Focusing on the face or getting distracted by social signals? The effect of distracting gestures on attentional focus in natural interaction

Jasmin Kajopoulos*[1,2,3,4], Gordon Cheng[3], Koichi Kise[5], Hermann J. Müller[4,6], Agnieszka Wykowska[7]

[1]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München

[2]International Research Fellow of the Japan Society for the Promotion of Science

[3]Institute for Cognitive Systems, Technical University of Munich

[4]General and Experimental Psychology unit, Dept. Psychologie, Ludwig-Maximilians-Universität München

[5]Dept. of Computer Science and Intelligent Systems, Graduate School of Engineering, Osaka Prefecture University

[6]School of Psychological Science, Birkbeck College, University of London, London, United Kingdom

[7]Social Cognition in Human-Robot Interaction unit, Istituto Italiano di Tecnologia, Genova

* Corresponding author:

Jasmin Kajopoulos, Technical University of Munich, Institute for Cognitive Systems,
Karlstrasse 45, 80333 Munich, Germany
Phone: +49 (89) 289 - 26801
Email: Jasmin.Kajopoulos@tum.de

Abstract

Attentional orienting towards others' gaze direction or pointing has been well investigated in laboratory conditions. However, less is known about the operation of attentional mechanisms in online naturalistic social interaction scenarios. It is equally plausible that following social directional cues (gaze, pointing) occurs reflexively, and/or that it is influenced by top-down cognitive factors. In a mobile eye-tracking experiment, we show that under natural interaction conditions overt attentional orienting is not necessarily reflexively triggered by pointing gestures or a combination of gaze shifts and pointing gestures. We found that participants conversing with an experimenter, who, during the interaction, would play out pointing gestures as well as directional gaze movements, continued to mostly focus their gaze on the face of the experimenter, demonstrating the significance of attending to the face of the interaction partner – in line with effective top-down control over reflexive orienting of attention in the direction of social cues.

Introduction

Gaze behaviour, and other nonverbal signals, such as pointing, can act as salient signals during social interactions. They provide critical information about an interaction partner's focus of attention, as well as her/his intentions, goals, or desires (Baron-Cohen, 1995; Mundy, 2017; Tomasello, 1995). In the laboratory, attentional orienting in response to directional social cues has been often investigated using the gaze-cueing paradigm (Friesen & Kingstone, 1998). It has been proposed that following gaze cues is based on a reflexive, stimulus-triggered orienting mechanism (Driver et al., 1999; Friesen & Kingstone, 1998). However, there is evidence that besides the bottom-up, reflexive component, top-down factors may also influence attentional orienting in response to gaze cues. For example, beliefs about the mental state(s) or inferred action plan(s) of the observed interaction partner (Perez-Osorio, Müller, Wiese, & Wykowska, 2015; Teufel, Fletcher, & Davis, 2010; Wykowska, Wiese, Prosser, & Müller, 2014).

In contrast to the typical laboratory setups designed to study directional social signals, it is important to note that in natural settings, directional gaze hardly ever occurs in isolation; rather, it is almost always accompanied by other non-verbal socially relevant behaviors, such as pointing gestures. It has been shown that like gaze, pointing acts as a strong non-verbal cue that induces reflexive attentional orienting towards the target of the pointing gesture (Hamilton, 2017; Langton & Bruce, 2000). However, during natural social interactions, gaze direction and pointing may not always be congruent; rather, the pointing signals may be irrelevant to the conversation. In essence, they may be a distractor that conveys no meaningful content, and it may even be disadvantageous to orient attention towards the direction of the pointing gesture.

Furthermore, previous studies have shown discrepancies between laboratory studies and real-world experiments. Accordingly, it has been suggested that investigating single social stimuli, such as gaze, without the appropriate social context, such as an interaction, may yield very different findings compared to naturalistic scenarios (Hayward, Voorhies, Morris, Capozzi, & Ristic, 2017; Kompatsiari et al., 2018; Schilbach et al., 2013; Skarratt, Cole, & Kingstone, 2010). In particular, gaze towards an interaction partner has been observed to occur less frequently than expected when the partner was real (Macdonald & Tatler, 2018), compared to previous reports of high amounts of fixations on the eye region when viewing images of people (Macdonald & Tatler, 2018; Birmingham, Bischof, & Kingstone, 2009). Similarly, gaze patterns while viewing the environment have been found to differ when being immersed in the environment as compared to when just viewing the environment on a screen (Foulsham, Walker, & Kingstone, 2011).

Freeth et al. (2013) used mobile eye tracking to measure overt attention during a real-world dyadic interaction in which the experimenter either maintained eye contact with the participant or looked away. Furthermore, they compared this naturalistic scenario to a laboratory setting in which participants viewed a video of an experimenter behaving in a similar manner. Results showed that only during the real-world scenario was participants' gaze affected by eye contact or gaze avoidance, that is, they looked more towards the interaction partner's face when she/he maintained eye contact (Freeth, Foulsham, & Kingstone, 2013). On this background, it is important to ask how attentional orienting mechanisms operate in real-world naturalistic interactive settings, with various directional signals – including not only gaze but also pointing, which can be irrelevant to the task at hand.

Aim of study

We examined how pointing alone or combined with gazing (pointing + gazing) would influence attentional orienting towards the direction of the signal(s), which are non-informative with respect to the ongoing conversation, and thus potentially distracting. In the traditional attentional-capture literature, there is a debate about whether attention is attracted reflexively to the most salient stimulus in the visual field (Theeuwes, 1994, 2004) or whether capture is modulable by top-down control (Bacon & Egeth, 1994; Müller, Geyer, Zehetleitner, & Krummenacher, 2009; Wykowska & Schubö, 2010), and if the latter, to what extent (Awh, Belopolsky, & Theeuwes, 2012; Theeuwes, 2010). Here, we transferred the question of attentional capture from artificial laboratory protocols to a natural, real-life social scenario, specifically: a conversation between an experimenter and a participant, where the experimenter produced directional gestures that were completely irrelevant to the topic of the conversation. Attentional capture on the part of the participant was measured in terms of her/him displaying saccadic activity towards the direction of the experimenter's pointing/gaze. The rationale was that if attention is reflexively engaged by the social gestures, participants should direct their gaze (saccadic activity) towards the location of the gaze and/or pointing gesture; by contrast, if attentional orienting is under top-down control, participants should remain fixated on the interaction partner, as the direction of their pointing/gaze is not relevant to the conversation. Thus, under top-down control, the reflexive gaze- (Driver et al., 1999; Friesen & Kingstone, 1998) or pointing-gesture-following (Hamilton, 2017; Langton & Bruce, 2000) would be suppressed due to being irrelevant to the task at hand, and observers would remain fixated on the face, as attending the speaker is relevant to the task.

Previous studies have argued that social signals, such as gaze or pointing, may be culturally determined, making it advisable, especially in the context of social interaction scenarios, to study gaze/pointing across different cultures (Ellsworth & Ludwig, 1972; Kleinke, 1986). The literature on gaze behavior is largely based on studies conducted in "Western cultures", that is, Europe or North America (Ellsworth & Ludwig, 1972; Kleinke, 1986), with implications for the reproducibility of the results. To ensure that our findings are "inter-culturally robust", we conducted the same study in Germany (Experiment 1) and in Japan (Experiment 2).

