Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–16
© Experimental Psychology Society 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1747021819829718
qjep.sagepub.com
SAGE

Responding to joint attention bids in schizophrenia: An interactive eye-tracking study

Nathan Caruana[1,2], Kiley Seymour[1,3], Jon Brock[1,2] and Robyn Langdon[1,2]

Abstract

This study investigated social cognition in schizophrenia using a virtual reality paradigm to capture the dynamic processes of evaluating and responding to eye gaze as an intentional communicative cue. A total of 21 patients with schizophrenia and 21 age-, gender-, and IQ-matched healthy controls completed an interactive computer game with an on-screen avatar that participants believed was controlled by an off-screen partner. On social trials, participants were required to achieve joint attention by correctly interpreting and responding to gaze cues. Participants also completed non-social trials in which they responded to an arrow cue within the same task context. While patients and controls took equivalent time to process communicative intent from gaze shifts, patients made significantly more errors than controls when responding to the directional information conveyed by gaze, but not arrow, cues. Despite no differences in response times to gaze cues between groups, patients were significantly slower than controls when responding to arrow cues. This is the opposite pattern of results previously observed in autistic adults using the same task and suggests that, despite general impairments in attention orienting or oculomotor control, patients with schizophrenia demonstrate a facilitation effect when responding to communicative gaze cues. Findings indicate a hyper-responsivity to gaze cues of communicative intent in schizophrenia. The possible effects of self-referential biases when evaluating gaze direction are discussed, as are clinical implications.

Keywords

Joint attention; eye-tracking; eye gaze; social cognition; social perception; schizophrenia

Received: 30 July 2018; revised: 28 November 2018; accepted: 13 December 2018

Social interactions are integral to our daily lives but can present significant challenges for many people with schizophrenia (Billeke & Aboitiz, 2013; Fiszdon, Fanning, Johannesen, & Bell, 2013). Impaired social cognition, particularly theory of mind (ToM)—the capacity to infer others' thoughts, intentions, or beliefs—is consistently associated with problematic social behaviours and reduced daily functioning in schizophrenia (Brüne, 2005; Couture, Penn, & Roberts, 2006). For this reason, it has been suggested that ToM impairments may partially explain why many individuals with schizophrenia find social interactions both confusing and stressful (Pallanti, Quercioli, & Hollander, 2004).

Researchers propose that the development of ToM depends on joint attention (Baron-Cohen, 1995). Joint attention is the ability to coordinate attention with a social interlocutor to attend to the same thing. This social skill involves one person initiating a joint attention bid—for instance, by intentionally shifting their gaze towards an object—and another person recognising that joint attention bid as intentional, before responding to achieve joint attention (Bruinsma, Koegel, & Koegel, 2004). Responding to joint attention bids is particularly interesting, as it requires the effective perception of a social cue (e.g., eye gaze), as well as an appropriate evaluation of its

---

[1]ARC Centre of Excellence in Cognition and Its Disorders, Macquarie University, Sydney, NSW, Australia
[2]Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia
[3]School of Social Sciences and Psychology, Western Sydney University, Sydney, NSW, Australia

Corresponding author:
Nathan Caruana, Department of Cognitive Science, Macquarie University, Balaclava Road, North Ryde, Sydney, NSW 2109, Australia.
Email: nathan.caruana@mq.edu.au

social significance and intentionality (Senju & Johnson, 2009). In this way, responsive gaze-based joint attention requires the synthesis of accurate eye-gaze perception and ToM processing of others' intentionality to successfully identify and respond to others' attempts for social interaction (Mundy & Newell, 2007).

Previous separate lines of research have examined gaze processing and ToM in schizophrenia. With regard to the former, a large and growing body of research has investigated visual processing and evaluation of faces in schizophrenia (see Watson, 2013 for review) with a subset of these studies focusing specifically on either the ability to accurately discriminate the direction of eye gaze and/or shift attention accordingly. Current research in this area suggests that there is no fundamental impairment in the early visual perception of eye-gaze information in schizophrenia (e.g., Palmer, Caruana, Clifford, & Seymour, 2018a, 2018b). Recent work using the continuous flash suppression (CFS) paradigm to examine early unconscious stages of gaze perception has revealed that patients and healthy controls demonstrated the same patterns of gaze processing (Seymour, Rhodes, Stein, & Langdon, 2016). Specifically, faces with direct gaze entered conscious awareness faster than faces with averted gaze in both patients and controls, and equally so, even though gaze direction was irrelevant to the task.

Other studies have evidenced intact gaze processing in schizophrenia using gaze-based Posner (1980) cueing tasks. Typically, healthy adults are faster to detect a peripheral target when its presentation is preceded by a congruent gaze cue and slower when preceded by an incongruent gaze cue (e.g., Frischen, Bayliss, & Tipper, 2007). Critically, these effects occur despite participants being instructed to ignore the gaze cue. Thus, the congruency effect indexes the extent to which attention is automatically oriented by gaze direction. To date, the only evidence of impaired automatic orienting to others' gaze in schizophrenia comes from studies using highly stylised or schematic "gaze" stimuli (e.g., Akiyama et al., 2008; Dalmaso, Galfano, Tarqui, Forti, & Castelli, 2013). Contrastingly, gaze-cueing studies that use biologically realistic stimuli (e.g., photographs) provide relatively consistent evidence of intact automatic orienting to both gaze (e.g., Magnee, Kahn, Cahn, & Kemner, 2011; Seymour et al., 2017) and head-orientation cues (Langdon, Corner, McLaren, Coltheart, & Ward, 2006) in schizophrenia.

There is also evidence from these cueing paradigms that patients with schizophrenia demonstrate not only *intact* but also an *enhanced* gaze congruency advantage in some conditions (see, e.g., Langdon et al., 2006, for findings related to head-orientation cues of gaze direction and Langdon, Seymour, Williams, & Ward, 2017, for findings related to eye-shift cues in a direct face). For instance, Langdon et al. (2017) found that relative to healthy controls, patients were faster to detect a visual target's location following the presentation of congruent, compared with incongruent, gaze-shift cues at longer (800 ms), but not shorter (100 and 300 ms), cue-target intervals. Importantly, differences between groups were not observed during a non-social control task where gaze cues were replaced by arrows, indicating that this was a gaze-specific advantage. The later temporal locus of these group differences suggests that conscious cognitive processes or biases which manifest later in the evaluation of social stimuli (such as when deliberative ToM processing occurs) may contribute to increased sensitivity to gaze cues in patients in certain contexts.

Interestingly, in the study by Langdon et al. (2017), patients were also slower than controls to detect targets at short cue-target intervals (100 ms) when preceded by a direct-gaze cue. Given that direct-gaze cues are spatially neutral, this effect suggests that direct gaze (or perceived eye contact) may result in the early capture of attention in patients. This early sensitivity to direct gaze has been interpreted as reflecting a "threat-avoidance" processing style in patients, whereby patients may be biased towards perceiving eye contact as a signal of potential threat (Franck et al., 2002).

