the british psychological society
promoting excellence in psychology

British Journal of Psychology (2021), 112, 301–314

www.wileyonlinelibrary.com

Stronger reactivity to social gaze in virtual reality compared to a classical laboratory environment

Marius Rubo[1,2]* and Matthias Gamer[2]

[1]Department of Psychology, Freiburg University, Switzerland
[2]Department of Psychology, Julius Maximilians University of Würzburg, Germany

People show a robust tendency to gaze at other human beings when viewing images or videos, but were also found to relatively avoid gaze at others in several real-world situations. This discrepancy, along with theoretical considerations, spawned doubts about the appropriateness of classical laboratory-based experimental paradigms in social attention research. Several researchers instead suggested the use of immersive virtual scenarios in eliciting and measuring naturalistic attentional patterns, but the field, struggling with methodological challenges, still needs to establish the advantages of this approach. Here, we show using eye-tracking in a complex social scenario displayed in virtual reality that participants show enhanced attention towards the face of an avatar at near distance and demonstrate an increased reactivity towards her social gaze as compared to participants who viewed the same scene on a computer monitor. The present study suggests that reactive virtual agents observed in immersive virtual reality can elicit natural modes of information processing and can help to conduct ecologically more valid experiments while maintaining high experimental control.

Humans pay attention to other human beings in many situations and in various ways: they preferentially gaze at human heads and eyes (Birmingham, Bischof, & Kingstone, 2008) as well as objects gazed at by conspecifics (Borji, Parks, & Itti, 2014) when viewing images or videos, automatically form representations of conspecifics' tasks even when there is no incentive to do so (Sebanz, Bekkering, & Knoblich, 2006), and show a similar inhibition when reaching towards objects which were previously touched by themselves or by another person (Welsh *et al*., 2007).

However, while several phenomena in the field of social cognition are typically investigated in real social situations (i.e., where a conspecific is physically present), gaze behaviour is often investigated in participants viewing images or videos of social situations. Interestingly, some studies measuring gaze allocation while other people were physically present found an avoidance of, instead of a preference for looking at them (Gallup, Chong, & Couzin, 2012; Laidlaw, Foulsham, Kuhn, & Kingstone, 2011), although other studies did not replicate this finding (Rubo, Huestegge, & Gamer, 2019). Furthermore, a recent study reported a modulation of social gaze in socially anxious persons, but only in a real-life situation and not in a matched laboratory condition (Rubo *et al*., 2019). This discrepancy

**Correspondence should be addressed to Marius Rubo, Department of Psychology, Freiburg University, Rue P.-A.-de-Faucigny 2, CH-1700 Freiburg, Switzerland (email: marius.rubo@unifr.ch).*

DOI:10.1111/bjop.12453

spawned claims that passive viewing of images or videos may represent inept proxies for real-world social attentional phenomena (Risko, Richardson, & Kingstone, 2016).

When investigating attention in real-life situations, however, precise experimental control is virtually impossible, and even confederates are not capable of reacting in identical ways towards each participant. As one methodological approach to simultaneously achieve high ecological validity as well as experimental control, researchers have suggested the use of virtual reality (VR) technology (Pan & Hamilton, 2018). Here, the idea is to model experiences in real-world social situations more closely compared to traditional laboratory-based set-ups, while the participants are able to freely look around in the scenario and virtual agents are naturally acting and responding to the participant. Critically, observing scenes in VR typically creates a sense of *presence* or *being there* in the scenario (Skarbez, Frederick Brooks, & Whitton, 2017), which is not the case when looking at depictions on a computer screen. Nonetheless, there is still only relatively scarce empirical evidence that social scenarios observed in VR really do elicit more natural modes of information processing (Kulik, 2018). Zimmer, Buttlar, Halbeisen, Walther, and Domes (2019) found comparable self-report, autonomic and endocrine stress markers in response to a standardized social stress test carried out by real or virtual conspecifics. Wienrich, Gross, Kretschmer, and Muller-Plath (2018) documented an inhibition of return effect towards virtual avatars – an effect which was previously believed to only be triggered by other human beings. In a study by Gallup, Vasilyev, Anderson, and Kingstone (2019), by contrast, contagious yawning was inhibited in the presence of a real conspecific, but not in the presence of a virtual agent. In the latter study, however, the virtual agent showed no meaningful social behaviour – a possible prerequisite for perceiving an agent as believable and activating natural modes of social information processing.

In the present study, a first group of participants was located in a complex social scenario in VR and had the opportunity to observe a reactive virtual agent with a naturalistic behavioural repertoire. A second group of participants viewed the same scene from the same perspective, but on a computer monitor (thus without being *present* in the virtual social situation). Assuming that a virtual agent witnessed in VR elicits naturalistic modes of information processing, we expected a relative avoidance of participants' gaze at the agent's face, but a reactivity to the agent's social behaviour (i.e., looking and smiling at the participant). By contrast, we expected such patterns of naturalistic gaze behaviour to be absent, or weaker when the same scene is viewed on a computer monitor. By comparing gaze behaviour between a situation seen in VR and on a computer screen in direct juxtaposition, the present study aims to reveal more detailed insights into the role of immersion or presence when experiencing social situations. This way, we aim to provide data for the discussion around the ecological validity of VR and more classical laboratory environments. We additionally tested whether inter-individual differences in attentional preferences were stable throughout the experiment – a largely unexplained phenomenon which was recently discovered in participants viewing images (de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2019; Guy et al., 2019), videos (Rubo & Gamer, 2018), or a real-life situation (Rubo *et al.*, 2019).

