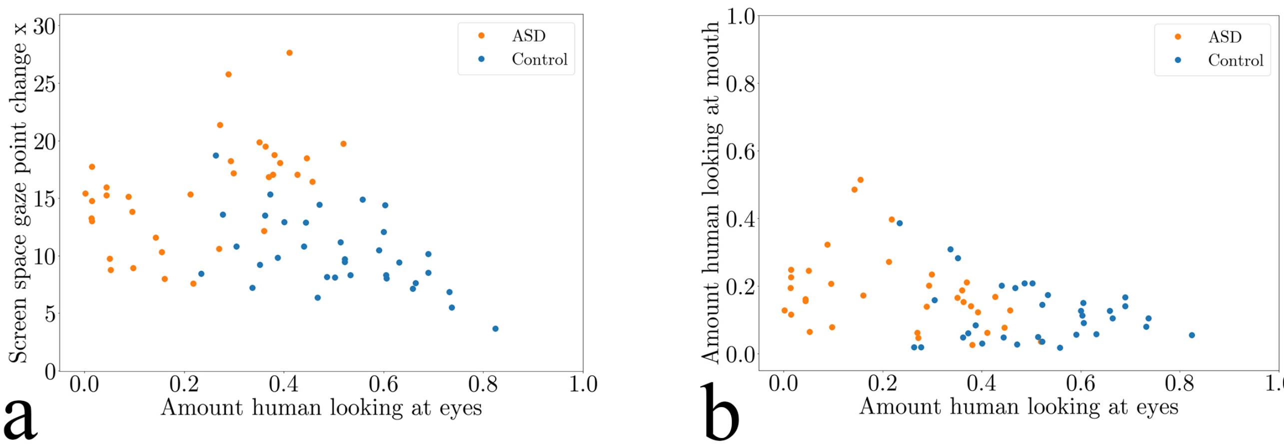


Fig. 5. **Separability**. a) Average change in horizontal focus vs. eyes AOI focus. b) Average focus on mouth area vs. eyes AOI focus.

C. Feature Analysis

In addition to training a classifier on all extracted features, separate classifiers have been trained on every single feature as well as every possible combination of two features. This allows an appraisal of the information content of each feature (see Tab. II). For the feature analysis, *we did not prevent* a mixture of an individual participant's data in training and test set, since we were interested in the individual contribution of each feature to a potential overall classification of ASD. A systematic analysis reveals the dwell time on the eyes to be most informative for the classification. A SVM trained on this single feature achieves an accuracy of 78.1% ($M = 0.781$, $SD = 0.038$). The second most informative feature is the average shift of the horizontal on-screen gaze point with an accuracy of 76.7% ($M = 0.767$, $SD = 0.046$). Evaluation of the feature combination demonstrates similar insights. In a 5-fold cross-validation, a SVM trained on these two features reaches an accuracy of 92.9% ($M = 0.929$, $SD = 0.078$). Hence, these features achieve the best results in a single analysis, and represent the best combination of features. Figure 5a+b visually illustrates the separability.

TABLE II

ACCURACY ($M \pm SD$) OF SVMS TRAINED ON SINGLE FEATURES.

Inner Head 0.574 ± 0.090	Small Padding 0.660 ± 0.067	Big Padding 0.707 ± 0.097
Eyes 0.781 ± 0.038	Mouth 0.660 ± 0.050	Background 0.733 ± 0.065
Screen x 0.767 ± 0.046	Screen y 0.643 ± 0.101	Head pox. x 0.550 ± 0.110
Head pos. y 0.457 ±0.073	Head pos. z 0.536 ± 0.111	Head rot. x 0.605 ± 0.089
Head rot. y 0.652 ± 0.194	Head rot. z 0.605 ± 0.155	