In other paradigms, we do see evidence of biases in the processing of eye gaze in schizophrenia. Most notably, it has been reported that patients are more likely than healthy adults to misperceive averted gaze as being directed towards themselves (Hooker & Park, 2005; Rosse, Kendrick, Wyatt, Isaac, & Deutsch, 1994; Tso, Mui, Taylor, & Deldin, 2012). Hooker and Park (2005) briefly presented participants with photographed faces (30 ms) followed by a scrambled face mask (75 ms). On each trial, participants were asked "*Are the eyes looking at you?*" Patients were more likely than controls to incorrectly indicate that the face was gazing at them when the gaze direction was averted. The authors claimed that this self-referential bias could not be explained by a fundamental deficit in low-level visual perception, given that patients' performance was commensurate with healthy controls in a non-social spatial discrimination task. Critically, the bias persisted in another control condition where eyes were edited out of the face stimuli. That is, patients continued to report the perception of direct gaze when eye gaze was unequivocally absent. These findings suggest that the bias is likely driven by later, conscious processes, rather than the early low-level visual processes that support gaze perception. In line with this interpretation, studies that do not involve a self-referential judgement, and simply instruct participants to report whether the eyes are averted to the left or right, have been unable to reliably identify a direct-gaze bias in patients (Franck et al., 1998; Franck et al., 2002; Seymour, Rhodes, McGuire, Williams, & Langdon, 2017). Summing up these findings, the early perceptual processing of gaze appears to be intact in schizophrenia, while later processes associated with evaluating its social significance may be disrupted (see, e.g., Langdon et al., 2006).

This is consistent with a growing although separate body of research which has reliably identified ToM impairments in patients (Brüne, 2005; Langdon, Still, Connors, Ward, & Catts, 2014; Sprong, Schothorst, Vose, Hox, & van Engeland, 2007). It is also consistent with other work which tentatively suggests that patients may be "hyper-primed" to detect or "over-perceive" intentionality in the behaviours of others (Blakemore, Sarfati, Bazin, & Decety, 2003; Harggard, Martin, Taylor-Clarke, Jeannerod, & Franke, 2003).

As argued earlier, responsive joint attention (RJA) requires the synthesis of accurate (unbiased) gaze perception and inferences of others' intentionality. Thus, studying RJA in schizophrenia will allow us to draw together the previous separate lines of research on ToM and gaze processing to examine how patients detect and interpret dynamic gaze cues during social interactions to infer the communicative intent of others, to respond contingently in a socially appropriate way. Yet, to date, no studies have examined RJA in schizophrenia.

This study aims to address this knowledge gap using a virtual interaction paradigm which critically requires the online synthesis of gaze perception and intention evaluation (Caruana, Brock, & Woolgar, 2015). Participants played a cooperative "Search" task with an avatar whom they believed to be controlled by a real person (named "Alan"). In reality, the avatar was pre-programmed to contingently respond to the participant's gaze (cf. Wilms et al., 2010). The participant and Alan worked collaboratively to locate a burglar who was hiding in one of six houses (see Figure 1), by each searching a row of houses. On some of these trials, the participant found the burglar during their search, and their job was to make eye contact with Alan and guide him by looking back at the burglar's location. Of critical interest to this study are the other trials when the participant did not find the burglar, so waited for Alan to finish searching his houses, after which Alan would establish eye contact to signal his communicative intent, and then guide the participant towards the house containing the burglar. The search aspect of our joint attention task is critical, as it creates a naturalistic context in which participants must differentiate between gaze shifts that are communicative from those that are not intentionally social. Although participants were required to initiate joint attention bids on trials where they did find the burglar, behaviour on these trials was not of interest in the current study.

We have previously used this paradigm to investigate joint attention in autistic adults (Caruana, Stieglitz Ham, et al., 2017). We found that autistic adults made significantly more responding errors (failing to respond or responding by fixating the wrong location) than control participants. However, there was no group effect on accuracy in a control condition in which participants responded to arrow cues, suggesting a specifically social impairment in evaluating intentional gaze cues.

Figure 1. Stimulus used in the interactive joint attention task, including the central avatar ("Alan") and the six houses in which the burglar could be hiding. Gaze areas of interest (GAOIs) are represented by blue rectangles. These were not visible to participants.

On correct trials, participants were significantly slower to respond to the avatar's eye gaze compared with arrow cues (see also Caruana et al., 2015; Caruana, McArthur, Woolgar, & Brock, 2017a; Caruana, Spirou, & Brock, 2017; also see Caruana, McArthur, Woolgar, & Brock, 2017b, for review). Notably, this saccadic reaction time (SRT) effect was exaggerated in autistic adults who were significantly slower than controls when they responded to gaze cues, but not arrows.

Subsequently, we have demonstrated that this SRT effect is sensitive to the eye contact processing and intention monitoring demands which precede RJA behaviour. We recently compared SRTs when typically developed adults completed the task described above (i.e., Search task) and a simplified version of the task which removed the initial search phase (and so participants knew that Alan would be signalling the burglar's location: NoSearch task; Caruana et al., 2017a). Thus, on NoSearch RJA trials, Alan only made eye contact and then one gaze shift to unambiguously initiate joint attention. Hence, evaluating eye contact was not needed to determine Alan's communicative intent, and participants were consequently faster to respond to gaze shifts on the NoSearch compared with the Search task. Importantly, the task manipulation had no impact on SRTs on non-social trials—indicating a specifically social effect of monitoring another's intention to use gaze as a directional signal.

Current study

The current study conducted the first experimental investigation of RJA in schizophrenia. We employed the same interactive task and Search versus NoSearch manipulation described above (cf. Caruana et al., 2017a) to investigate whether patients with schizophrenia experience any difficulty, compared with controls, responding to joint attention bids during gaze-based interactions, particularly on the Search trials when there is a need to disambiguate gaze shifts that do and do not signal communicative intent. We also assessed classic offline ToM abilities in participants.

If patients with schizophrenia—like autistic adults—experience difficulty in responding to joint attention bids, we would expect to see a group (Patient vs. Control) × condition (Social vs. Control) effect for both SRTs and accuracy. If similar to the pattern seen in autism, this would be characterised as slower SRTs and more errors on RJA trials, but no differences between groups on non-social control (RJAc) trials.

Furthermore, if any group differences in RJA, if found, are specifically related to how patients use eye contact as an ostensive cue of communicative intent, then we would expect the condition (Social vs. Control) × task (Search vs. NoSearch) interaction previously observed in healthy adults to be modulated by group. Specifically, if patients require more time to determine whether gaze shifts signal communicative intent (consistent with previous evidence of ToM impairment in schizophrenia), we might expect to see greater differentiation between groups in SRTs on the Search versus NoSearch social (i.e., RJA) trials. Contrastingly, a reduction in this effect (i.e., smaller group differences in SRTs across Search and NoSearch RJA trials) would indicate an enhanced sensitivity to detect communicative intent signalled by gaze in patients, consistent with other evidence suggesting that patients may be hyper-primed to perceive intentionality.

Method

Ethical statement

The study was approved by the Human Research Ethics Committee at Macquarie University (MQ; reference no. 5201200021). Participants received payment for their time and provided written consent before participating.