Methods

Participants, apparatus, and software

We tested 80 participants of whom 40 viewed a social scene in VR and 40 viewed the same scene on a computer screen (PC group). Participants in the VR group (32 females, mean age = 24.62 years, *SD* = 6.83 years) watched the scene using an HTC Vive system (2,160 × 1,200 pixel resolution, 110° field of view) while a built-in eye-tracker (SMI Eye Tracking VR HMD) recorded their gaze direction from both eyes at a sampling rate of 60 Hz. Experimental stimuli were displayed using the Unity 3D Game Engine (https://unity3d.com), and the virtual agent's behaviour was controlled using in-house software described at https://github.com/mariusrubo/Unity-Humanoid-TransportObjects which was built using an inverse kinematics algorithm package (www.root-motion.com). The sample size for this group and the experimental procedure were determined a priori and were pre-registered (AsPredicted #14290, see https://aspredicted.org/5wj8u.pdf). Participants in the PC group (30 females, mean age = 26.46 years, *SD* = 6.46 years) watched video clips from the same virtual scene on a 24-inch LCD monitor (Asus VG248QE, 53.136 × 29.889 cm, 1,920 × 1,080 pixels, refresh rate 60 Hz) at a viewing distance of approximately 50 cm. Video clips had a resolution of 1,280 × 720 pixels (resulting in a visual angle for the videos of 39.01° horizontally × 22.54° vertically) and a frame rate of 30 Hz. In the PC group, eye movements were recorded from the right eye using an EyeLink 1000 Plus system (SR Research, Ottawa, ON, Canada) at a sampling rate of 250 Hz, but data were averaged within video frames. Head location was fixed using a chin rest and a forehead bar. All participants had normal or corrected-to-normal vision. The study was approved by the ethics committee of the Department of Psychology, University of Würzburg, and conducted in accordance with the Declaration of Helsinki.

Procedure

Participants in both groups were invited to the laboratory individually, informed about the purpose of the study and completed an informed consent form. They were instructed to merely observe the scenario (see Figure 1) for about 6 min, and participants in the VR group were furthermore asked not to walk away from the location where they were positioned. The virtual scenario consisted of a suburban neighbourhood in which a garage sale was being prepared on a driveway in front of a house. Participants in the VR group found themselves located at the driveway while a virtual character – a woman in the age of roughly 60 years – was engaged in carrying household objects (e.g., a radio, a lamp, a toy car) from her house onto two tables positioned approximately 2 m in front of the participant. Participants in the PC group viewed the scene from the same visual perspective as participants in the VR group.

Altogether, the virtual character carried 10 items onto the tables during the experiment, each time approaching the participant in a similar manner: after leaving the veranda with an item in her hands, the character walked a straight path with a distance of approximately 12 m through the front yard before placing the item in front of the character. The character then turned around and walked the same itinerary back and into her house, picking up the next item. Participants were therefore confronted with 10 similarly structured opportunities to gaze at the character or the object carried in her hands as she approached (ten *trials*). The character furthermore glanced at the participant and smiled in random trials (but never showing the same behaviour in more than two consecutive trials), starting at a distance of approximately 4 m.

Prior to the start of the scenario, the eye-tracker was calibrated using a 5-point calibration technique in the VR group and a 9-point calibration and validation technique in the PC group. In the VR group, gaze measurement validation was performed using in-house software to allow for a drift correction procedure and to achieve detailed insight into potential problems with gaze measurement. The general implementation of the

