



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

Mathis Jording^{1,2*}, Arne Hartz^{3,4}, Gary Bente⁵, 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 (Vermetti 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; Dalmasso 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 alternately 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 Foundation¹) 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 ± 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 ± 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>

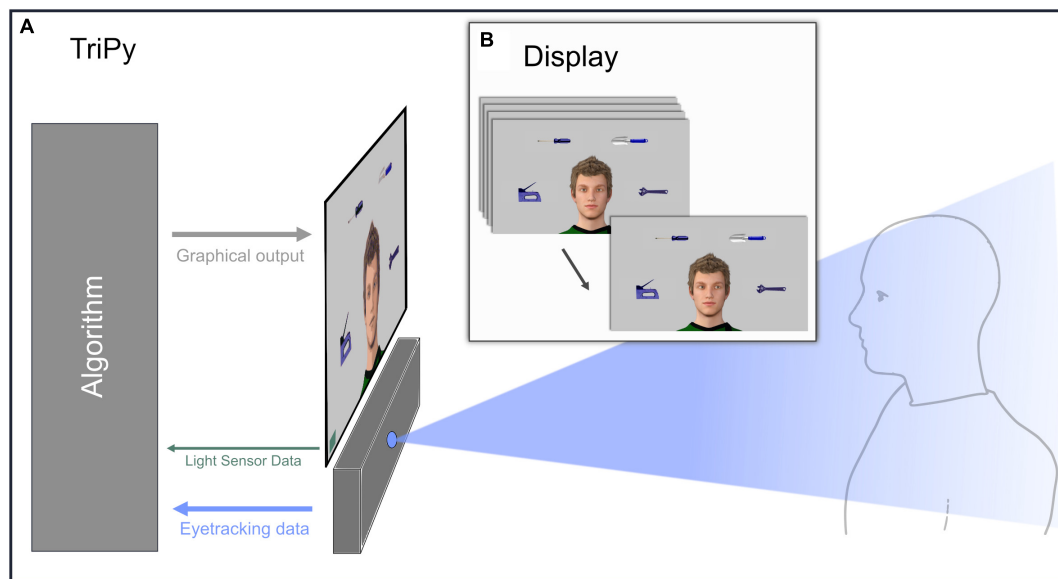


FIGURE 1 | Illustration of the technical setup and the participants' perspective during the experiment. **(A)** Illustration of a participant interacting with the agent controlled by the platform TriPy. **(B)** The behavior of the agent created by TriPy as seen from the perspective of the participant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

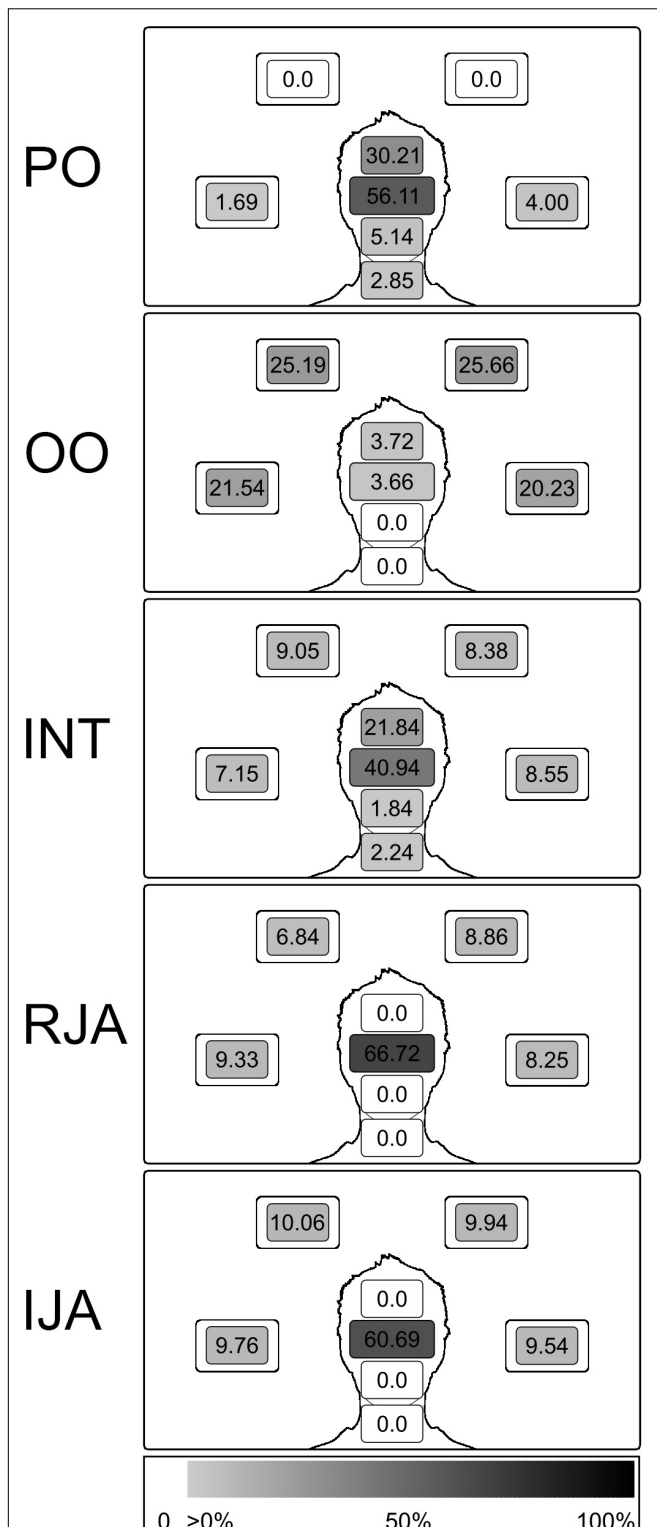


FIGURE 2 | Illustration of the distribution of the agent's visual attention separately cumulated for the different gaze states. Numbers express the rate in percent with which the agent looked at the Aols in total in the specific state as portion of all fixations, color schemes coding serve as additional illustration (white, Aols 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 Aols "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