Participants

Patients with schizophrenia. A total of 21 clinically stable outpatients (5 females, M_{age}=48.29, *SD*=10.31) were recruited from the Australian Schizophrenia Research Bank (Loughland et al., 2010) and the Macquarie Belief Formation Volunteer Register. Diagnosis of schizophrenia was confirmed using the Diagnostic Interview for Psychosis (DIP; Castle et al., 2006). Symptom severity was also assessed using the Scales for the Assessment of Positive and Negative Symptoms (SAPS and SANS; Andreasen, 1983, 1984). Clinical symptom severity scores obtained using the SAPS and SANS are summarised in Table 1 for the patient group. Age of diagnosis ranged from 15 to 55 (*M*=26.57, *SD*=9.58) years, and duration of illness ranged from 6.75 to 42.83 (*M*=20.12, *SD*=10.43) years. In total, 19 patients were receiving a stable dose of second-generation anti-psychotic medication at the time of testing, while 1 patient was receiving a stable dose of a typical neuroleptic and another had discontinued anti-psychotic medication and was only receiving anti-epileptic medication.

Table 1. Symptom ratings for patients on the SAPS and SANS.

	M	*SD*	Range
Negative symptoms (SANS)[a]			
Affective flattening or blunting	2.76	1.09	0-4
Alogia	0.90	1.61	0-5
Apathy	2.52	1.47	0-5
Anhedonia	3.19	1.08	0-5
Attention	1.14	1.49	0-4
Positive symptoms (SAPS)[a]			
Hallucinations	0.57	1.08	0-3
Delusions	1.43	1.25	0-4
Bizarre behaviour	1.00	1.10	0-3
Positive thought disorder	1.10	1.37	0-4

SAPS: Scale for the Assessment of Positive Symptoms; SANS: Scale for the Assessment of Negative Symptoms; SD: standard deviation.
[a]0 = *not present*, 1 = *questionable*, 2 = *mild*, 3 = *moderate*, 4 = *marked*, 5 = *severe*.

Healthy controls. A total of 21 age-, sex-, and IQ-matched healthy controls (5 females, M_{age}=48.94, *SD*=8.08) were recruited from the general community and through the CCD Adult Register (www.ccd.edu.au/services/registers/). Controls were screened using a structured interview based on the affective, psychotic, and substance abuse screening modules from the Structured Clinical Interview for Axis 1 Disorders previously outlined under DSM-IV (SCID-1; First, Spitzer, Gibbon, & Williams, 2002). Participants with chronic medical conditions, history of nervous system disease or head injury, general psychological problems, or self-reported drug or alcohol abuse were excluded from the study. Three control participants, not included in the 21 participants mentioned above, were excluded based on responses provided during this interview indicating chronic medical conditions (including repeated treatment for cancer) and histories of depression and anxiety. Control participants also completed the brief version of the Schizotypal Personality Questionnaire (SPQ-B; Raine & Benishay, 1995). The range of scores obtained (*M*=5.76, *SD*=3.73) was consistent with previous studies involving non-clinical community samples (e.g., Compton, Chien, & Bollini, 2007).

There were no significant differences between patients (*M*=109.10, *SD*=9.30) and controls (*M*=110.71, *SD*=10.05) on estimated premorbid full-scale IQ measured using the National Adult Reading Test (NART; Nelson & Willison, 1991). All participants in both groups had vision that was either normal or corrected-to-normal.

Offline ToM (non-verbal and verbal) assessments

Participants also completed two offline tasks of ToM processing to evaluate whether the patients in our sample demonstrated difficulties in making inferences about the beliefs and intentions of others in non-interactive contexts.

Picture-sequencing task (non-verbal). Participants were required to arrange four cartoon cards in the correct order to depict a logically sequenced story (18 stories in total, including 2 practice stories; see Langdon, Ward & Coltheart, 2008, for a detailed description). There were four sequences for each of four story types: (1) false-belief stories testing ToM processing, (2) social script stories testing logical social script reasoning, (3) mechanical stories testing cause-and-effect reasoning, and (4) capture stories testing inhibitory control. A maximum score of 6 was awarded for each story depending on the number of correctly sequenced cards. Scores were averaged across each story type.

False-belief/deception story comprehension task (verbal). Participants read four short stories (see Langdon, Connors & Connaughton, 2017, for a detailed description). Two stories assessed first-order ToM (the ability to make inferences about others' thoughts) and two assessed second-order ToM (the ability to make inferences about how others think about others). Each story involved a character who was either deceived or had some false belief. Participants read each story and answered three questions. The first two questions assessed ToM ability and required participants to predict the character's actions by inferring their mental state (scored "1" point if correct, else "0") and to justify their answer ("2" if explicitly identified the character's false belief, intentions, or incomplete knowledge; "1" if the response implied but did not explicitly identify the other person's mental state [e.g., false belief], else "0"). Participants were also asked a control comprehension question that did not require ToM processing (scored "1" point if correct, else "0"). Scores were summed across each pair of first- and second-order stories. Participants could achieve a maximum score of "6" on ToM questions and "2" on control questions in each set.

Joint attention task

Social conditions. Participants played a cooperative game with an on-screen avatar believed to be controlled by another person named "Alan." Participants believed Alan was interacting with them from the neighbouring eye-tracking laboratory using live infrared eye-tracking; however, the avatar's gaze was controlled by a gaze-contingent algorithm (see Caruana et al., 2015, for a detailed description of this algorithm and a video demonstration of the task). During the game, participants worked with their partner to catch a burglar who was hiding inside one of the six houses presented on the screen. The participant found the burglar on half the trials, while Alan found the burglar on the other half—the critical RJA trials. This created the necessary collaborative search context. Participants completed a "Search" and "NoSearch" version of the task during two separate blocks.

Search task. Search trials began with a "search phase" where participants were required to search through the row of houses with blue doors, while Alan searched the houses with brown doors. One row was located in the upper portion of the screen, and the other in the lower portion of the screen. Whether the participant searched an upper or lower row of houses was counterbalanced across subjects. Participants were instructed that the contents of each house would only be revealed to the person allocated to search that row of houses. Whoever found the burglar was required to guide the other person to the burglar's location by initiating joint attention, and the other person was required to respond appropriately. Once joint attention was achieved, the burglar would be captured, and participants received feedback with the burglar appearing behind prison bars at the correct location.

Participants could search their allotted houses in any order they chose. Once fixated, the door would open to reveal the burglar, or an empty house. At the same time, Alan's gaze would shift to search his own allotted houses in a randomised order. Importantly, we varied whether 0-2 of the participant's houses were already opened and empty at the beginning of the trial. This meant that the search pattern engaged by participants was varied across trials, making Alan's random search behaviour appear realistically unpredictable.

On the RJA trials, participants discovered that all of their allotted houses were empty (Figure 2, row 1). Once the participant fixated back on the avatar's face, Alan was programmed to search 0-2 more houses before making eye contact. Alan then initiated joint attention by gazing towards one of his allotted houses. Participants were required to respond by fixating that house to capture the burglar. On these search trials, establishing eye contact with Alan was critical as it enabled participants to differentiate between the preceding gaze shifts which reflected the completion of Alan's search, and the gaze shifts following eye contact which signalled an intentional joint attention bid.