V. DISCUSSION

Our results confirm *H1: the nonverbal behavior is significantly different between autistic persons and typically developed persons*. While head motion was not a clear discriminator for our data, we confirmed *H2: gaze behavior was a significant discriminator*, which is in line with previous work. More specifically, the focus on the eyes AOI as well as the horizontal gaze movement were the most informative features. Answering *R1: Can we classify ASD on the basis of the expressed nonverbal behavior (gaze, voice, head motion) acquired through an avatar-mediated communication system?*, we found that a classification through the presented system reached a high accuracy of up to 92.9% using a neural network for our dataset. These results from the automatic classifications are promising with respect to the applicability as screening instrument in diagnostic procedures. While this approach cannot replace clinical interviews by educated clinicians, it may be applied as pre-screening of patients. Due to the simplicity of application and fully automated preprocessing it can potentially shorten waiting periods, save costs, and take strain from patients and health care workers. In addition, it can substantially enrich the diagnostic procedures by complementing the clinical interview with objective, quantitative data from the domain of nonverbal communication. Since the production and perception of nonverbal cues are normally unconscious processes, clinicians specialized on the diagnostics of autism have to learn to pay attention, describe and interpret the nonverbal behavior of patients. Additional quantitative measures can substantiate the clinicians descriptions and assessments. Furthermore, these measures might enhance comparability of cases and help standardize the diagnostic procedure. Due to different limitations (see below) direct comparisons to the state of the art screening and diagnostic tools are not yet possible. However, given the disappointing reliability of existing instruments [27] the results are promising. Another advantage of the system introduced here is that it does not involve specific test items, but is based on a mere conversation, irrespective of the conversation topic. Thus, patients do not have to solve additional test items, further lengthening the diagnostic process, but it can in principle be integrated in the standard clinical interview. From a scientific perspective we were able to replicate findings about generally reduced visual attention towards faces in persons with ASD [14]. In addition, the results corroborate the notion of diminished attention towards the eye region and seem to suggest increased fixations on the mouth region in ASD [15], [16]. Both of these effects are especially pronounced in interactive situations as recent reviews and meta-analysis have demonstrated [40]–[43]. Thus, the classification approach probably benefits from the interactive nature of the setting. The result of increased amounts of gaze shifts in the horizontal plane fits findings about enlarged horizontal gaze movement in children with ASD in naturalistic settings [44]. However, since we had no a priori hypothesis regarding direction specific differences in the amount of gaze shifts in ASD, this result remains to be confirmed in future replications. In general, the results substantiate the validity of the approach and its applicability to research in social psychology and social neuroscience.

A. Limitations

Some limitations arise. First, the study sample was limited and the exclusion of invalid data points during preprocessing further reduced the data available for classification. Despite the fact that it is challenging to recruit ASD patients for research studies, future studies should try to increase the data sample to open the gate for more sophisticated machine learning approaches that utilize larger data sets. Furthermore, more data had to be excluded for the ASD group then the control group. However, this was the result of ASD participants more often averting their gaze and turning their head away from their conversation partner, typical for dysfunctional communication behavior in ASD. Thus, one could argue that the excluded data would actually have been most informative for differentiating the two groups, potentially even enhancing the classification accuracy. Second, we only investigated ASD participants without any cognitive impairments or learning disabilities. Therefore, the applicability of the system will have to be tested separately for patients with additional impairments or comorbidities. Third, random sampling resulted in a considerable difference between groups with regard to the age and gender distribution. Forth, we did not have the chance to study mixed dyads in which a person with ASD interacts with a person without ASD.

B. Future Work

We hope to extend the present data set concerning sample size and variability, including mixed dyads. While we cannot publish raw data (institutional data protection policies), we have prepared and uploaded anonymized aggregates and the analysis code, see Section VIII. In addition to further data collection and analyses, we also consider including a greater fidelity and degree of replicated behaviors [7], which would allow for a greater feature space. Further, we propose to transfer the present application to a head-mounted display-based Virtual Reality application, by using head-mounted displays that support eye tracking in combination with user-embodiment [45], which also allows to integrate the tracking and replication of full body movements [8]. This may increase the precision of the approach since participants could move their head freely, while the tracking of gaze is still possible. A benefit of such a system would also be the greater movement space. Future work will also consider diagnosis approaches based on user-agent interactions (e.g., [46]), expecting that a single user interaction with an intelligent virtual agent will provide a higher degree of control, a better comparability of reactions, and will lead to a more efficient and automated screening procedure. Such a system may include pretested and validated animations for agent behavior, which may further increase the degree of experimental control [9], [47].