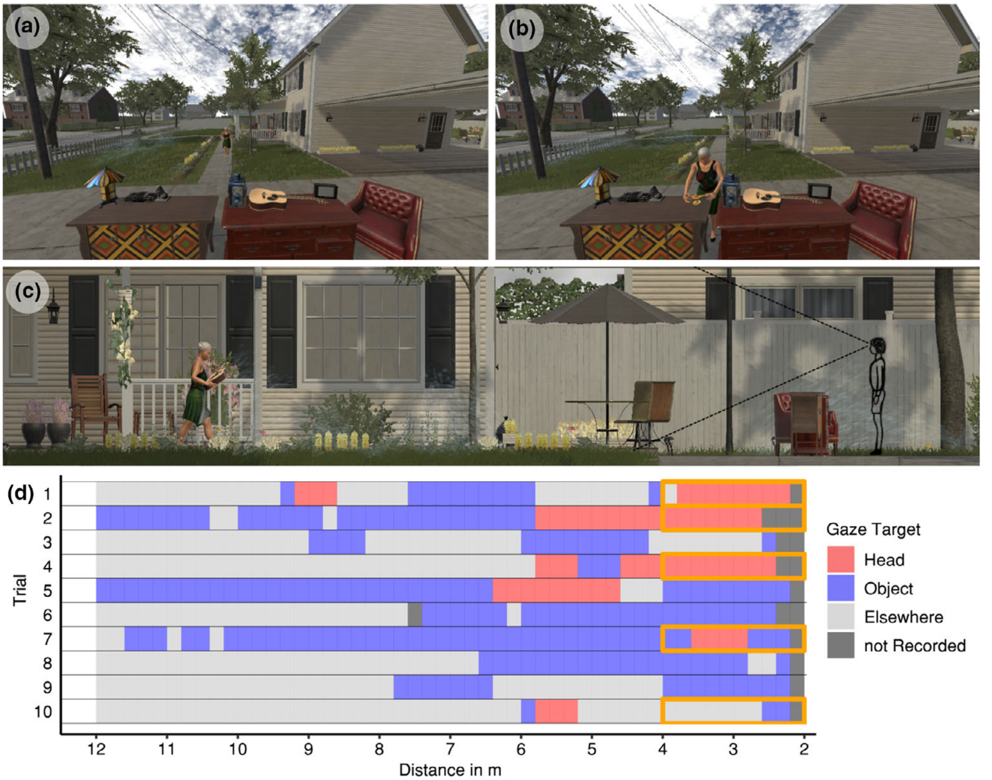

Figure 1. The social situation taking place in the virtual scenario. Images (a) and (b) are from the perspective of the participants in moments when the virtual character has just left her house carrying an object and when she places the object on the table in front of the participant, respectively. (c) Shows the scene from a lateral perspective (as an orthographic projection), with the participant and her typical viewing angle (black dotted lines) pictured on the right of the image. (d) Illustrates one exemplary participant's gaze while the virtual character is approaching the participant. The distance scale corresponds to (c). For this visualization, gaze is categorized as being directed towards the character's head or the object if either is being gazed within a distance bin of 20 cm (e.g., while the avatar was at a distance between 3.60m and 3.80m from the observer). Orange frames indicate moments when the virtual character looked at the participant and smiled. [Colour figure can be viewed at wileyonlinelibrary.com]

procedure is described at https://github.com/mariusrubo/Unity-EyeTracking-RegionsOfInterest. Specifically, we presented a red sphere in the virtual scenario with a diameter of 10 cm and a distance to the participants of about 4 m. Participants were instructed to gaze at the sphere until it vanished about three seconds later. This allowed us to compare the participant's gaze rays from both eyes as they were estimated by the eye-tracker with hypothetical gaze rays perfectly hitting the sphere's centre. We geometrically transposed this arrangement to represent the measured and the ideal rays' relative slope as a two-dimensional deviation, comparable to the shot on a target and parallel to monitor-based eye-tracking where gaze deviations are described along the screen's *x*- and *y*-axes. This validation process was performed directly after the initial calibration and repeated several times throughout the experiment. Specifically, validation was performed in moments when the virtual avatar was walking back into her house, since the participants' gaze behaviour in these moments was not of interest for the present study. If validation was unsuccessful, calibration was performed again. Data from all validation procedures were used to correct for drifts in gaze measurement (see Data processing).

Upon completing the experiment, participants in the VR group were asked to report on their sense of *presence* in the virtual environment (Skarbez *et al.*, 2017) using the German version of the *Igroup Presence Questionnaire* (IPQ; Schubert, Friedmann, and Regenbrecht (2001)) as well as on sensations of simulator sickness using the *Simulator Sickness Questionnaire* (SSQ, Kennedy, Lane, Berbaum, and Lilienthal (1993)). Participants gave moderate ratings for presence ($M = 57.25$, $SD = 11.06$, on a scale from 14 to 98) and low ratings for simulator sickness ($M = 3.35$, $SD = 3.29$, on a scale from 0 to 48). Participants in both groups furthermore filled out a sociodemographic questionnaire.

Data processing and statistical analyses

Data were analysed using R for statistical computing (version 3.2; R Development Core Team, 2015). We were only interested in participants' gaze behaviour while the virtual character approached the participants, that is, from the moment the character stepped onto the garden path (at a distance of approximately 12 m from the participants) to the moment the character placed the object on the table in front of the participants and turned around.

Gaze data collected during the validation procedures in the VR group were used to correct for drifts in gaze measurements resulting from shifts in the HMD's position on the participants' head. To exclude moments when participants did not look at the validation sphere during the validation procedures (e.g., when the sphere had only just appeared and participants were not yet looking at it), drift correction was performed using only data points when participants' gaze from both eyes missed the validation sphere by less than 15° (this was the case in 93.25%, $SD = 9.59\%$, of data points in all validation procedures). In one participant, the eye-tracker only correctly tracked the left eye (during all validation procedures, observed gaze from the left eye missed the validation sphere by 2.45°, while observed gaze from the right eye missed the sphere by 64.69°). In this case, only data from the left eye were used for further analyses.

We used a recursive outlier removal algorithm to estimate drift in each validation process. Separately for the deviation measurements in x- and y-directions of each eye, the lowest and highest values were both removed from the distribution, individually compared to the distribution of the remaining data and entered again if they were located within 3 standard deviations from the mean. This process was recursively applied to the remaining data until both the highest and the lowest data points met the criterion to be re-entered to the distribution (for a similar procedure see End & Gamer, 2017; Rubo & Gamer, 2018). The values in the remaining distributions were averaged to arrive at *baseline deviation* values.

Since eye-tracking in VR is not yet a standard research method, we furthermore used the data recorded during the validation procedure (i.e., while participants were asked to gaze at a sphere) to describe the method's measurement precision. For each validation phase and for both eyes individually, we computed the average deviation of valid recordings from the whole measurement's mean and aggregated these values within participants (collapsing across both eyes and all validation phases). Across all participants, measurements during the calibration phase varied by $M = 1.29^{\circ}$ ($SD = 1.22^{\circ}$).

In the VR group, gaze data were represented as deviations from two regions of interest (ROIs) – the character's face and the object she was carrying – in an x- and y-direction and separately for both eyes (see Figure 2). We subtracted baseline deviation values from both eyes' measurements individually and then averaged deviations from both eyes to arrive at a deviation representing both eyes' gaze. Gaze data which could not be tracked during the