2 Material and Methods

Participants

Two experiments were conducted with a similar setup, collecting eye-tracking data from 25 German participants (mainly students or former students; mean age: 25.16, age range 18-35, 6 male) at the Technical University Munich (TUM), Institute for Cognitive Systems (ICS), in Experiment 1 and from 26 Japanese participants (all students; mean age: 20.92, age range, 19-23, 12 male) at Osaka Prefecture University in Experiment 2. 5 German and 4 Japanese participants had to be excluded, as, due to technical failure, the gaze data were not recorded properly. 2 Japanese participants were excluded as they indicated that they were not naive to the purpose of the study. In total, we recruited a usable sample size of 40, aiming for 85% power at an alpha of .05 on the basis of a moderate effect size of Cohen's f = 0.25. This was based on previous research on attentional orienting following pointing cues (Hamilton, 2017) and gaze signals (Wiese, Wykowska, & Müller, 2014) that yielded moderate to large effect sizes ($\eta_p^2 = .346$ - $\eta_p^2 = .909$, required estimated sample size = 38; Wiese et al., 2014).

The experimental procedures consisted of purely behavioral data collection (i.e., gaze activity) and sound-recording of voices; they did not involve any invasive or potentially dangerous methods, and were in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All participants gave written informed consent regarding their participation in the study prior to commencing the experiment. Data were stored and analyzed anonymously.

Apparatus and experimental setup

Gaze activity was recorded with mobile eye-tracker glasses (SMI ETG, SensoMotoric Instruments GmbH (SMI), Germany, cf. Fig. 1) [scene camera resolution = 30Hz, gaze camera resolution = 60Hz], which afford a positional accuracy of 0.5º. Sound was recorded through the integrated microphone. The experimenter was seated opposite to the participant, at a distance of approximately 1 m. The experimenter held a pen in her hand, which was used for producing pointing gestures; to justify the presence of the pen, there was a sheet of paper in front of the experimenter, though this was not used during the pointing phase of the experiment.

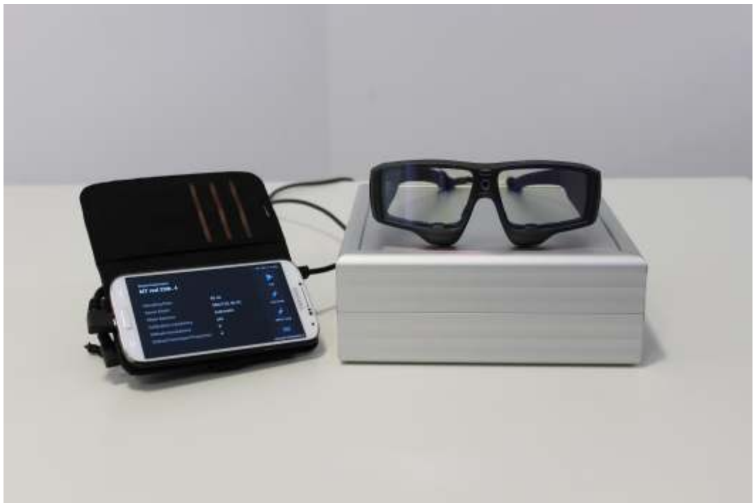

Fig. 1 Mobile eye-tracker glasses (SMI ETG, SensoMotoric Instruments GmbH (SMI), Germany) used for recording

Procedure

The experiment was first conducted in Germany (with a native German-speaking experimenter) and then replicated in Japan (with a native Japanese speaking experimenter). During the interaction, the experimenters wore a neutral black shirt and neutral blue jeans. The interaction script instructed them to mostly show a neutral facial expression, gaze straight ahead towards the participants, and keep their hands in front of them while holding a pen during all phases of the experiment – except for when explicitly instructed to act differently by the interaction script (pointing trials). Both experimenters, in Germany and in Japan, were female.

The participant and the experimenter were seated across from each other at a table (see Fig. 2). After the participant put on the eye-tracker, 1-point calibration was carried out (in rare cases when 1-point calibration failed, 3-point calibration was used). The recording was started immediately afterwards. At the beginning of the recording, participants were asked to fixate different points in the environment, which were later used in a second, offline calibration process (see Online Resource, calibration procedure).

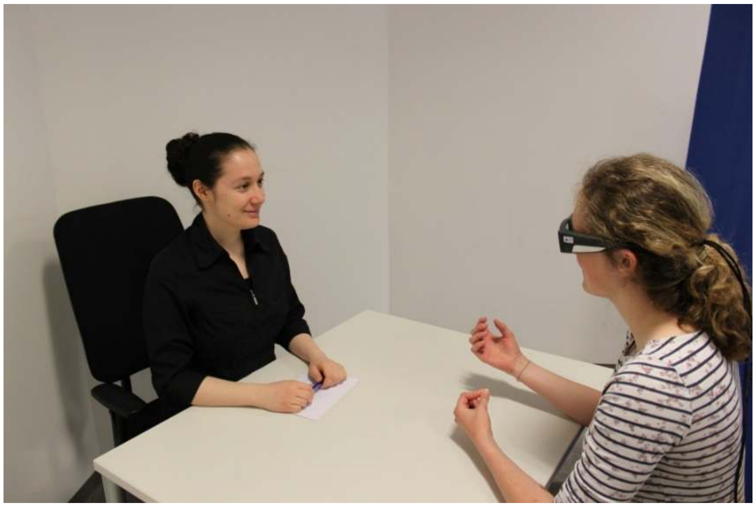

Fig. 2 Experimental set-up. An experimenter (left) and a participant (right) wearing the mobile eye-tracker are sitting across from each other

After the calibration process, participants were verbally told by the experimenter that the experiment involved their filling-out of four questionnaires, during which their gaze would be tracked with the mobile eye-tracker – with the purpose of examining gaze patterns during reading. This was a cover story for participants to make sure that their gaze behaviour during the critical time of the experiment (interaction with the experimenter) would be as natural as possible. That is, the cover story was designed to make participants believe that the critical gaze-tracking procedure was conducted during their filling-out of the questionnaires, while in reality the critical period for gaze tracking was the initial interaction with the experimenter. In the cover story, participants were told that their gaze data was being calibrated before the administration of the questionnaires, which would take a little time; and while the calibration procedure was taking place, the experimenter would have a chat with the participant and ask a few questions. In reality, this was the period of gaze tracking that was the focus of our analysis. Participants were fully debriefed about this procedure and the cover story after the experiment (see Fig.3).