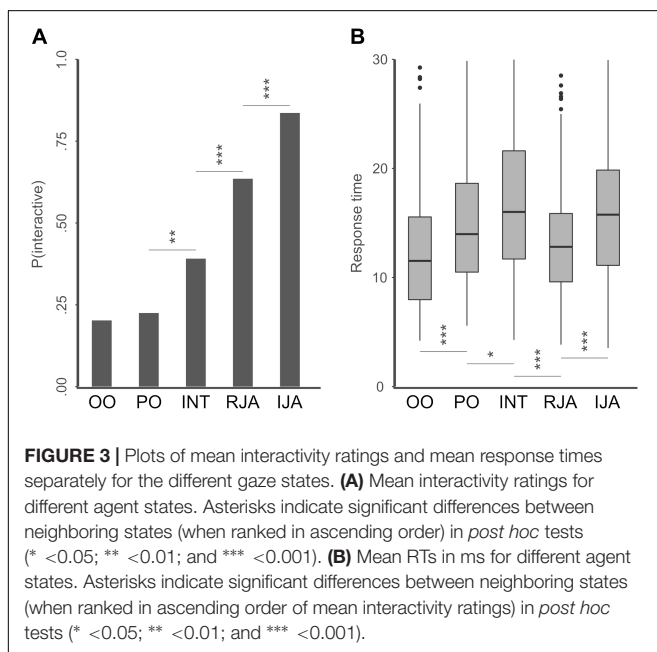
reached significance. In order to investigate whether the quality of the participants' ratings would increase with longer decision time we computed mean correctness scores (RC; correct = "non-interactive" for PO, OO, and INT or "interactive" for RJA and IJA) for each participant. We found a significant 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: $\chi^2(1) = 1.43$, $p < 0.245$].

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



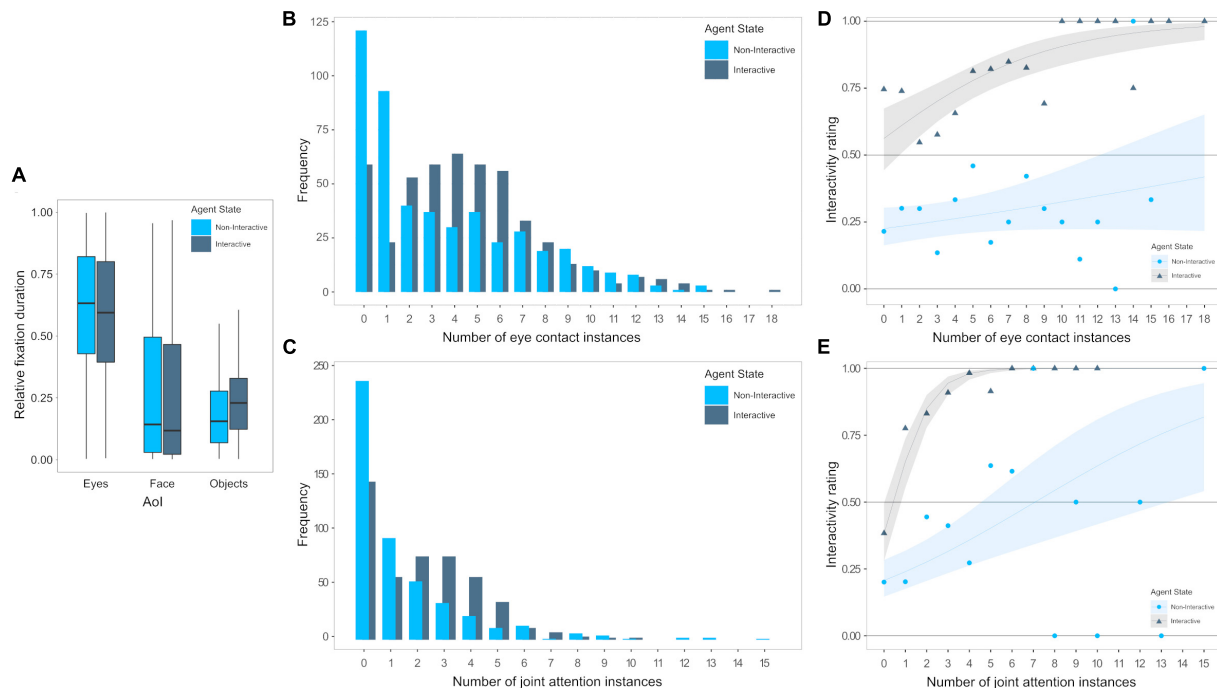


FIGURE 4 | Illustration of the participants gaze behavior and instances of eye contact and joint attention between participant and agent in connection to the participant's rating of the agents interactivity, separately for an agent behaving non-interactively (light blue) vs. interactively (dark blue). **(A)** Boxplots of relative fixation durations as the portion of time spent on the Aols 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, males and 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
- Beccio, 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.
- Dalmajer, 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/s13428-013-0422-2
- 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 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).
- 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
- Oberwilling, 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
- Oberwilling, 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
- Pearce, 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.

Copyright © 2019 Jording, Hartz, Bente, Schulte-Rüther and Vogeley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.