NoSearch task. On these trials, it was immediately obvious to participants at the beginning of each trial whether the

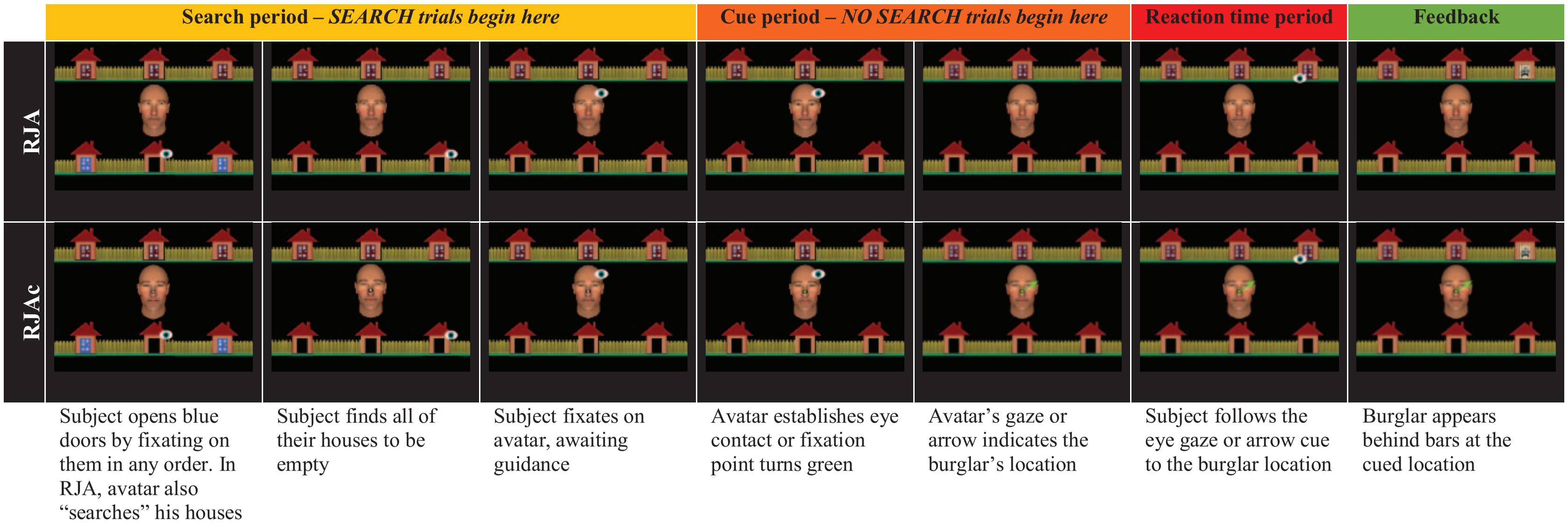

Figure 2. Event timeline for RJA and RJAc trials.

burglar was hiding in one of their allotted houses. This is because the participant's allotted houses were either all open and empty (RJA trials) or there would be a single closed blue door. Participants were told that a single blue door always contained the burglar. At the beginning of trials, Alan's eyes were closed and then opened after 500-1,000 ms (jittered with a uniform distribution) to establish eye contact with the participant. If the participant fixated Alan's gaze to establish eye contact, Alan's gaze only shifted once thereafter to guide the participant to the correct location after a further 500-1,000 ms on RJA trials. Importantly, while the perceptual properties of the Search and NoSearch tasks were identical, Alan establishing eye contact before guiding participants on RJA trials was only useful in conveying the communicative intent of the gaze cue on the Search and not the NoSearch trials.

Control conditions (RJAc). For both versions of the task (Search and NoSearch), we implemented arrow control trials matched on non-social task demands (e.g., attentional, oculomotor, and inhibitory control). Participants were instructed that they would complete the same task on their own and that changes in the stimulus on the screen were controlled by a computer programme. The avatar's face remained on the screen during control trials, with his eyes closed. On Search trials, a grey fixation point was presented on the avatar's nose until the participant completed their search and fixated upon it. The fixation point turned green (analogous to the avatar making eye contact). The Search and NoSearch tasks were otherwise identical. On RJAc trials, a green arrow, subtending three degrees of visual angle, cued the burglar's location (analogous to the avatar's averted gaze cue; see Caruana et al., 2015, for a video with example trials).

Procedure

Joint attention task. The experiment was presented using Experiment Builder 1.10.165 (SR Research, 2004). Participants completed two blocks of trials (one for the Search task, and the other for the NoSearch task), each comprising 108 trials, of which half required RJA from participants. Block order was counterbalanced across participants and matched between groups. Half of the participants per group were required to monitor the upper row of houses and the other half the lower row of houses in both blocks.

Each block comprised 27 trials from each condition (i.e., RJA and RJAc). Trial order randomisation was constrained to ensure that the location of the burglar, the location of blue doors, and the number of gaze shifts made by the avatar were matched within each block and condition (cf. Caruana et al., 2015). Order of blocks was predetermined and counterbalanced across participants and matched between groups. Alternating clusters of six trials were presented throughout each block. Each cluster began with a 1,000-ms cue that was presented over the avatar stimulus which read "Together" for RJA trials or "Alone" for RJAc trials.

Post-experimental interview. After the experiment, participants completed a post-experimental interview where they were asked to rate their subjective experience during the task on a number of dimensions (described below).

Eye-tracking. A desktop-mounted EyeLink 1000 Remote Eye-Tracking System (SR Research Ltd, Ontario, Canada) was used to record the right eye with a sampling rate of 500 Hz. A chinrest was used to stabilise head movements and standardise viewing distance. A 9-point sequence eye-tracking calibration was conducted at the beginning of each block. Seven gaze areas of interest (GAOIs) over the houses and avatar stimulus were used by our gaze-contingent algorithm and for subsequent analyses (see Caruana et al., 2015, for details).

Dependent variables

SRTs. We measured the latency (in ms) between the presentation of the gaze (RJA) or arrow cue (RJAc) and the onset of the participant's responding saccade towards the correct burglar location (see Figure 2, SRT period). Thus, this is a measure of saccade programming, rather than saccade execution. Trials with incorrect responses or saccadic RTs <150 ms were excluded as these were likely to be anticipatory saccades, rather than deliberate responses to the gaze or arrow cue (Carpenter, 1988). Trials naturally timed-out if the participant failed to respond within 3,000 ms. These were considered errors and were thus removed from saccadic RT analyses.

Accuracy. Participants could make three types of errors, (1) *Location errors* (fixating the wrong house when responding to a joint attention bid), (2) *Time-out errors* (taking longer than 3 s to achieve joint attention after establishing eye contact), or (3) *Search errors* (spending more than 3 s looking away from task-relevant stimuli—i.e., away from Alan or the houses). On location and time-out error trials, the burglar appeared in red at his true location, to indicate that he had escaped. On search error trials, red text reading "Failed Search" appeared on the screen to provide feedback. Accuracy was calculated as the proportion of trials where the participant succeeded in catching the burglar, excluding trials that required a recalibration or resulted in a Search error.