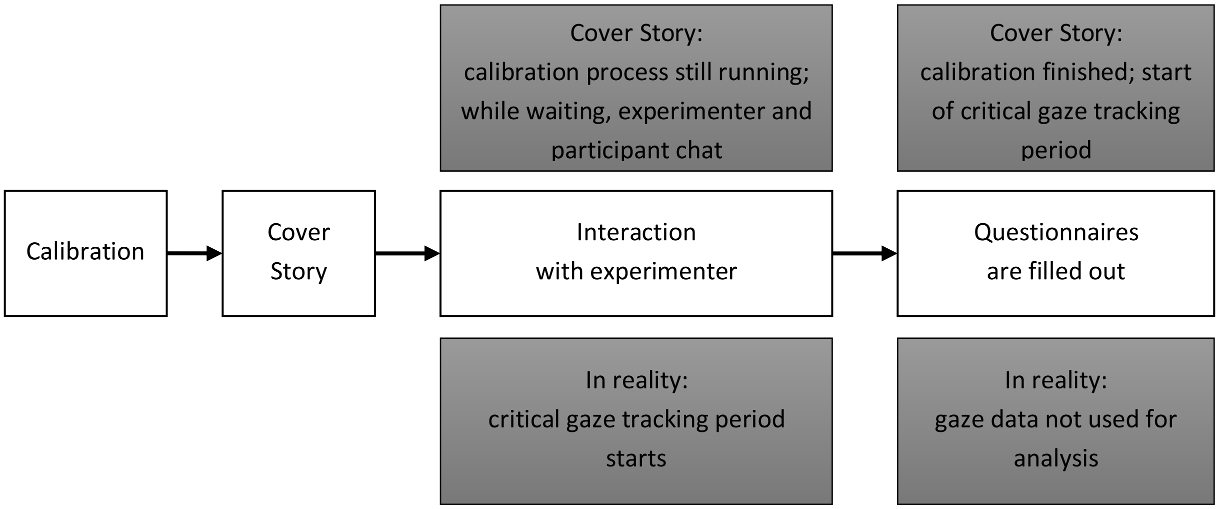
Cover Story:
calibration process still running; while waiting, experimenter and participant chat
Cover Story:
calibration finished; start of critical gaze tracking period
Calibration
Cover Story
Interaction with experimenter
Questionnaires are filled out
In reality:
critical gaze tracking period starts
In reality:
gaze data not used for analysis

Fig. 3 Experimental Procedure

During the period of the natural "chat" with the participants, the experimenter followed a step-by-step interaction script. Before the critical pointing phase, participants were generally familiarized with the interview situation. Thus, after first listening to the experimenter talk about the experiment (L1=Listening, trial 1), the experimenter asked a question (Q1= being Questioned, trial 1). Following this, the participants' answered (A1=Answering, trial 1), which varied in length and content depending on the individual participant. This was repeated 5 times [L1, Q1, A1, L2, Q2, A2, L3, Q3, A3, L4, Q4, A4, L5, Q5, A5], so that by the end of this phase (end of A5), participants would have become used to the interview situation.

Next, the most important phase for the purposes of this study – interaction with pointing – started. Here, the experimenter pointed repeatedly to non-meaningful locations in the environment while talking about various topics. For pointing, the experimenter used a pen in her hand instead of an extended index finger (in order to avoid any culture-related misinterpretations of an extended-finger gesture). The pointing movement was clearly displayed as an aimed directional movement to the left, the right, or downwards. The smaller directional gestures were somewhat exaggerated in their extent, so that pointing would not be mistaken for mere tapping or lifting of the hand.

Following A5 and A6, the experimenter always pointed near (PN=listening trials with pointing near) (Fig. 4b); and after A7 and A8, the experimenter always pointed near while also following the pointing with her gaze (PNg=listening trials with pointing near + gaze) to give a more powerful signal while speaking (Fig. 4d). After A9 and A10, the experimenter always pointed far (PF=listening trials with pointing far, Fig. 4a); and following A11 and A12, the experimenter always pointed far and simultaneously gazed far (PFg=listening trials with pointing far + gaze, Fig. 4c). [...A5, PN, Q6, A6, PN, Q7, A7, PNg, Q8, A8, PNg, Q9, A9, PF, Q10, A10, PF, Q11, A11, PFg, Q12, A12, PFg, Q13, A13...]. (For details see: Online Resource, experimental procedure). Importantly, in order to assure a smooth flow of the interaction, the recording started immediately after the real calibration (before L1). However, for the purposes of this study, only pointing trials (as described above) were included in the analysis.

At the end, and irrelevant to the pointing experiment, the interaction script continued with a Question and Answer-only phase until A25. [Q14, A14, Q15, A15, Q16, A16, Q17, A17, Q18, A18, Q19, A19, Q20, A20, Q21, A21, Q22, A22, Q23, A23, Q24, A24, Q25, A25]. In cases where participants had already answered a question on a previous trial, that question was skipped. All questions asked were neutral with respect to emotional content. The questions were of the following type: “What do you think about e-readers?”; “If you had a tablet, how would you use it?”; ”How do you use your smartphone?” etc. Furthermore, as this experiment was focused on the gaze behavior of participants who listened to the experimenter and watched her make pointing movements, the questions were intended only as fillers in-between the listening sequences to render an ecologically valid and smooth interaction.

The numbers of critical pointing events conducted by the experimenters (on which the data analysis was based) per participant were as follows: for pointing near PN trials, on average 10.92 (sd = 2.54) critical pointing events; for both pointing and gaze PNg trials, 11.00 (sd = 2.14) events; for pointing far PF trials, 12.77 (sd = 3.24) events; and for both pointing and gaze far PFg trials, 11.25 (sd = 1.67) events. A degree of variability in number of pointing events occurred due to the real-world nature of the experiment, that is: while the experimenter followed a script, slight departures from the script could and did happen by accident (i.e., human error).