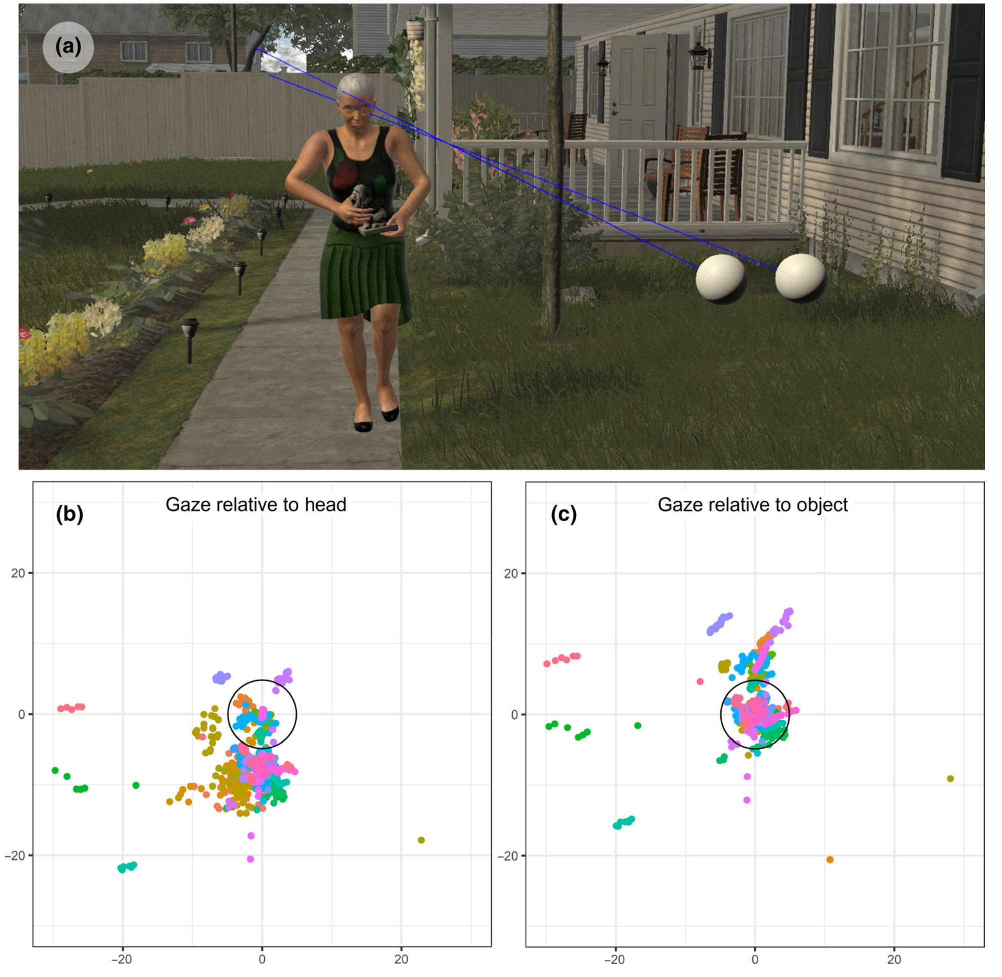

Figure 2. Relating gaze recordings to the two ROIs (head or object) in the VR condition. (a) In 3D space, gaze is represented as a ray consisting of an origin position and a direction vector. The image (a) shows gaze rays for one exemplary participant in one moment in time. The white spheres indicate the position of the two eyes with the blue line showing gaze direction. In this case, the gaze rays are colliding with the virtual agent's head. In order to classify gaze directions as being directed towards one of the two ROIs, we transposed gaze rays to represent angular deviations from each ROI in the horizontal and vertical directions using an approach available at https://github.com/mariusrubo/Unity-EyeTracking-RegionsOfInterest which also allowed to perform drift corrections. (b) Exemplarily shows this angular deviation for agent's head for all participants (differentiated by colour) while the agent was carrying one specific object and was between 3 m and 4 m away from the participant. (c) Shows the same data, but transposed to represent deviations from the object. Data points inside the black circle are labelled as being directed towards the respective ROI (specifically, the shown circle corresponds to the decision limit when the distance is 3.5 m; for smaller and larger distances, the circle is bigger or smaller, respectively). ROI = regions of interest; VR = virtual reality. [Colour figure can be viewed at wileyonlinelibrary.com]

relevant phases of the experiment due to eye blinking or technical issues (M = 7.83%, SD = 7.15% of the time) were removed from the dataset. For the valid data, we classified for each time point if gaze was directed towards the virtual agent's head, the object in her hands, or elsewhere. To achieve similar a priori probabilities for gaze at both ROIs, gaze was classified to hit either of the two if it collided with a hypothetical sphere of a diameter of 30 cm in the position of each ROI. This procedure implies that the angular range in which gaze is classified as being directed towards one of the two ROIs increases as the virtual agent approaches (illustrated in Figure 3).

In order to investigate viewing behaviour in same situation viewed on a computer screen, we recorded videos of the virtual scene from the participants' perspective along with videos from the same perspective, but where only spheres of 30 cm diameter around both ROIs (avatar's head and object) were visible (*ROI* videos). These videos were used to relate gaze data from the stationary eye-tracker – which represents points on the 2-dimensional plane that makes up the monitor – with the scene's three-dimensional geometry which is naturally bent and flattened when rendered onto a video. Using this approach, decisions about when gaze was directed towards either or none of the ROIs followed the same logic in both the VR and the PC conditions (see Figure 4). In the PC group, M = 3.45% (SD = 2.24%) of gaze data were not available due to eye blinking or technical issues and were removed from the dataset.

For each participant and trial, we aggregated relative gaze dwell times towards both ROIs within distance bins with a width of one metre, and, for analyses across trials, further aggregated gaze dwell times across trials. This approach allows to analyse the relative gaze dwell times at the two ROIs as a function of the distance towards the ROIs. We tested the influence of individual parameters on gaze allocation by means of linear mixed models where participant ID was inserted as a random effect. All parameters were tested for significance using an F-test, with α set to 0.05. Reported correlations are Neyman–Pearson correlations.