Subjective ratings. During the post-experimental interview, participants from both groups rated how difficult, natural, intuitive, and pleasant they found each task and condition on a 10-point Likert-type scale (1 = *Not at all*, 10 = *Extremely*). Participants also rated how cooperative their partner was and how "human-like" the avatar felt generally, as well as how human-like he appeared and behaved, using the same scale. Participants then indicated whether they preferred (1) completing the task alone (non-social trials) or together with their partner (social trials), (2) completing the NoSearch task or the Search task, and (3) interacting with a stranger through a virtual interface or face-to-face. Then, participants rated the strength of each preference on a 10-point scale (1=*completely prefer together/Search/face-to-face*, 10=*completely prefer alone/NoSearch/virtual*). Finally, participants were debriefed and rated how convinced they were that Alan was a real person on a 10-point scale (1=*Not at all*, 10=*Extremely*). Interviews were recorded using an audio recorder. Notable comments were recorded as close to verbatim during the interview. Any transcriptions noted in this manuscript were then checked against the audio recordings for accuracy.

Statistical analyses

Group differences in our offline measures of ToM were tested with mixed-design analyses of variance (ANOVAs; described for each test below). We used DataViewer software (SR Research Ltd) to export interest area and trial reports. All subsequent analyses were performed in R using a custom script to screen data and conduct statistical analyses.

Joint attention data were analysed in R. All our raw data, R code (with annotated descriptions), and analysis output can be downloaded from the Open Science Framework (osf.io/fhmyb). Accuracy and SRT data were subjected to logistic (accuracy) and linear (SRT) mixed random-effects analyses, using the lme4 R package (Bates, 2005) to test for interacting effects between group, task, and condition. Mixed random-effects analyses were implemented using the maximum likelihood estimation method. Mixed random-effects models were used as they are robust to missing data and can account for both subject and item-level variance (i.e., random effects) when estimating fixed effects and interactions. This is unlike a traditional ANOVA which compares mean data, aggregated across trials for each subject (Quene & van den Bergh, 2004, 2008). In the current study, all three variables of interest (i.e., group, task, and condition) were treated as binary fixed factors, coded as ±0.5.

In line with recommendations for implementing mixed random-effects models, we adopted "maximal" random factor structures, with random intercepts for trial, and by-subject random slopes for the fixed effects and interactions (cf. Barr, Levy, Scheepers, & Tily, 2013). Specifically, we originally defined a saturated model including by-subject random slopes for the group × condition × task effect. Unsurprisingly, the "maximum likelihood" of this highly complex model could not be estimated given the available data and "failed to converge" (see Barr et al., 2013, p. 10, for an explanation). Therefore, we simplified our random-effects parameters to define the most saturated and parsimonious model. The accuracy model was simplified to only include random intercepts for subject and trial. The SRT model was simplified to include by-subject random slopes for the effect of condition and task, and random intercepts for trial.

The values of p were calculated using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014). For SRT data, we used an inverse transform on trimmed data. First, we screened trials with SRTs faster than 150 ms as these were likely to be anticipations. Second, we trimmed standardised residuals greater than 2.5. The inverse transformation is recommended for ensuring that reaction time residuals meet the assumption of normality (see Balota, Aschenbrenner, & Yap, 2013) as skewed data biases the estimation of model parameters. Removing standardised residuals greater than 2.5 is also recommended when reaction time outliers contribute to non-normality (see Baayen, 2008; Baayen, Davidson, & Bates, 2008). We have included the output for SRT analyses on untrimmed and trimmed transformed data (including Q-Q normality plots; see osf.io/fhmyb). Note, the output reveals that (1) transforming the data alone was not sufficient for normalising the SRT data, however transforming *and* trimming residuals greater than 2.5 was, and (2) trimming residuals did not change the overall pattern of results.

To quantify the variance explained by each of our fixed effect factors and interactions of interest, we compared a series of mixed random-effects models which comprised the maximally defined random effects possible for our accuracy and SRT models and added one of our fixed effects at a time (see RMarkdown document at osf.io/fhmyb for a detailed description and output). Chi-square likelihood ratios were calculated for each comparison to determine the extent to which each fixed effect parameter improved model fit and whether that improvement was significant. These analyses are presented in lieu of traditional effect size statistics, which are unable to account for the variance explained by each fixed effect, over-and-above variance already explained by the defined random effects.

For subjective ratings, we used non-parametric Mann–Whitney's U-tests to investigate the effect of group for each rating. A significance criterion of $p < .05$ was used for all analyses, except for follow-up pairwise comparisons, where we implemented a Bonferroni correction for each family of follow-up tests. This applied to the offline ToM tasks, accuracy, and SRT analyses, each comprising four follow-up comparisons ($\alpha = .0125$).

Results

Offline ToM tasks

For both ToM tasks, Greenhouse–Geisser corrections are reported as the assumption of sphericity was violated.

Table 2. ToM performance by task and group.

	Patients		Healthy controls		*T(df)*	*p*-value	*d*
	M	*SD*	*M*	*SD*			
Picture sequencing							
False belief	4.45	1.38	5.00	0.91	1.41 (40)	.164	0.437
Social script	5.71	0.63	5.83	0.31	0.78 (40)	.441	0.240
Mechanical	5.43	0.84	5.79	0.43	1.74 (40)	.091	0.535
Capture	3.81	1.21	4.26	1.16	1.24 (40)	.223	0.382
Total score	77.76	11.91	83.52	7.84	1.85 (40)	.071	0.572
Story comprehension							
First-order ToM	4.52	1.89	5.62	0.80	2.45 (27.04)	.021[a]	0.755
First-order comprehension	1.86	0.36	1.90	0.30	0.47 (40)	.644	0.144
Second-order ToM	3.19	1.86	5.29	1.10	4.44 (32.49)	<.001	1.37
Second-order comprehension	1.90	0.30	1.95	0.22	0.59 (40)	.560	0.181

ToM: theory of mind; SD: standard deviation.

[a] This comparison is not significant after implementing a Bonferroni correction for multiple comparisons ($\alpha = .013$).

Picture-sequencing task. A 2 (group: patient/control) × 4 (story condition) mixed-design ANOVA revealed a significant main effect of story condition, $F(2.49, 99.57) = 43.84$, $p < .001$, $\eta_p^2 = .523$, with no suggestion of an effect of group, $F(1, 40) = 3.43$, $p = .071$, $\eta_p^2 = .079$, nor a story condition × group interaction, $F(2.49, 99.57) = 0.50$, $p = .647$, $\eta_p^2 = .012$ (see Table 2). Thus, patients showed general picture-sequencing difficulty and no evidence of a ToM-specific impairment on this task.