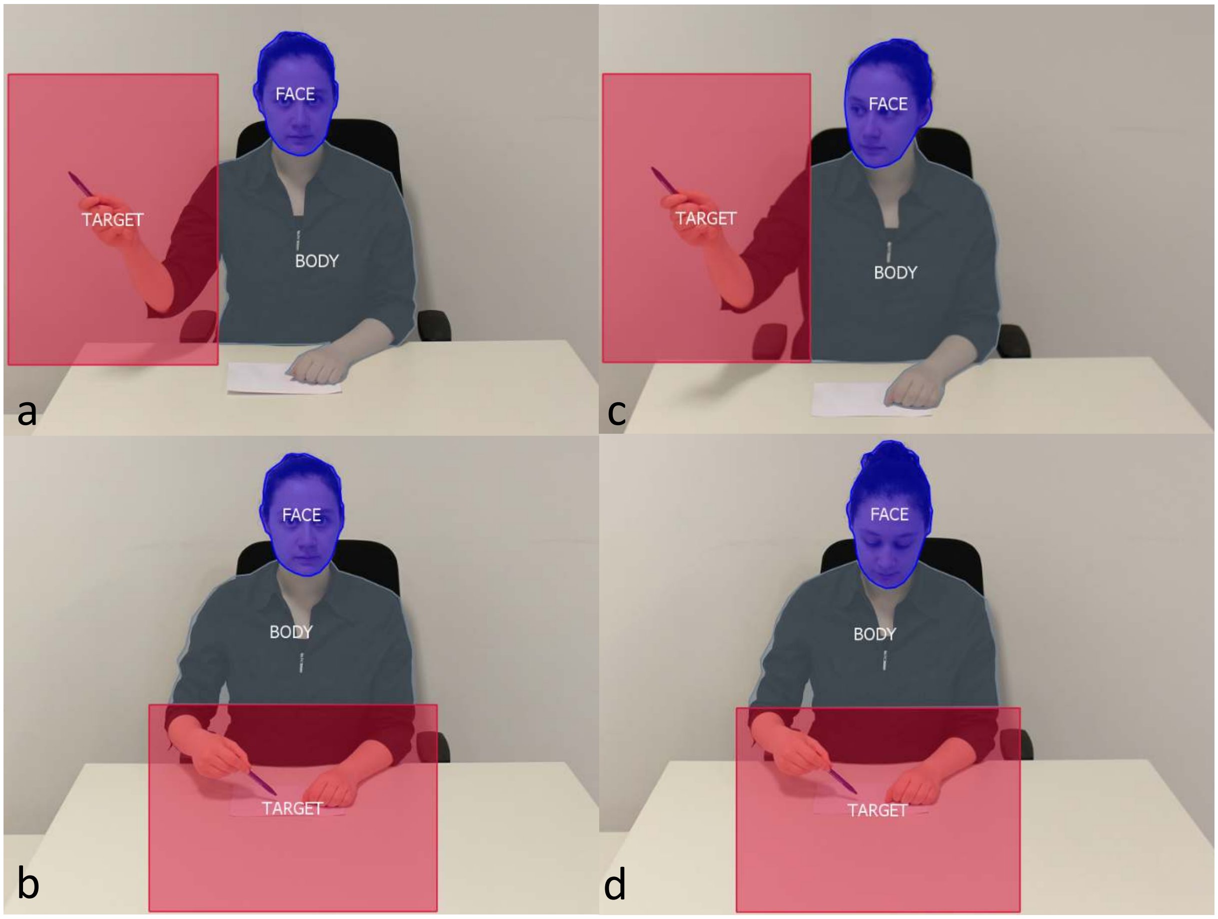

Fig. 4 Sketch of Experimental Procedure, Pointing Phase, Area of Interest (Face, Body, Target, White Space). a = Pointing Far (PF); b = Pointing Near (PN); c = Pointing Far while gazing (PFg); d = Pointing Near while gazing (PNg). Example for Area of Interest Face (blue), Area of Interest Body (grey), Area of Interest Target (red), everything that is not Face, Body or Target is Area of Interest White Space. Importantly, this figure only illustrates the approximate Areas of Interest (AOI). For the Target AOI, all gaze events in the direction of pointing movement not falling on the Body AOI were taken into account; i.e., the target AOI encompassed a larger area than depicted here.

Data Analysis

Initially, data was pre-processed using the BeGaze software (SensoMotoric Instruments GmbH (SMI), Germany) to label all relevant fixation events per Area of Interest (see Fig.4). All gaze events were tagged with an AOI label manually.

All gaze events located on the face of the experimenter were labelled with the Area of Interest (AOI) – Face. All gaze events located on the experimenter's pointing hand and also in the direction of the pointing movement (= all fixations after the first saccade towards the pointing movement with onset of pointing movement and prior to hand returning to starting position; see Fig. 4) were labelled as AOI – Target (collectively, Target AOIs took up 11.5 % of the whole reference image in Fig. 4). The gaze events on the upper body, excluding the face and target, were labelled as AOI – Body. All other gaze events were labelled as AOI – White Space. Of note, AOIs were defined dynamically, taking into account both changes of the moving stimulus (experimenter) and changes of the moving (mobile) eye-tracker. That is, the AOIs were based on landmarks of the stimulus in the real world, rather than being based on coordinates of the scene camera. Hence, rather than being defined in fixed measurements (pixel size or square centimeters), AOIs were defined in relation to the moving stimulus (experimenter) and a reference image (see Fig. 4 for an example) used to label the gaze points.

Fixation-Time [%] Analysis

Fixation Times (sum of durations of all fixations), for all AOIs, for Pointing Distance (Far, Near), and Gaze (With Gaze, Without Gaze), were calculated per participant. If no gaze event occurred inside a particular AOI (mainly for AOI-Target or AOI-Body), this was also included in the analysis as 0-second fixation time for the respective AOI. Data analysis was conducted on mean Fixation Time [%], i.e., mean fixation time / trial duration (for each critical event, one of the epochs PN, PNg, PF, PFg). The 'trial durations', that is, the durations of each pointing epoch, were as follows: for pointing near PN trials, on average 14.27s (sd= 5.52); for pointing near with gaze PNg trials, 12.03s (sd = 2.97); for pointing far PF trials, 13.27s (sd = 3.88); and for pointing far with gaze PFg trials, 10.86s (sd = 3.15). As this experiment was based on a real-world scenario, the trial durations could not be exactly controlled during the experiment. Thus, we divided fixation time by trial duration (per participant!) to calculate Fixation Time [%], thereby controlling for any differences in trial duration.

The resulting Fixation Time [%] scores were examined in a mixed-design Analysis of Variance (ANOVA) with Distance (Far, Near), Gaze (With Gaze, Without Gaze), and AOI (Face, Body, Target, White Space) as within-subjects factors and Experiment (Experiment_1(Germany), Experiment_2 (Japan)) as between-subjects factor. Mauchly's test for sphericity was conducted on all data and, where significant, the *Greenhouse-Geisser corrected* p-values *p[GG]* were calculated (see Online Resource for further assumptions regarding the ANOVA). For the condition means (of interest), 95% Confidence Intervals (CIs) are depicted in the figures.

An additional mixed-design ANOVA was performed with participants' gender (male, female) instead of Experiment as a between-subjects factor; all other factors remained the same.

Probability of Gaze Movements to Areas of Interest

Lastly, the average probability of fixations falling into AOIs Face, Target, Body, or White Space for the first four fixations after onset of directional cues was calculated for every Distance (Far, Near) and Gaze (With, Without Gaze) combination. Probability was calculated by dividing the number of times each participant's fixation fell into either of the AOIs for any of the first four fixations by the participant's number of fixations across all AOIs for any of the first four fixations.

3 Results

Analysis of Variance - Fixation Time %

The mixed-design ANOVA revealed a significant main effect of Area of Interest (Face, Body, Target, White Space), $F(3, 114) = 335.94$, $\eta^2 = 0.86$, *p[GG]* < .001 (Fig.5) and significant interactions between AOI and Gaze, $F(3, 114) = 6.45$, $\eta^2 = 0.01$, *p[GG]* = .006, as well as between AOI and Pointing Distance, $F(3, 114) = 8.26$, $\eta^2 = 0.02$, *p[GG]* = .001.