VI. Conclusion

We presented a truly interactive approach that allows to document and analyse the communication behavior of two interlocutors and classify ASD through nonverbal behavior analysis of avatar-mediated interactions with high accuracy. The presented approach is specifically beneficial as it can be fully automated and is ecologically valid. Our approach demonstrates that nonverbal behavior-based classification could be a potential tool to support clinicians in the diagnosis of ASD by providing objective data about the communication behavior of persons under diagnosis.

VII. Acknowledgements

We thank Kathrin Liese, Christina Adams, Leif Pursche, and Carolin Esser for their help in the data collection and Roman Eyck for his initial contribution to the software.

VIII. Data Availability

Classification data and code are available online: https://osf.io/u9pz3/ [48].

References

[1] D. Matsumoto, M. G. Frank, and H. S. Hwang, "Reading People," *Nonverbal Communication: Science and Applications: Science and Applications*, p. 1, 2012.

[2] K. Vogeley and G. Bente, ""Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication," *Neural Networks*, vol. 23, no. 8-9, pp. 1077–1090, Oct. 2010.

[3] Y. S. Choi, H. M. Gray, and N. Ambady, "The glimpsed world: Unintended communication and unintended perception," in *The New Unconscious*, R. R. Hassin, J. S. Uleman, and J. A. Bargh, Eds. New York: Oxford University Press, 2005, pp. 309–333.

[4] J. K. Burgoon, D. B. Buller, and W. G. Woodall, *Nonverbal communication: the unspoken dialogue*. New York: Harper & Row, 1989.

[5] L. Schilbach, B. Timmermans, V. Reddy, A. Costall, G. Bente, T. Schlicht, and K. Vogeley, "Toward a second-person neuroscience 1," *Behavioral and Brain Sciences*, vol. 36, no. 4, pp. 393–414, Aug. 2013.

[6] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente, "Hybrid avatar-agent technology–a conceptual step towards mediated "social" virtual reality and its respective challenges," *i-com*, vol. 14, no. 2, pp. 107–114, 2015.

[7] D. Roth, K. Waldow, M. E. Latoschik, A. Fuhrmann, and G. Bente, "Socially immersive avatar-based communication," in *2017 IEEE Virtual Reality (VR)*. IEEE, 2017, pp. 259–260.

[8] D. Roth, G. Bente, P. Kullmann, D. Mal, C. F. Purps, K. Vogeley, and M. E. Latoschik, "Technologies for social augmentations in user-embodied virtual reality," in *25th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3359996.3364269

[9] S. Lammers, G. Bente, R. Tepest, D. Roth, and K. Vogeley, "Introducing acass: an annotated character animation stimulus set for controlled (e) motion perception studies," *Frontiers in Robotics and AI*, vol. 6, p. 94, 2019.

[10] A. L. Georgescu, B. Kuzmanovic, D. Roth, G. Bente, and K. Vogeley, "The Use of Virtual Characters to Assess and Train Non-Verbal Communication in High-Functioning Autism," *Frontiers in Human Neuroscience*, vol. 8, Oct. 2014.

[11] World Health Organization, Ed., *International statistical classification of diseases and related health problems (11thed.)*. Geneva: World Health Organization, 2019. [Online]. Available: https://icd.who.int/

[12] A. Senju, T. Hasegawa, and Y. Tojo, "Does perceived direct gaze boost detection in adults and children with and without autism? The stare-in-the-crowd effect revisited," *Visual Cognition*, vol. 12, no. 8, pp. 1474–1496, Nov. 2005.

[13] S. Baron-Cohen, S. Wheelwright, and T. Jolliffe, "Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome," *Visual Cognition*, vol. 4, no. 3, pp. 311–331, Sep. 1997.

[14] G. Bird, C. Press, and D. C. Richardson, "The Role of Alexithymia in Reduced Eye-Fixation in Autism Spectrum Conditions," *Journal of Autism and Developmental Disorders*, vol. 41, no. 11, pp. 1556–1564, Nov. 2011.