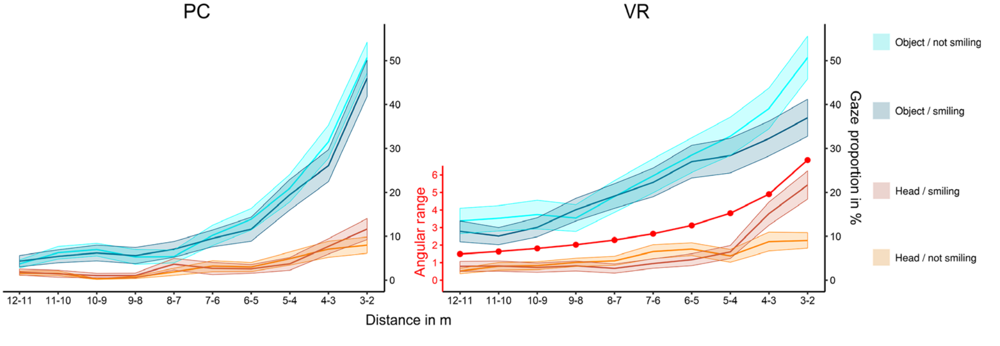

Figure 3. Gaze allocation as a function of the virtual character's distance and its social behaviour (smiling at the participant while approaching or merely looking at her hands and not smiling). Gaze data are allocated within distance bins with a width of one metre. Ribbons represent SEM. For the VR group, the red points (and the corresponding y-axis on the left of the figure, also printed in red) highlight the angular range in degrees within which the participants' gaze is classified as being directed towards each of the two regions of interest. Note that as the virtual character approaches, this angular range enlarges, thereby possibly increasing the measurement's sensitivity. The effect is similar in the PC group, but cannot easily be measured in degrees due to geometric distortions when rendering a 3D scene to a video. [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 4. Relating gaze recordings to the two ROIs (head or object) in the PC condition. Along with a recording of the scene, we also recorded videos where only the ROIs were visible (superimposed here as green and red circle). This approach allows to automatically assign gaze (data from several participants depicted as black crosses) to the ROIs along the same logic as in the VR condition (in spite of the scene's distortion which naturally occurs when rendering to an image). This still frame is taken from an example video available at https://osf.io/nqv8x/. ROI = regions of interest; VR = virtual reality. [Colour figure can be viewed at wileyonlinelibrary.com]

Results

Gaze allocation as a function of the Agent's distance

We first aimed to broadly describe the gaze pattern in both groups as a function of the character's distance (see Figure 3), therefore including the ROI (the character's head or the object in her hands), the distance, the group (VR or PC) as well as interactions between these factors as fixed effects into a linear mixed model and the percentage of gaze dwell time as dependent variable. We found statistically significant effects of ROI ($F(1, 3{,}114) = 914.40$, $p < .001$), distance ($F(1, 3{,}114) = 832.51$, $p < .001$), group ($F(1, 78) = 9.30$, $p = .003$), the ROI × distance interaction ($F(1, 3{,}114) = 270.51$, $p < .001$), ROI × group interaction ($F(1, 3{,}114) = 39.26$, $p < .001$), and the ROI × distance × group interaction ($F(1, 3{,}114) = 6.47$, $p = .011$), but no distance × group interaction ($F(1, 3{,}114) = 0.09$, $p = .769$). To follow up on group effects, we then performed congeneric analyses for both ROIs individually. With regard to gaze on the head, we found a statistically significant group effect ($F(1, 78) = 5.84$, $p = .018$), a distance effect ($F(1, 1{,}518) = 221.28$, $p < .001$), and a distance × group interaction ($F(1, 1{,}518) = 7.28$, $p = .007$). With regard to gaze on the object, we found a group effect ($F(1, 78) = 7.07$, $p = .010$), a distance effect ($F(1, 1{,}518) = 886.29$, $p < .001$), and a marginally significant distance × group interaction ($F(1, 1{,}518) = 3.48$, $p = .062$). Specifically, participants in the VR group showed more gaze allocation towards the virtual agent's head (VR: $M = 6.05\%$, $SD = 10.60\%$; PC: $M = 3.49\%$, $SD = 8.19\%$) as well as the object (VR: $M = 23.29\%$, $SD = 24.38\%$; PC: $M = 14.81\%$, $SD = 20.61\%$). Across both groups, there was a negative correlation between distance and gaze on the head ($r = -.31$, 95% CI = [−0.35, −0.26], $p < .001$) and, even more strongly so, between distance and gaze on the object (PC: $r = -.47$, 95% CI = [−0.50, −0.43], $p < .001$). The latter correlation was stronger in the PC group (head: $r = -.29$, 95% CI = [−0.36, −0.23], $p < .001$, object: $r = -.55$, 95% CI = [−0.60, −0.50], $p < .001$) than in the VR group (head: $r = -.33$, 95% CI = [−0.39, −0.26], $p < .001$, object: VR: $r = -.41$, 95% CI = [−0.47, −0.35]). In sum, participants in both groups spent more time looking at the object compared to the character's head and furthermore increased gaze dwell times towards both ROIs as the character approached. Participants in the VR group spent more time looking at both ROIs, but participants in the PC group showed a stronger relative increase in gaze towards the objects as the agent approached.

Reactivity to the Agent's social behaviour

We then focused on time points when the virtual avatar was closer than 4 m to the participant – the distance range in which she looked and smiled at the participant in random trials. Here, we describe the relative gaze dwell time as a function of the ROI, the character's social behaviour (smiling or not smiling), the group, and an interaction between these factors. We found a statistically significant effect of ROI ($F(1, 554) = 343.53$, $p < .001$), a ROI × smiling interaction ($F(1, 554) = 19.90$, $p < .001$), and a ROI × smiling × group interaction ($F(1, 554) = 4.30$, $p = .039$), but no main effect of the group ($F(1, 78) = 1.35$, $p = .248$) and no ROI × group ($F(1, 554) = 1.65$, $p = .200$) or smiling × group interaction ($F(1, 554) = 0.12$, $p = .726$).