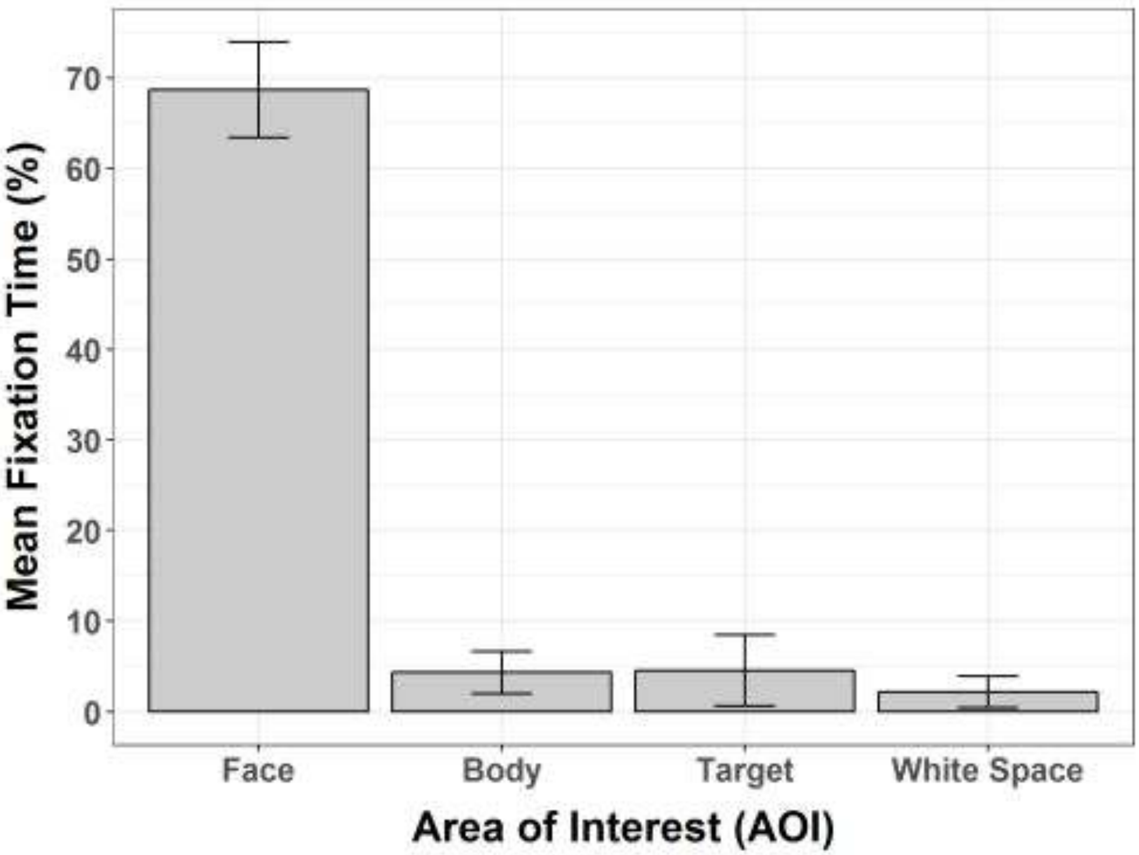

Fig. 5 Mean Fixation Time (%) for Area of Interest (Face, Body, Target, White Space); Error bars: 95% confidence interval (CI).

These effects were modulated by a significant three-way interaction of Distance, Gaze, and AOI, $F(3, 114) = 16.48$, $\eta^2 = 0.03$, $p[GG] < .001$ (Fig. 6).

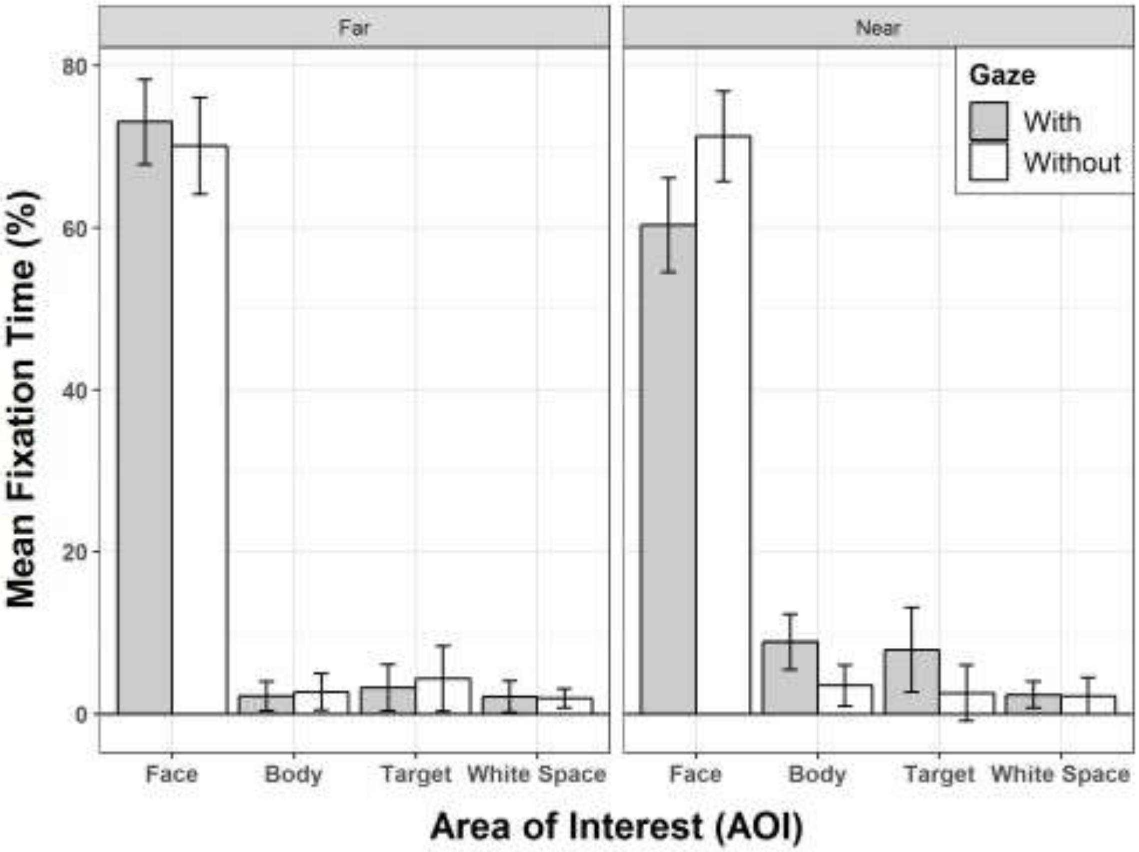

Fig. 6 Mean Fixation Time (%) for Area of Interest (Face, Body, Target, White Space), Distance (Far, Near), and Gaze (With, Without); Error bars: 95% CI.

As can be seen from Figure 6, when the Pointing Distance was Far, there was little difference in the amount of fixation activity on the various AOIs depending on whether or not the pointing gesture was accompanied by an eye movement; in both cases, fixations were more likely directed to the Face than to the Body, Target, or White Space.

This was confirmed by a separate ANOVA for the 'Far' distance: the AOI x Gaze interaction was non-significant ($F(3,114) = 1.55$, $p = .21$ (whereas the main effect of AOI was significant: $F(3,114) = 360.64$, $\eta^2 = 0.89$, $p[GG] < .001$).

By contrast, gaze mattered when the Pointing Distance was Near, as confirmed by a separate ANOVA for the ‘Near’ distance, which revealed the AOI x Gaze interaction to be significant (*F*(3, 114) = 17.62, $\eta^2 = 0.07$, *p[GG]* < .001 (the main effect of AOI was also significant: *F*(3, 114) = 234.50, $\eta^2 = 0.83$, *p[GG]* < .001): Pointing Near accompanied by gaze, as compared to Pointing Near without gaze, reduced the fixation activity on the Face significantly (which, however, still exhibited the highest amount of fixational activity) (With Gaze *M* = 60.32 vs. Without Gaze *M* = 71.26), *t*(39) = -4.29, *p* = .002. At the same time, it increased fixation activity significantly on the Body (With Gaze *M* = 8.87 vs. Without Gaze *M* = 3.50), *t*(39) = 4.84, *p* < .001, and on the Target (With Gaze *M* = 7.90 vs. Without Gaze *M* = 2.59), *t*(39) = 3.49, *p* = .02. By contrast, there were similar amounts of fixation activity on the White Space whether or not Pointing Near was accompanied by gaze (With Gaze *M* = 2.37 vs. Without Gaze *M* = 2.23), *p* > .999.