[15] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," *Archives of General Psychiatry*, vol. 59, no. 9, pp. 809–816, 2002.

[16] D. Neumann, M. L. Spezio, J. Piven, and R. Adolphs, "Looking you in the mouth: Abnormal gaze in autism resulting from impaired top-down modulation of visual attention," *Social Cognitive and Affective Neuroscience*, vol. 1, no. 3, pp. 194–202, Dec. 2006.

[17] K. A. Pelphrey, N. J. Sasson, J. S. Reznick, G. Paul, B. D. Goldman, and J. Piven, "Visual Scanning of Faces in Autism," *Journal of Autism and Developmental Disorders*, vol. 32, no. 4, pp. 249–261, Aug. 2002.

[18] J. Baio, L. Wiggins, D. L. Christensen, M. J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. Robinson, Rosenberg, T. White, M. S. Durkin, P. Imm, L. Nikolaou, M. Yeargin-Allsopp, L.-C. Lee, R. Harrington, M. Lopez, R. T. Fitzgerald, A. Hewitt, S. Pettygrove, J. N. Constantino, A. Vehorn, J. Shenouda, J. Hall-Lande, K. Van, Naarden, Braun, and N. F. Dowling, "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014," *MMWR. Surveillance Summaries*, vol. 67, no. 6, pp. 1–23, Apr. 2018.

[19] L. D. Wiggins, J. Baio, and C. Rice, "Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample," *Journal of developmental and behavioral pediatrics: JDBP*, vol. 27, pp. 79–87, Apr. 2006.

[20] C. M. Corsello, "Early Intervention in Autism," *Infants & Young Children*, vol. 18, no. 2, pp. 74–85, 2005.

[21] M. J. Hollocks, J. W. Lerh, I. Magiati, R. Meiser-Stedman, and T. S. Brugha, "Anxiety and depression in adults with autism spectrum disorder: a systematic review and meta-analysis," *Psychological Medicine*, vol. 49, no. 4, pp. 559–572, Mar. 2019.

[22] P. Howlin, "Social Disadvantage and Exclusion: Adults With Autism Lag Far Behind in Employment Prospects," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 52, no. 9, pp. 897–899, Sep. 2013.

[23] V. Brooke, A. M. Brooke, C. Schall, P. Wehman, J. McDonough, K. Thompson, and J. Smith, "Employees with Autism Spectrum Disorder Achieving Long-Term Employment Success: A Retrospective Review of Employment Retention and Intervention," *Research and Practice for Persons with Severe Disabilities*, vol. 43, no. 3, pp. 181–193, Sep. 2018.

[24] F. J. Scott, S. Baron-Cohen, P. Bolton, and C. Brayne, "The CAST (Childhood Asperger Syndrome Test): Preliminary Development of a UK Screen for Mainstream Primary-School-Age Children," *Autism*, vol. 6, no. 1, pp. 9–31, Mar. 2002.

[25] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians," *Journal of autism and developmental disorders*, vol. 31, no. 1, pp. 5–17, 2001.

[26] C. Lord, M. Rutter, P. C. DiLavore, S. Risi, K. Gotham, and Bishop, *Autism Diagnostic Observation Schedule-2*. Los Angeles: Western Psychological Services, 2012.

[27] C. M. Conner, R. D. Cramer, and J. J. McGonigle, "Examining the

Diagnostic Validity of Autism Measures Among Adults in an Outpatient Clinic Sample," *Autism in Adulthood*, vol. 1, no. 1, pp. 60–68, Mar. 2019.

[28] M. Jiang and Q. Zhao, "Learning Visual Attention to Identify People with Autism Spectrum Disorder," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 3287–3296.

[29] V. Yaneva, L. A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov, "Detecting Autism Based on Eye-Tracking Data from Web Searching Tasks," in *Proceedings of the Internet of Accessible Things*, ser. W4A '18. Lyon, France: ACM, 2018, pp. 16:1–16:10.

[30] A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, and I. Castiglioni, "Use of machine learning to identify children with autism and their motor abnormalities," *Journal of autism and developmental disorders*, vol. 45, no. 7, pp. 2146–2156, 2015.