False-belief/deception story comprehension task. In contrast, patients did reveal a ToM-specific difficulty, compared with controls for both first- and second-order stories. This was demonstrated by a significant main effect of group, $F(1, 40) = 20.70$, $p < .005$, $\eta_p^2 = .341$, and story condition, $F(2.09, 83.73) = 102.16$, $p < .001$, $\eta_p^2 = .719$, as well as a group × condition interaction, $F(2.09, 83.73) = 9.38$, $p < .001$, $\eta_p^2 = .189$, in which patients performed significantly poorer that controls for second-order ToM questions, but not on control comprehension questions, or first-order ToM questions once correcting for multiple comparisons (see Table 2).

SRT

Full model. Figure 3 summarises the SRT data by group, task, and condition. Participants in both groups were slower to respond on RJA trials than RJAc trials (main effect of condition, $\beta = -0.69$, $SE = .05$, $t = 14.85$, $p < .001$). Patients were also significantly slower than controls overall (main effect of group, $\beta = -0.43$, $SE = .12$, $t = 3.68$, $p = .001$). We replicated the intention monitoring effect identified in our previous work (Caruana et al., 2017a) in which the condition effect on responding (i.e., slower to respond to arrows than social eye gaze) was greater during the Search task than the NoSearch task, as participants must also disambiguate which gaze shift signals an intentional joint attention bid, resulting in slower response times (condition × task interaction, $\beta = -0.22$, $SE = .04$, $t = 5.46$, $p < .001$). We conducted a follow-up analysis to confirm that this condition × task interaction was also significant when data were analysed separately for controls ($\beta = -0.29$, $SE = .06$, $t = 5.07$, $p < .001$) and patients ($\beta = -0.16$, $SE = .06$, $t = 2.89$, $p = .004$).

We also found a significant group × condition interaction ($\beta = 0.43$, $SE = .08$, $t = -5.24$, $p < .001$) in which patients exhibited a different condition effect compared with controls. However, we found no evidence for a significant main effect of task, group × task, or group × condition × task interaction (all $ps > .110$). To fully characterise the observed group × condition interaction effect, we tested four additional models, two models to test the condition effect for each group separately and two models to test the group difference in SRTs for the RJA and RJAc conditions separately. Task was not included as a factor in any of the follow-up models as there were no observed significant interactions between task and group. We found that the condition effect observed in our previous studies using this task was significant in both controls ($\beta = -0.86$, $SE = .05$, $t = 18.61$, $p < .001$) and patients ($\beta = -0.46$, $SE = .08$, $t = 5.93$, $p < .001$), with slower responses to gaze cues on RJA trials, compared with arrow cues on RJAc trials. Of most interest, we found that while there was no evidence for a significant group difference in response times on RJA trials ($\beta = -0.20$, $SE = .11$, $t = 1.74$, $p = .089$), patients were significantly slower than controls to respond to arrow cues on RJAc trials ($\beta = -0.61$, $SE = .13$, $t = 4.71$, $p < .001$). These data suggest that patients are slower than controls when required to orient attention to non-social arrow cues, despite demonstrating no difficulty, relative to controls, when orienting to social gaze cues.

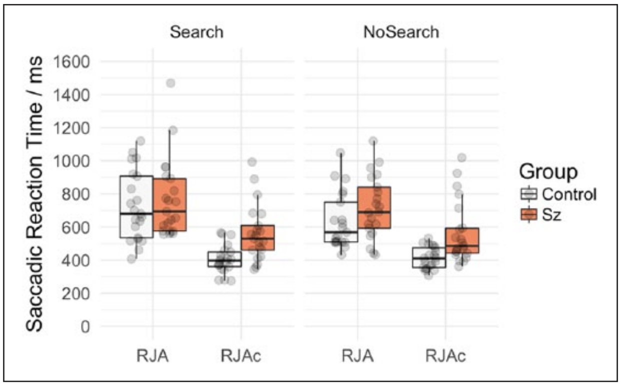

Figure 3. Boxplot with individual data points depicting saccadic reaction times on correct trials by group (i.e., Control, Sz), condition (i.e., RJA, RJAc), and task (i.e., Search, NoSearch). In all boxplot figures, whiskers extend (as in a conventional Tukey's boxplot) to the furthest data points that are within 1.5 times the length of the box, from the end of the box.

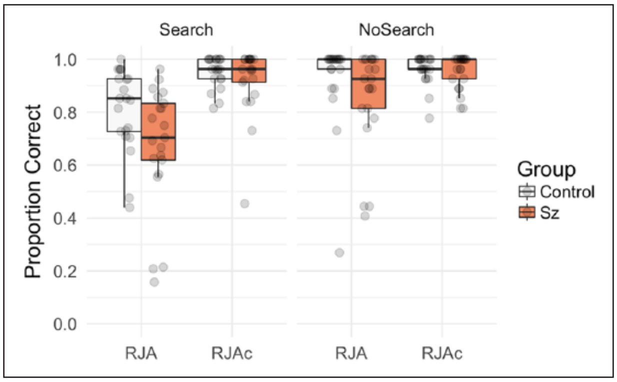

Figure 4. Boxplot with individual data points depicting the proportion of correct trials by group (i.e., Control, Sz), condition (i.e., RJA, RJAc), and task (i.e., Search, NoSearch).

Accuracy

Full model. Figure 4 summarises the accuracy data by group, task, and condition. Participants in both groups made more errors on RJA trials than RJAc trials (main effect of condition, $\beta=-1.49$, $SE=.14$, $z=-11.02$, $p<.001$) and on Search trials compared with NoSearch trials (main effect of task, $\beta=-0.87$, $SE=.12$, $z=-7.13$, $p<.001$). Specifically, errors were more frequently made on RJA trials during the Search task compared with the NoSearch task and were less frequent overall for RJAc trials than RJA trials (condition × task interaction, $\beta=-0.85$, $SE=.24$, $z=-3.52$, $p<.001$).

Importantly, we found that the condition effect above (i.e., more errors when responding to gaze cues than arrow cues) was significantly larger in patients than controls (group × condition interaction, $\beta=-0.63$, $SE=.25$, $z=-2.55$, $p=.011$) suggesting a specifically social impairment in accurately processing directional information from gaze cues to direct attentional shifts and share attention with others. However, we found no evidence for a significant main effect of group, group × task, or group × condition × task interaction (all $ps>.123$). In line with our SRT analyses, we tested four additional models to fully characterise the observed group × condition effect on accuracy, including two models to test the condition effect for each group separately and two models to test group differences for the RJA and RJAc conditions separately. Task was not included as a fixed effects factor in any of the follow-up models as there were no observed significant interactions between task and group. We found that the condition effect observed in our previous studies using this task was significant in both controls ($\beta=-1.23$, $SE=.18$, $z=-7.02$, $p<.001$) and patients ($\beta=-1.77$, $SE=.17$, $z=-10.24$, $p<.001$), with more errors made when required to respond to gaze cues (RJA trials) than arrow cues (RJAc trials). Patients made more errors than controls when required to respond to gaze cues ($\beta=-0.79$, $SE=.36$, $z=-2.21$, $p=.027$) but this did not reach statistical significance once correcting for multiple comparisons ($\alpha=.0125$). Furthermore, there were no significant differences between groups in accuracy when responding to arrow cues ($\beta=-0.27$, $SE=.18$, $z=-0.056$, $p=.575$). A visualised breakdown of the types of errors by group, task, and condition can be found in the RMarkdown document published on the Open Science Framework (osf.io/fhmyb).