Also of note, without gaze, the pattern was very similar to that obtained for the Far pointing movement (without gaze, face, near vs. far; without gaze, body, near vs. far; without gaze, target, near vs. far; and without gaze, white space, near vs. far: all *p* > .999). In contrast, with gaze, the pattern slightly differed from that obtained for the Far pointing movement (with gaze, face, near vs. far: *t*(39) = 5.24, *p* < .001; with gaze, body near vs. far: *t*(39) = -4.07, *p* = .004; with gaze, target, near vs. far: *p* = .11; with gaze, white space, near vs. far, *p* >.999).

Thus, the tree-way interaction is explained by the reduced gaze activity on the face and the increased activity on the body and the target in the Pointing Near With Gaze condition compared to all other conditions.

No other main effects or interaction effects reached the level of significance. There was also no significant effect of Experiment (Germany, Japan), all ps > .16.

As for the additional mixed-design ANOVA with participants' gender (male, female) as between-subjects factor, no significant main or interaction effects involving gender were found, all *ps* > .17.

Probability of Gaze Movements to Areas of Interest

This is confirmed by an analysis of the average probability of participants gazing at the Areas of Interest (AOIs) Face, Body, Target, or White Space for the first four fixation positions in participants' fixation sequence following the onset of the directional social cues. For this analysis, we calculated the average frequency with which participants' fixations fell within a certain AOI for a certain position in the fixation sequence, for each of the four types of social cue (Distance and Gaze combination) – see Figure 7 for the results.

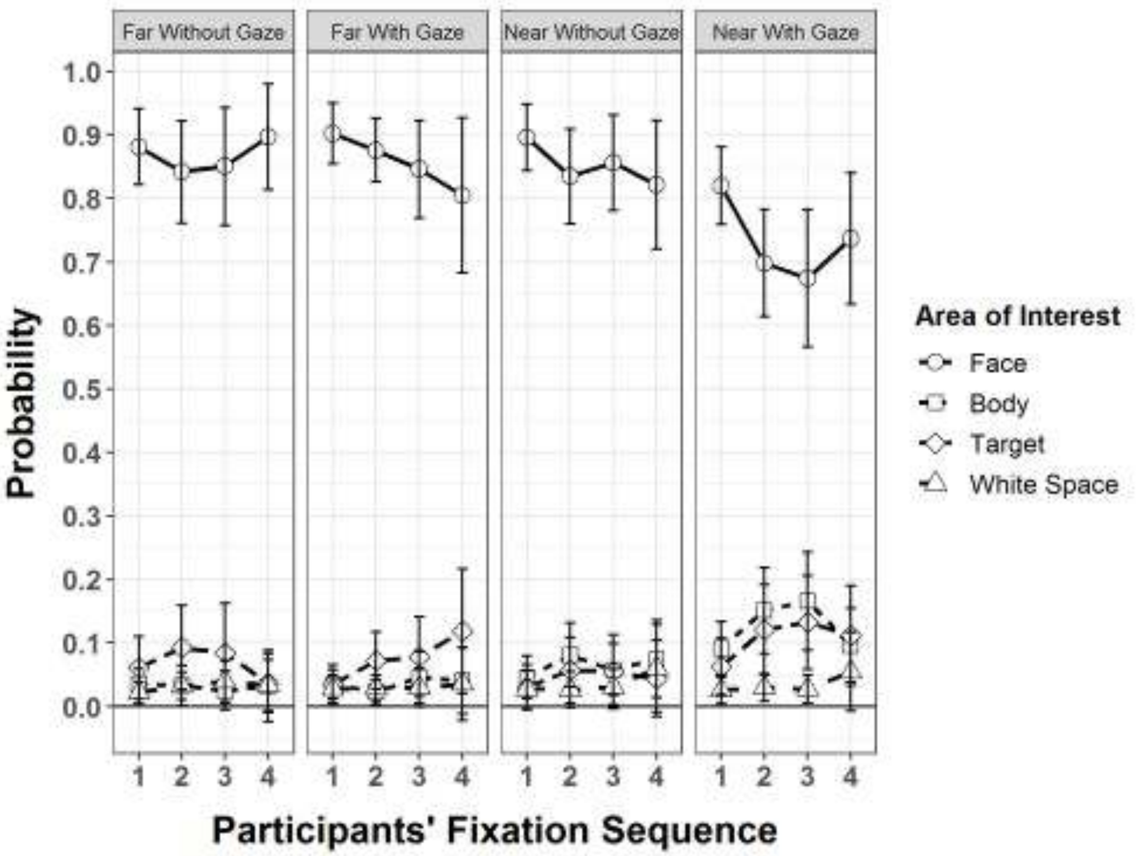

Fig. 7 Probability for fixation falling on Area of Interest Face, Body, Target, or White Space across the first 4 fixations, i.e., the first 4 gaze movements participants made after the onset of the directional cue, for every Distance (Far, Near) and Gaze (With, Without) combination; Error bars: 95% CI.

As can be seen, for all combinations, the first fixation after the onset of the directional cue falls within the face AOI, by a huge margin (without overlap of the 95%-CIs), rather than the Target or any other AOI; and, in fact, among the non-Face AOIs, the target exhibits no clear difference to the remaining AOIs. Essentially the same applies to the second, third, and fourth fixations. This more detailed analysis effectively rules out that, within the sampling period following the directional social cue, participants divert their gaze briefly (and preferentially) to the task-irrelevant Target and then rapidly redirect it to the task-relevant Face of their interaction partner – as would have been predicted, for instance, by the 'capture and rapid disengagement' hypothesis advocated by Theeuwes and colleagues (e.g. Theeuwes, 2010; Theeuwes, Atchley, & Kramer, 2000). We see little evidence of such a pattern in the present, live interaction scenario.

4 Discussion

The goal of the present study was to investigate the responsiveness of participants, indexed by overt attentional orienting in terms of eye movements, to irrelevant pointing gestures and gaze shifts displayed during a natural interaction. It has been proposed that social cues such as gaze and pointing trigger an automatic and reflexive shift of attention, as they potentially provide important information about the other's focus of interest (Friesen & Kingstone, 1998; Friesen, Moore, & Kingstone, 2005; Gluckman & Johnson, 2013; Hamilton, 2017; Langton & Bruce, 2000; Moore & Dunham, 2014; Wykowska et al., 2015). On the other hand, various studies have reported that reflexive shifting of attention may be modulated by higher-order mechanisms of social cognition, based on certain kinds of contextual information (Wiese, Zwickel, & Müller, 2013) as well as assumptions about the cue provider, specifically, the attribution of mental states to the observed agent (Teufel et al., 2010; Wiese, Wykowska, Zwickel, & Müller, 2012; Wiese et al., 2013; Wykowska et al., 2014).

In this context, we aimed at understanding whether, in natural interactions, pointing movements and gaze shifts would trigger reflexive gaze following even when these cues are irrelevant to the content of an ongoing conversation. To this end, we implemented an "attentional-capture" paradigm in a real-life interaction scenario.

Participants conversed with an experimenter, who followed a conversation script wherein pointing gestures alone, or accompanied by gaze shifts, were acted out while conversing.