We again performed congeneric analyses for both ROIs to follow up on group effects. With regard to gaze on the head, we found a group effect ($F(1, 78) = 4.84$, $p = .031$), a smiling effect ($F(1, 238) = 20.86$, $p < .001$), and a smiling × group interaction ($F(1, 238) = 8.29$, $p = .004$). With regard to gaze on the object, we found a smiling effect ($F(1, 238) = 14.36$, $p < .001$), but no group effect ($F(1, 78) = 0.06$, $p = .810$) and no smiling × group interaction ($F(1, 238) = 1.65$, $p = .200$).

Specifically, across all trials, the head was gazed at more in the VR group ($M = 13.67\%$, $SD = 16.44$) than in the PC group ($M = 8.61\%$, $SD = 12.48$), while gaze towards objects was similar in both groups (VR: $M = 39.75\%$, $SD = 28.54$; PC: $M = 38.56\%$, $SD = 25.63\%$). The virtual agent's smiling resulted in a stronger increase in gaze towards the head in the VR group (without smiling: $M = 8.93\%$, $SD = 12.31\%$; with smiling: $M = 18.41\%$, $SD = 18.64\%$) than in the PC group (without smiling: $M = 7.53\%$, $SD = 11.61\%$; with smiling: $M = 9.68\%$, $SD = 13.28\%$). The decrease of gaze on the object following smiling was largely similar between the VR group (without smiling: $M = 44.88\%$, $SD = 30.34\%$; with smiling: $M = 34.61\%$, $SD = 25.78\%$) and the PC group (without smiling: $M = 41.09\%$, $SD = 24.75\%$; with smiling: $M = 36.03\%$, $SD = 26.39\%$)

To sum up, participants in both groups gazed more at the object compared to the head while the virtual agent was near. However, participants in the VR group gazed more at the character's head compared to participants in the PC group and more strongly shifted their gaze preference towards the head in trials when the agent was smiling.

Variations in Gaze Behaviour throughout the experiment

We subsequently analysed the variation of gaze patterns throughout the course of the experiment (see Figure 5). Again focusing on time points when the virtual avatar was closer than 4 m, we describe gaze dwell time as a function of the trial, the region of interest

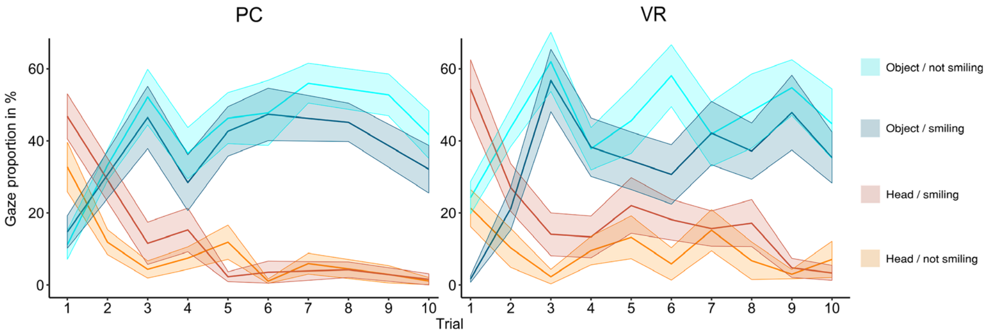

Figure 5. Gaze allocation along the experiment's 10 trials while the virtual agent was nearer than 4 m. Note that in each trial and for each participant, the virtual agent either did or did not smile at the participant while approaching, and data are therefore aggregated along different subsets of participants within each trial and condition (smiling vs. not smiling). Error ribbons represent SEM. [Colour figure can be viewed at wileyonlinelibrary.com]

(ROI), the avatar's smiling, the group, and interactions between these factors. This analysis revealed a general effect of ROI ($F(1, 1{,}504) = 460.58, p < .001$), a trial × smiling interaction ($F(1, 1{,}504) = 6.21, p = .013$), a trial × ROI interaction ($F(1, 1{,}504) = 93.67, p < .001$), a smiling × ROI interaction ($F(1, 1{,}504) = 33.82, p < .001$), a marginally significant trial × smiling × ROI interaction ($F(1, 1{,}504) = 3.40, p = .065$), and a smiling × ROI × group interaction ($F(1, 1{,}504) = 4.79, p = .029$), but no statistically significant main effects of trial ($F(1, 1{,}504) = 0.51, p = .477$), the agent's smiling behaviour ($F(1, 1{,}504) = 0.51, p = .476$), and group ($F(1, 78) = 0.55, p = .462$) as well as no smiling × group interaction ($F(1, 1{,}504) = 0.00, p = .945$), trial × group interaction ($F(1, 1{,}504) = 0.04, p = .849$), ROI × group interaction ($F(1, 1{,}504) = 2.00, p = .158$), trial × smiling × group interaction ($F(1, 1{,}504) = 0.47, p = .494$), trial × ROI × group interaction ($F(1, 1{,}504) = 2.07, p = .150$), or trial × smiling × ROI × group interaction ($F(1, 1{,}504) = 1.27, p = .260$).