Subjective ratings

Figure 5 provides a summary of the subjective task ratings. Almost all participants reported that they were completely convinced that their virtual partner was controlled by another person ($M=8.76$, $SD=2.29$). No group differences were significant for ratings of how natural or how pleasant participants found the tasks. Patients rated all conditions and versions of the task as more difficult and less intuitive than did controls (all $ps<.019$; see RMarkdown for full summary of test statistics, osf.io/fhmyb). Compared with controls, patients also rated Alan as being significantly less cooperative ($W=354.5$, $p<.001$) and indicated a stronger preference for interacting with a stranger—like Alan—through a virtual interface than engaging face-to-face ($W=119$, $p=.01$).

Discussion

In this study, we conducted the first experimental investigation of RJA in schizophrenia. We extended our previous work with autistic adults by implementing an interactive task developed to capture the social and non-social demands involved when intentionally responding to communicative joint attention bids (cf. Caruana et al., 2017a). Specifically, we aimed to investigate whether patients with schizophrenia experience any difficulty, compared

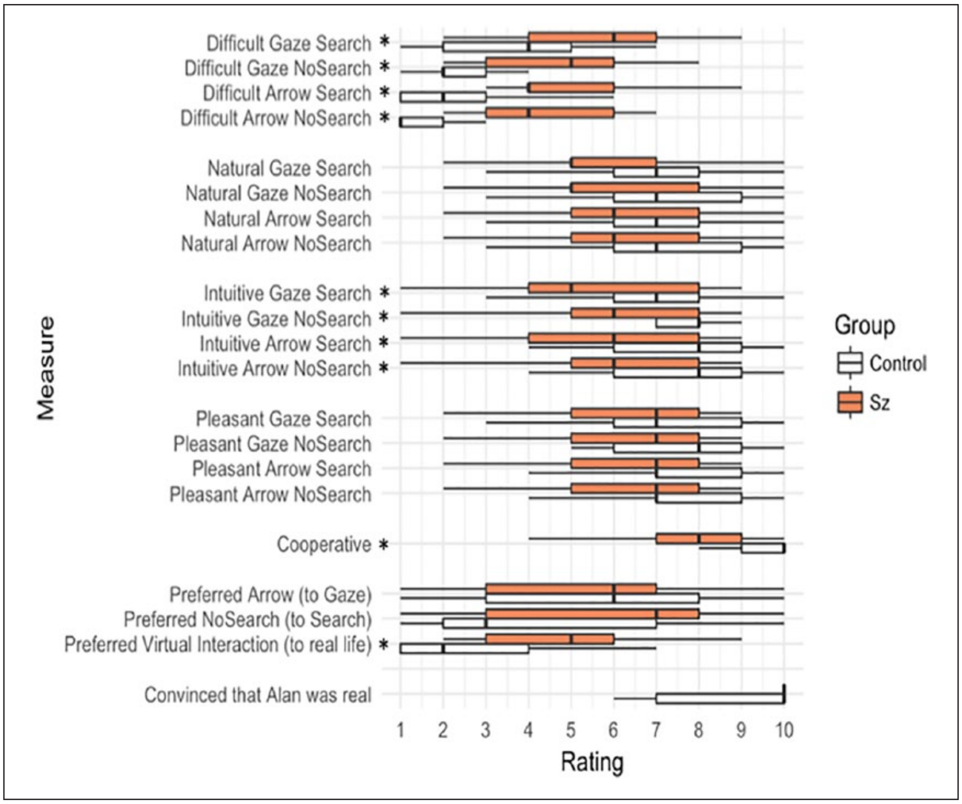

Figure 5. Tukey's boxplots depicting responses to subjective rating questions.

with controls, when responding to joint attention bids during gaze-based interactions. We particularly endeavoured to evaluate whether RJA difficulties could be explained by divergent processing of the communicative intent of others. Performance in classic offline ToM tasks was also assessed.

On the verbal offline tests of false belief and deception understanding, patients in our sample showed a *selective ToM impairment*. On our interactive joint attention task, accuracy results also showed that patients made selectively *more social errors* (i.e., failures to process and respond appropriately to gaze-direction information on social trials) compared with controls, irrespective of whether there was a need to monitor gaze for signals of communicative intent (i.e., across Search and NoSearch tasks). In contrast, the SRT results revealed what could be interpreted as a *social facilitation* effect for patients; on correct trials, patients were slower than controls to respond to non-social arrow cues (RJAc) but were no slower when responding to gaze cues (RJA). In other words, they were relatively faster than controls in responding to communicative gaze cues, once correctly identified. The SRT results also revealed that the task effect (i.e., Search vs. NoSearch tasks) was commensurate in patients and controls, indicating that the gaze advantage in responsive attentional shifts for patients was not additionally modulated by time taken to monitor communicative intent. We offer more detailed interpretations of these findings in the following subsections.

Search versus NoSearch task results

Responding to joint attention bids requires the effective perception of a social cue (e.g., eye gaze), as well as an appropriate evaluation of its social significance and intentionality (Senju & Johnson, 2009). While both the Search and NoSearch tasks require the former process, only the Search task requires participants to engage the later process. In this way, comparing response times on these tasks provides an index of the time taken to process the communicative intent of gaze cues. In line with our previous findings using the Search and NoSearch versions of this task, we found that participants in both groups were significantly slower to respond to Alan's gaze cues during the Search task than the NoSearch task (Caruana et al., 2017a). Interestingly, our data reveal that patients are no slower than controls in evaluating the communicative intent of others when responding to gaze cues in the Search task. In fact, a closer inspection of the SRT data across social and non-social versions of the task in both groups suggests that patients demonstrated a gaze advantage (discussed further below). The following subsections focus on the SRT and accuracy data showing significant group × condition (social vs. non-social) interactions, which were not modulated by task (Search vs. NoSearch).

SRTs

While the accuracy effects observed in this study mirror those of our previous work with autistic adults—suggesting a similar selective difficulty in evaluating gaze direction and accurately responding to gaze cues to achieve joint attention with others in both schizophrenia and autism (discussed further below)—we see the opposite pattern of results in our SRT data. Specifically, our previous work revealed that when autistic adults completed our Search task, they were initially slower than controls when responding to gaze cues, but not arrow cues. This suggested that autistic adults experienced a specific difficulty when shifting attention responsively to gaze cues that was independent of their ability to orient attention or control eye movements in non-social contexts.