Results showed that participants were substantially (~10 times) less likely to gaze towards the pointing direction ("Target": < 5% Fixation Time) than maintaining fixation on the face of the experimenter (> 65% Fixation Time, Fig. 5). Importantly, this pattern was evident independently of whether or not the pointing gesture was accompanied by a gaze shift, and independently of the tested sample (German or Japanese participants). That is, even when the experimenter moved her gaze congruently with the pointing direction, participants only very rarely shifted attention away from the experimenter's face. Neither did participants look elsewhere in the environment or at the body of the experimenter, as results indicated also a substantially lower fixation amount for the "White Space" or "Body" compared to the "Face" AOI (Fig. 5). Thus, despite the pointing movements (with or without gaze shifts) during the conversation, participants remained largely fixated at the experimenter's face, exhibiting only minimal orienting of attention towards the directional cues (Figs. 5+6). An even finer-grained, fixation-sequence temporal analysis revealed that following the directional social cue, participants' immediate eye movements (i.e., the first, second, third, and fourth fixations in the sequence) were, by a large margin, more likely to be directed to the face of the experimenter than to the target (which itself was not consistently more likely to be fixated than the remaining AOIs) (Fig.7).

However, when pointing distance and gaze were both taken into consideration (Fig.6), an interesting pattern of gaze behavior emerged. When the experimenter pointed and, at the same time, shifted her gaze (congruently) to a near location, participants were more likely to shift their gaze away from the experimenter's face to either the target or the body. This indicates that gaze might be a more potent signal in capturing attention to irrelevant locations when the gazed-at location is in the "near" space, as delimited by the pointing gesture accompanying the gaze. This is an interesting result indicating that gaze is indeed a relatively potent attention-orienting signal, but perhaps – in some contexts (e.g., a social context like that implemented in the present experiment) – only within a limited spatial region. It appears (Fig. 4) that in the Gaze + Point Near condition, the social shared space was more concentrated and centered around the two interaction partners, while in the Gaze + Point Far condition, the surrounding space was being opened up and expanded (by the pointing gesture) to encompass the external environment. One might speculate that the Gaze + Point Near condition was perceived as more "social"/ "shared". As a result, participants engaged more in joint attention with the experimenter (by following the gaze/pointing directional cue), as compared to a situation where the social context involved a larger area of space not necessarily shared between the interaction partners. This result might therefore suggest that, in natural situations, gaze cueing is a strong attention-orienting mechanism, but only when a sufficient degree of joint social context is established. Interestingly, during the Gaze +Point Near condition some fixations also fell on the body. However, as the body was in-between the face and the target, these fixations might have landed on the body on the way to the target due to the continuous nature of the pointing movement.

It cannot be ruled out, though, that the lack of a similar effect in the condition in which the experimenter was pointing far reflects (in part) an order effect, as the 'pointing-far' trials were always administered after the 'pointing-near' trials. That is, the initial effect of orienting attention in response to the ‘pointing + gaze’ signals might simply have washed out with time on the task. Although, in this case, one would have expected at least some (numerical) effect with ‘pointing far’, the possibility of order effects requires further examination in future research.

One might argue that in our paradigm, the pointing gesture performed by the experimenter was a meaningless signal, not even being a directional cue. However, previous findings indicated that as soon as a stimulus is believed to have eyes, which was the case in our paradigm, automatic shifts of attention cannot be suppressed (Ristic & Kingstone, 2005). Furthermore, there is evidence suggesting that an oculomotor, gaze-following response occurs involuntarily as soon as a gaze cue is perceived (Kuhn & Kingstone, 2009; Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002) – which has led to the proposal that we reflexively mimic our interaction partner’s oculomotor behavior even when detrimental to performance (Kuhn & Kingstone, 2009).

Our study provides striking support for accounts which propose that gaze following in social situations involves a substantial top-down component. It should be noted that the real-life interaction scenario employed here might have played a key role in participants not yielding to attentional capture by effectively ignoring the irrelevant directional cues. Schilbach et al. (2013) pointed out that our understanding of the mechanisms underlying social cognition is often based on detached observations involving static stimuli, with participants employing cognitive mechanisms from an observer’s perspective. However, in order to tap into the actual mechanisms of social cognition, participants ought to be involved in truly interactive social environments. This would enable a component of emotional engagement, which in turn would influence individuals’ response – thus accounting for possible differences between ecologically valid studies with real interactions and laboratory studies with static stimuli (Hayward, Voorhies, Morris, Capozzi, & Ristic, 2017; Schilbach et al., 2013; Skarratt, Cole, & Kingstone, 2010). Indeed, previous studies have shown that being immersed in a social context or the environment in general matters when examining gaze behavior. During an interview situation with a real human as compared to a video opposite, participants engaged more, as measured by fixations on the face, with the interaction partner when eye contact was established. In contrast, no such effect of eye contact was observed with screen-based stimuli (Freeth, Foulsham, & Kingstone, 2013). Similarly, walking around in the environment elicited different gaze patterns compared to viewing the same scenery on a screen (Foulsham, Walker, & Kingstone, 2011). With regard to the present study, our results suggest that, in a natural social interaction setting, task-irrelevant directional cues are hardly ever (reflexively) followed with overt attention. Participants follow directional cues only to some extent and only when they are embedded in a very concentrated space shared closely with their interaction partner.

Thus, it would appear that interacting with a real interaction partner and therefore being immersed in a social context matters with regard to the social signals that are being elicited. This has important implications for the emerging field of social robotics (for reviews, see, e.g., Cabibihan, Javed, Ang, & Aljunied, 2013; Broekens, Heerink, & Rosendal, 2009). With social robots, we are able to bridge the gap from controlled laboratory protocols to embodied interactions (Kompatsiari et al., 2018b), as robots may be programmed to move in a controlled, human-like manner, while being physically present and able to respond to participants in a real interaction (Kompatsiari et al., 2018a, b; Schellen & Wykowska, 2019; Wiese, Metta & Wykowska, 2017; Willemse & Wykowska, 2019; Wykowska, Chaminade & Cheng, 2016; Kajopoulos et al., 2015).

Of course, it might also be the case that, in natural interaction protocols such as the present one, reflexive orienting to ("attentional capture" by) pointing and gaze cues may not be triggered at all (rather than being top-down modulated). Reflexive orienting might only be triggered if a certain threshold of neural activation is reached, which is influenced by stimulus energy, context, and learning history (Cole, Smith, & Atkinson, 2015; Zehetleitner, Hegenloh, & Müller, 2011; Zehetleitner, Krummenacher, & Müller, 2009). Thus, in isolation, such as in laboratory conditions, pointing gestures as well as gaze shifts might be sufficient to elicit attentional orienting. However, in real-life scenarios where there is much more competition from possibly conflicting signals, as well as much richer sensory input, the attention-engaging signals become less salient and may no longer suffice to reach the threshold at which attentional capture is triggered. This possibility needs to be addressed in further research on stimulus saliency in natural environments.

Importantly, culture did not seem to affect gaze behavior in the present scenario. As differences between face and target were quite large and manifested independently of culture, one may speculate that other factors, such as personality or gender, may similarly have had little influence on our results. Specifically, the gender of the experimenter has been suggested to affect participants' performance (Chapman, Benedict, & Schiöth, 2018). While our results show that our participants' gender did not affect their gaze behavior, we cannot rule out effects of the experimenter's gender, as we only had one, female experimenter per participant. Thus, future research may examine more closely for effects of gender (of the experimenter and/or the participants) on attentional orienting in naturalistic interaction scenarios.