Following up on group effects, we again performed congeneric analyses for both ROIs individually. With regard to gaze allocation on the head, we found a main effect of the trial ($F(1, 713) = 97.83, p < .001$) and the character's smiling ($F(1, 713) = 23.55, p < .001$) as well as a trial × smiling interaction ($F(1, 713) = 15.77, p < .001$) and a smiling × group interaction ($F(1, 713) = 4.67, p = .031$), but no main effect of the group ($F(1, 78) = 2.67, p = .106$), trial × group interaction ($F(1, 713) = 2.31, p = .130$), or trial × smiling × group interaction ($F(1, 713) = 0.07, p = .796$).

With regard to gaze allocation on the objects, we found statistically significant main effects of trial ($F(1, 713) = 36.28, p < .001$) and the character's smiling ($F(1, 713) = 19.24, p < .001$), but no main effect of the group ($F(1, 78) = 0.00, p = .984$), trial × smiling interaction ($F(1, 713) = 0.35, p = .554$), smiling × group interaction ($F(1, 713) = 2.06, p = .152$), trial × group interaction ($F(1, 713) = 0.66, p = .418$), or trial × smiling × group interaction ($F(1, 713) = 1.13, p = .288$).

Across both groups, gaze on the head decreased with the trial, and more so in trials when the agent smiled ($r = -.40$, 95% CI = [−0.48, −0.31], $p < .001$) compared to when it did not smile ($r = -.22$, 95% CI = [−0.31, −0.13], $p < .001$). This was the case both in the VR group (smiling: $r = -.36$, 95% CI = [−0.47, −0.23], $p < .001$; not smiling: $r = -.14$, 95% CI = [−0.27, 0.00], $p = .050$) and in the PC group (smiling: $r = -.47$, 95% CI = [−0.59, −0.34], $p < .001$; not smiling: $r = -.30$, 95% CI = [−0.41, −0.18],

$p < .001$). By contrast, gaze on the objects increased similarly in trials when the agent smiled ($r = .16$, 95% CI = [0.06, 0.26], $p = .002$) and did not smile ($r = .19$, 95% CI = [0.10, 0.28], $p = .002$). This was the case both in the VR group (smiling: $r = .17$, 95% CI = [0.03, 0.30], $p = .018$; not smiling: $r = .13$, 95% CI = [−0.01, 0.26], $p = .069$) and in the PC group (smiling: $r = .15$, 95% CI = [−0.01, 0.30], $p = .060$; not smiling: $r = .25$, 95% CI = [0.13, 0.37], $p < .001$).

Consistency of Gaze behaviour within participants

Additionally, we estimated the consistency of gaze preferences within individual observers for phases when the virtual agent was near. To this end, we randomly split the 10 trials into two halves for each participant, aggregated relative gaze dwell times on both ROIs across both trial groups and computed split-half correlations of gaze dwell times between the two halves. We repeated this procedure 1,000 times and determined the 2.5th and 97.5th percentile rank among correlation estimates in all iterations as an unbiased estimate of the 95% confidence interval. In both groups, inter-individual gaze consistency was somewhat higher with regard to the objects (VR: $r = .73$, 95% CI = [0.62, 0.83]), PC: $r = .65$ [0.53, 0.78]) compared to the head (VR: $r = .41$ [0.21, 0.63], PC: $r = .45$ [0.22, 0.64]).

Discussion

The present study confronted participants with a naturalistic social situation and a reactive social agent presented in a virtual environment, either viewed in VR or on a computer monitor. Participants in both groups spent more time fixating objects in the agent's hands compared to the agent's head. Participants in both groups increased gaze dwell times towards both of these regions of interest while the agent approached, and did so more strongly with regard to gaze on objects (especially in the PC group). The higher amount of gaze allocated towards the objects – especially after the first trials, when gaze towards the head was still more frequent – is most parsimoniously explained by a novelty effect (Desimone, 1995): while participants saw the same character throughout the experiment, the object changed every trial. The increase in gaze towards either the character's head or the object as it approached may be due to a better visibility to the participants as well as to increased measurement accuracy, and this effect may interact with the generally larger interest in gazing at the objects.

When the virtual agent was near, its social behaviour (smiling at the participant in random trials) became a relevant variable. Here, contrary to our prediction, participants in the VR group spent more time gazing at the character's head compared to participants in the PC group. This aspect of gaze alteration in VR may be seen to speak against a more natural mode of information processing in this environment, since some (Gallup *et al.*, 2012; Laidlaw *et al.*, 2011) – although not all (Rubo *et al.*, 2019) – previous studies found less gaze towards conspecifics in a more real compared to a more artificial social situation. However, other studies pointed to a more complex relation between gaze on conspecifics and the situation's perceived realness. For example, in a study by Gobel, Kim, and Richardson (2015), participants avoided gaze contact with higher-ranked individuals, but sought gaze contact with lower-ranked individuals when they were led to believe that the gaze recipients could later view the recorded gaze patterns. Summing up, we feel that the interpretation of this finding requires a deeper understanding of the variables that may increase or decrease gaze contact throughout a social situation.