[31] A. L. Georgescu, J. C. Koehler, J. Weiske, K. Vogeley, N. Koutsouleris, and C. Falter-Wagner, "Machine learning to study social interaction difficulties in asd," *Frontiers in Robotics and AI*, vol. 6, p. 132, 2019.

[32] H. Drimalla, T. Scheffer, N. Landwehr, I. Baskow, S. Roepke, B. Behnia, and I. Dziobek, "Towards the automatic detection of social biomarkers in autism spectrum disorder: Introducing the simulated interaction task (SIT)," *npj Digital Medicine*, vol. 3, no. 1, p. 25, Dec. 2020.

[33] D. M. Riby and P. J. Hancock, "Viewing it differently: Social scene perception in Williams syndrome and Autism," *Neuropsychologia*, vol. 46, no. 11, pp. 2855–2860, Sep. 2008.

[34] A. de Marchena and I.-M. Eigsti, "Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency," *Autism Research*, vol. 3, no. 6, pp. 311–322, Dec. 2010.

[35] D. Roth, P. Kullmann, G. Bente, D. Gall, and M. E. Latoschik, "Effects of Hybrid and Synthetic Social Gaze in Avatar-Mediated Interactions," in *Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018*. Munich, Germany: IEEE, 2018.

[36] S. Y. Oh, J. N. Bailenson, N. Krämer, and B. Li, "Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments," *PLoS ONE*, vol. 11, no. 9, p. e0161794, 2016.

[37] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg, "Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations," *Human Communication Research*, vol. 34, no. 2, pp. 287–318, Apr. 2008.

[38] C. Ahlstrom, T. Victor, C. Wege, and E. Steinmetz, "Processing of Eye/Head-Tracking Data in Large-Scale Naturalistic Driving Data Sets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 553–564, Jun. 2012.

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

[40] M. Chita-Tegmark, "Social attention in ASD: A review and meta-analysis of eye-tracking studies," *Research in Developmental Disabilities*, vol. 48, pp. 79–93, Jan. 2016.

[41] T. W. Frazier, M. Strauss, E. W. Klingemier, E. E. Zetzer, A. Y. Hardan, C. Eng, and E. A. Youngstrom, "A Meta-Analysis of Gaze Differences to Social and Nonsocial Information Between Individuals With and Without Autism," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 56, no. 7, pp. 546–555, Jul. 2017.

[42] Q. Guillon, N. Hadjikhani, S. Baduel, and B. Rogé, "Visual social attention in autism spectrum disorder: Insights from eye tracking studies," *Neuroscience & Biobehavioral Reviews*, vol. 42, pp. 279–297, May 2014.

[43] T. Falck-Ytter and C. von Hofsten, "How special is social looking in ASD: A review," in *Progress in Brain Research*, O. Braddick, J. Atkinson, and G. M. Innocenti, Eds. Elsevier, Jan. 2011, vol. 189, pp. 209–222.

[44] B. Noris, J. Nadel, M. Barker, N. Hadjikhani, and A. Billard, "Investigating Gaze of Children with ASD in Naturalistic Settings," *PLoS ONE*, vol. 7, no. 9, p. e44144, Sep. 2012.

[45] D. Roth and M. E. Latoschik, "Construction of the Virtual Embodiment Questionnaire (VEQ)," *IEEE Transactions on Visualization and Computer Graphics*, pp. 3546–3556, 2020.

[46] M. E. Latoschik, J.-L. Lugrin, M. Habel, D. Roth, C. Seufert, and S. Grafe, "Breaking bad behavior: immersive training of class room management," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016, pp. 317–318.

[47] R. Reynolds, E. Novotny, J. Lee, D. Roth, and G. Bente, "Ambiguous bodies: The role of displayed arousal in emotion [mis] perception," *Journal of Nonverbal Behavior*, vol. 43, no. 4, pp. 529–548, 2019.

[48] D. Roth, M. Jording, T. Schmee, P. Kullmann, N. Navab, and K. Vogeley, "Data and code repository: Towards computer aided diagnosis of asd using virtual environments," Oct 2020. [Online]. Available: 10.17605/OSF.IO/U9PZ3