In summary, the present study showed that, in a naturalistic setting of a real social interaction, attentional capture measured by overt attention does not necessarily occur towards irrelevant social directional gestures. Instead, humans tend to continuously fixate at the face of their interaction partner. These effects appear to be largely independent of participants' cultural background. This finding is in line with literature showing that attentional orienting has a strong top-down component (Perez-Osorio, Müller, Wiese, & Wykowska, 2015; Teufel, Fletcher, & Davis, 2010; Wykowska, Wiese, Prosser, & Müller, 2014), and does not rely solely on bottom-up reflexive mechanisms (Theeuwes, 1994, 2004). Interestingly, however, it should be noted that various measures might be differently sensitive to various mechanisms of attentional orienting (Kerzel, Zarian, & Souto, 2009; Prinzmetal, McCool, & Park, 2005). For example, gaze cues might affect reaction times more than accuracy. Overt attentional orienting (eye movements) might be more prone to top-down control than covert attention, which may be an interesting question for future research. Future research should also focus on examining further how findings on various cognitive mechanisms might transfer from the screen-based "observational" protocols to more ecologically valid experimental scenarios.

Data Availability Statement

The data are not publicly available due to them containing information that could compromise research participant privacy/consent, but upon reasonable cause a request may be made to the author [J.K.].

Compliance with Ethical Standards

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in cognitive sciences, 16*(8), 437-443.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics, 55*(5), 485-496.

Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and the theory of mind*. Boston: MIT Press/Bradford Books.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Get real! Resolving the debate about equivalent social stimuli. *Visual cognition, 17*(6-7), 904-924.

Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive social robots in elderly care: a review. *Gerontechnology, 8*(2), 94-103.

Cabibihan, J.-J., Javed, H., Ang, M., Jr., & Aljunied, S. (2013). Why Robots? A Survey on the Roles and Benefits of Social Robots in the Therapy of Children with Autism. *International Journal of Social Robotics, 5*(4), 593-618.

Chapman, C. D., Benedict, C., & Schiöth, H. B. (2018). Experimenter gender and replicability in science. *Science advances, 4*(1), e1701427.

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

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.

Ellsworth, P. C., & Ludwig, L. M. (1972). Visual Behavior in Social Interaction. *Journal of Communication, 22*(4), 375-403.

Eraslan, S., Yesilada, Y., & Harper, S. (2015). Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research, 9*(1).

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920-1931.

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *Plos one, 8*(1), e53286.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495.

Friesen, C. K., Moore, C., & Kingstone, A. (2005). Does gaze direction really trigger a reflexive shift of spatial attention? *Brain and cognition, 57*(1), 66-69.

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology, 4*.

Hamilton, S. J. (2017). The Effects of Pointing Gestures on Visual Attention. *University Honors Program Theses, 243*.

Hayward, D. A., Voorhies, W., Morris, J. L., Capozzi, F., & Ristic, J. (2017). Staring reality in the face: A comparison of social attention across laboratory and real world measures suggests little common ground. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 71*(3), 212.

Kajopoulos, J., Wong, A. H. Y., Yuen, A. W. C., Dung, T. A., Kee, T. Y., & Wykowska, A. (2015). Robot-Assisted Training of Joint Attention Skills in Children Diagnosed with Autism. *Social Robotics: 7th International Conference*, *ICSR 2015*, Paris, France, 296-305.

Kerzel, D., Zarian, L., & Souto, D. (2009). Involuntary cueing effects on accuracy measures: Stimulus and task dependence. *Journal of Vision, 9*(11), 16-16.

Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, A., Wykowska, A. (2018a). On the role of eye contact in gaze cueing. *Scientific Reports, 8*, 17842.

Kompatsiari, K., Pérez-Osorio, J., De Tommaso, D., Metta, G., Wykowska, A. (2018b). Neuroscientifically-grounded research for improved human-robot interaction. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, 3403-3408.

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

Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics, 71*(2), 314-327.

Langton, S. R., & Bruce, V. (2000). You must see the point: automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 747.

Moore, C., & Dunham, P., J. (2014). *Joint Attention: Its Origins and Role in Development*. New York: Psychology Press.

Macdonald, R. G., & Tatler, B. W. (2018). Gaze in a real-world social interaction: A dual eye-tracking study. *Quarterly Journal of Experimental Psychology, 71*(10), 2162-2173.

Müller, H. J., Geyer, T., Zehetleitner, M., & Krummenacher, J. (2009). Attentional capture by salient color singleton distractors is modulated by top-down dimensional set. *Journal of Experimental Psychology: Human Perception & Performance, 35*(1), 1-16.

Mundy, P. (2017). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*.

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

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General, 134*(1), 73.

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: exploring the tendency to imitate another individual's gaze. *Neuroreport, 13*(17), 2259-2264.

Ristic, J., & Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition: International Journal of Cognitive Science, 94*(3), B55-65.

Schellen, E., Wykowska, A. (2019). Intentional mindset toward robots—open questions and methodological challenges. *Frontiers in Robotics and AI 5*, 139.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and brain sciences, 36*(4), 393-414.

Skarratt, P. A., Cole, G. G., & Kingstone, A. (2010). Social inhibition of return. *Acta Psychologica, 134*(1), 48-54.

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in cognitive sciences, 14*(8), 376-382.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20*(4), 799-806.

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. M. J. Driver (Ed.), *Attention & Performance*, *18*. (pp. 105−125). Cambridge: MIT Press.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review, 11*(1), 65-70.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77-99.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103-130). USA: Lawrence Erlbaum Associates, Inc.

Wiese, E., Metta, G., Wykowska, A. (2017). Robots as Intentional Agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology, 8*,1663.

Wiese, E., Wykowska, A., & Müller, H. J. (2014). What we observe is biased by what other people tell us: Beliefs about the reliability of gaze behavior modulate attentional orienting to gaze cues. *Plos one, 9*(4), e94529.

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

Wiese, E., Zwickel, J., & Müller, H. J. (2013). The importance of context information for the spatial specificity of gaze cueing. *Attention, Perception, & Psychophysics, 75*(5), 967-982.

Willemse, C., & Wykowska, A. (2019). In natural interaction with embodied robots we prefer it when they follow our gaze: a gaze-contingent mobile eyetracking study. *Philosophical Transactions of the Royal Society B., 374*, 20180036.

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society London: B. Biological Sciences, 371, 20150375.*

Wykowska, A., Kajopoulos, J., Obando-Leitón, M., Chauhan, S., Cabibihan, J.-J., & Cheng, G. (2015). Humans are Well Tuned to Detecting Agents Among Non-agents: Examining the Sensitivity of Human Perception to Behavioral Characteristics of Intentional Systems. *International Journal of Social Robotics*, 1-15.

Wykowska, A., & Schubö, A. (2010). On the temporal relation of top-down and bottom-up mechanisms during guidance of attention. *Journal of Cognitive Neuroscience, 22*(4), 640-654.

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *Plos one, 9*(4).

Zehetleitner, M., Hegenloh, M., & Müller, H. J. (2011). Visually guided pointing movements are driven by the salience map. *Journal of Vision, 11*(1), 24-24.

Zehetleitner, M., Krummenacher, J., & Müller, H. J. (2009). The detection of feature singletons defined in two dimensions is based on salience summation, rather than on serial exhaustive or interactive race architectures. *Attention, Perception, & Psychophysics, 71*(8), 1739-1759.