Interestingly, while participants across both groups shifted their gaze preference towards the head when the agent smiled at them, this behaviour was far more pronounced in the VR compared to the PC group (see Figure 3). In our view, the clear emergence of this everyday behaviour in the VR group – and the relative lack thereof in the PC group – more unequivocally points to a more natural mode of information processing and behaviour in VR compared to when observing a scene on a computer screen. At the same time, it must be noted that in this relatively lifelike scenario, a variety of variables were not individually controlled, possibly diluting attributions of statistical effects to clear mechanisms. For instance, we did not systematically vary the agent's facial expression when looking at the participant (she always began to smile in these occasions), leaving us unable to decide if participants reacted to the agents *looking* at them or *smiling* at them. Moreover, in rotating the head when looking at the participant, the virtual character also causes an additional movement in the scene, and as motion itself is known to attract gaze (Itti, 2005; Mital, Smith, Hill, & Henderson, 2010), this factor and its differential effects between different viewing conditions may contribute to differences in viewing behaviour as well. We therefore suggest that future studies should include a stronger variability of behaviour – that is, smiling behaviour including stronger or weaker head movements – and introduce measures of physical saliency which have not yet been incorporated into VR research. In our view, however, future research should not make the systematic variation of individual variables the only priority in designing experiments since this strategy would bear the risk of creating stilted and unnatural scenarios which cannot harvest the methodology's potential to stimulate naturalistic behaviour. Instead, we argue that social attention research using VR should employ even less tabulated scenarios than the one used in the present study, possibly even abandoning the repetition of homogeneous trials and instead immersing participants in social situations as semantically rich as many situations we encounter in real life. Confronting participants with a variety of semantically coherent situations may nonetheless allow to carve out contributions of individual factors to attention allocation, but will also help to identify broad and robust patterns which can remain hidden in overly standardized experiments (Würbel, 2000).

Besides predictors of viewing behaviour across individuals, we additionally found substantial inter-individual differences in gaze allocation towards both ROIs in both groups, which is in line with previous findings of stable inter-individual differences when viewing images (de Haas *et al.*, 2019; Guy et al., 2019), videos (Rubo & Gamer, 2018) and real-life situations (Rubo *et al.*, 2019). The stability of these differences was similar in both groups, but somewhat higher with regard to objects compared to the head – which may be explained by larger and differential reactions to the character's smiling.

The present study compared social attention in VR and a paralleled PC condition using a lifelike virtual scenario where a reactive virtual agent staged a complex process taken from real life. Most strikingly, participants in the VR group were more reactive towards the agent's social behaviour – looking and smiling at the participant – compared to participants in the PC group. We suggest that this form of behaviour hints towards the perceived naturalness of the VR situation, which, unlike viewing a scene on a computer monitor, allows participants to act in a more ecological manner. Adding to previous findings in the field, this study demonstrates the potential for heightened levels of ecological validity in VR compared to more traditional laboratory research.

Acknowledgements

This work was supported by the European Research Council (ERC-2013-StG #336305). We thank Jasin Sahraoui and Aki Schumacher for their help in data collection.

Conflicts of interest

All authors declare no conflict of interest.

Author contributions

Marius Rubo (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing) Matthias Gamer (Conceptualization; Funding acquisition; Investigation; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing)

Pre-registration

The study protocol was defined prior to the experiment (AsPredicted #14290, see https://aspredicted.org/5wj8u.pdf).

Data availability statement

The data that support the findings of this study are available at https://osf.io/nqv8x/.

References

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition*, *16*(2–3), 341–355. https://doi.org/10.1080/13506280701434532

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

De Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences*, *116*(24), 11687–11692.

Desimone, R. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.neuro.18.1.193

End, A., & Gamer, M. (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, *8*, 418. https://doi.org/10.3389/fpsyg.2017.00418

Gallup, A., Chong, A., & Couzin, I. (2012). The directional flow of visual information transfer between pedestrians. *Biology Letters*, *8*, 520–522. https://doi.org/10.1098/rsbl.2012.0160

Gallup, A., Vasilyev, D., Anderson, N., & Kingstone, A. (2019). Contagious yawning in virtual reality is affected by actual, but not simulated, social presence. *Scientific Reports*, *9*(1), 294. https://doi.org/10.1038/s41598-018-36570-2

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

Guy, N., Azulay, H., Kardosh, R., Weiss, Y., Hassin, R. R., Israel, S., & Pertzov, Y. (2019). A novel perceptual trait: Gaze predilection for faces during visual exploration. *Scientific Reports*, *9*(1), 359–364. https://doi.org/10.1038/s41598-019-47110-x.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*, 1093–1123. https://doi.org/10.1080/13506280444000661

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Kulik, A. (2018). Virtually the ultimate research lab. *British Journal of Psychology*, *109*, 434–436. https://doi.org/10.1111/bjop.12307

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

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2010). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, *3*, 5–24. https://doi.org/10.1007/s12559-010-9074-z

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

R Core Team (2015). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing.

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science. *Current Directions in Psychological Science*, *25*, 70–74. https://doi.org/10.1177/0963721415617806

Rubo, M., & Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-22127-w

Rubo, M., Huestegge, L., & Gamer, M. (2019). Social anxiety modulates visual exploration in real life but not in the laboratory. *British Journal of Psychology*, *111*, 233–245. https://doi.org/10.1111/bjop.12396

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Skarbez, R., Frederick Brooks, J., & Whitton, M. (2017). A survey of presence and related concepts. *ACM Computing Surveys*, *50*, 1–39. https://doi.org/10.1145/3134301

Welsh, T. N., Lyons, J., Weeks, D. J., Anson, J. G., Chua, R., Mendoza, J., & Elliott, D. (2007). Within- and between-nervous-system inhibition of return: Observation is as good as performance. *Psychonomic Bulletin & Review*, *14*, 950–956. https://doi.org/10.3758/bf03194127

Wienrich, C., Gross, R., Kretschmer, F., & Muller-Plath, G. (2018). Developing and proving a framework for reaction time experiments in VR to objectively measure social interaction with virtual agents. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 191–198). https://doi.org/10.1109/VR.2018.8446352

Würbel, H. (2000). Behaviour and the standardization fallacy. *Nature Genetics*, *26*(3), 263–263. https://doi.org/10.1038/81541.

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., & Domes, G. (2019). Virtually stressed? A refined virtual reality adaptation of the trier social stress test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*, *101*, 186–192. https://doi.org/10.1016/j.psyneuen.2018.11.010

Received 4 April 2019; revised version received 5 May 2020