In the current study, we find the opposite pattern of SRT results in which patients are significantly slower than controls when responding to arrow cues, but do not differ in their response times to gaze cues. The condition effect on response times was smaller for patients ($M_{RJA\text{-}RJAc} = 187.19$ ms, $SD = 161.01$) than controls ($M_{RJA\text{-}RJAc} = 261.63$ ms, $SD = 97.19$). That is, patients differed less in their response times to arrows and gaze, suggesting that they experience a relative advantage when responding to gaze-cued joint attention bids. This advantage may compensate for the non-social attention orienting and/or oculomotor control deficits experienced by patients. This interaction effect is unlikely to reflect a simple deficit of responding to arrows, given that the non-social task demands are equal in both conditions. In fact, the social condition is objectively more demanding, which is demonstrated by the condition effects observed in both groups, where participants made more errors and were slower to respond to gaze cues than arrow cues within the same task context.

One explanation for this relative social advantage is that patients have an increased sensitivity to process and respond to signals of intentionality (e.g., eye contact followed by an averted gaze shift to signal communicative intent). This interpretation aligns with findings of an increased tendency for patients to respond to non-biological motion as if it were biological and animate (e.g., Kim et al., 2011; see Billeke & Aboitiz, 2013, for review) and evidence that patients are faster than controls to reflexively shift attention in response to head-orientation cues at short stimulus onset asynchronies (SOAs; Langdon et al., 2006). This hyper-responsivity may have been particularly evident in the current study, where gaze was deliberatively evaluated in a truly social and interactive context and conveyed information that was of direct self-relevance to the participant. In other words, the self-referential hyper-responsivity in patients is likely the result of effects manifesting at later stages of deliberative cognitive processing. This aligns with evidence that patients have also demonstrated faster response times when detecting targets that are congruently cued by gaze at longer SOAs—where there is time for deliberative self-referential processes to influence gaze evaluation (Langdon et al., 2017)—and the interpretation that the direct-gaze bias effect observed in patients (when averted gaze is mistakenly judged direct) likely reflects later stages of cognitive processing and self-referential evaluation, and not fundamental impairments of visual perception (Franck et al.,1998; Franck et al., 2002; Seymour et al., 2017).

This account is generally consistent with accumulating evidence that patients over-attribute social meaning and intent to social (and non-social) cues—and this has been proposed as potentially contributing to commonly experienced persecutory and paranoid delusions (i.e., mistaken inferences of others' malicious intent; Abu-Akel & Bailey, 2000; Frith, 2004). Indeed, the over-attribution of intent in schizophrenia may be one aspect of a broader deficit in mental state attribution or ToM which may have a top-down influence on social perception and responsivity (Brüne, 2005; Sprong et al., 2007), consistent with our findings of poorer performance by patients on our verbal ToM task (indexing a failure to take appropriate account of others' beliefs when explaining their behaviour), alongside patients' relatively faster response times to gaze signals on our joint attention task. Unlike the offline ToM tasks, our joint attention task required participants to interpret social cues to infer communicative intentions during a second-person, rather than a third-person, perspective. As such, this hypersensitivity to perceive the intentions of others in schizophrenia may be specific to scenarios where cues are self-relevant. It is also important to recognise that while patients may be more likely to evaluate cues as intentional and communicative, such a hypersensitivity does not necessarily mean that this evaluation will be accurate.

Accuracy

Our accuracy analyses revealed that patients exhibited larger condition effects on accuracy ($M_{RJA\text{-}RJAc} = 0.17$, $SD = 0.15$)—characterised by more errors when responding to gaze cues than arrow cues—compared with controls ($M_{RJA\text{-}RJAc} = 0.08$, $SD = 0.11$). This group × condition interaction effect resembles our previous findings in autistic adults (Caruana, Stieglitz Ham, et al., 2017). These data suggest that patients with schizophrenia—like autistic adults—may find it difficult to accurately identify and respond to the correct location cued by gaze-direction signals of communicative intent. Several patients explicitly reflected on the challenges of understanding and responding to the direction signalled by Alan's gaze. For example, one patient explained:

> Working with Alan was probably the hardest part, direction—wise picking up where I was looking with his eyes was very difficult—but working with the arrows was very easy—I was able to relax. But working with Alan was very hard. When it was green arrows it was dead easy.

Despite the similarities in the subjective experiences of autistic adults and patients with schizophrenia, the reason for their divergent sensitivity to gaze cues is likely to be different. In our previous work, we hypothesised that autistic individuals may be *hypo*sensitive to gaze-based signals of communicative intent (Caruana, Stieglitz Ham, et al., 2017; also see Böckler et al., 2014; Senju & Johnson, 2009). However, in patients, who were not differentially delayed when monitoring Alan's communicative intent on Search compared with NoSearch tasks, there may be a self-referential *hyper*sensitivity to gaze signals of communicative intent (as described above). While this could underlie the patients' faster response times to communicative gaze signals when they are correctly identified, it may also lead to more errors. For instance, it is possible that patients may have been more likely to incorrectly interpret alan's *searching* gaze shifts as intentional joint attention bids, resulting in pre-emptive orienting to the incorrect location, and thus a location error. In line with this interpretation, one patient explained:

> It was difficult interacting with Alan, because his eyes when I was searching kept looking at the three different places—and sometimes he had to look twice at the same spot—and then I thought that's what he was gesturing. When I was alone I had to wait for the green dot and that was a lot clearer to me.

Another patient said: "*Whenever I was working with him it took a lot of effort . . . figuring out whether he was searching or actually looking.*"

However, a breakdown of the types of errors patients made on social trials (available in the RMarkdown, osf.io/fhmyb) reveals a mixture of both location errors (i.e., responding by looking at the incorrect location) and time-out errors (i.e., failing to respond). Therefore, it is possible that there may be multiple ways in which patients are more confused on social trials, thus leading to more errors. A second possibility is that the self-referential nature of the task triggers the direct-gaze bias effect. In other words, patients may misperceive sustained direct eye contact and so miss Alan's gaze shift to initiate joint attention—which ultimately results in their failure to respond, and so a time-out error. The likelihood of confusing direct and averted gaze in this task may also be exacerbated by the use of gaze-shift cues in a static direct face, which occur in isolation of head-orientation cues in our task (cf. Langdon et al., 2006).

Some of the subjective comments made by patients are consistent with the interpretation that they often found it difficult to identify relevant gaze shifts. One explained "*Quite often Alan would find it, and I'd have to look back at him, but he'd give me an indication, but I wouldn't pick up on it.*" Another said: "*I found it hard to tell where he was looking.*" And the third said: "*Sometimes I just didn't follow them. Not used to eyes I suppose. I often look away.*" One avenue for future work would be to examine whether the accuracy differences between groups are abolished when joint attention is initiated using head-turn cues which are less susceptible to the direct-gaze bias and perceptual confusion.

Finally, it is possible that the perceived presence of another real person controlling the avatar (Alan)—independent of task demands—may have compromised accuracy on social trials by establishing a context which may have triggered social anxiety or stress in some participants. This was supported by comments made by two patients in our study. One patient explained that they associate human interactions with negative experiences: "*I struggle with human interaction. Always have my whole life . . . Just bad experiences in the past.*" Another provided an insightful account of the discomfort they often experience during face-to-face interactions:

> I find social interactions overwhelming because I can't pick up on micro—gestures. I can talk to someone for hours over the phone but after about an hour of face—to—face—I'm "full—up"—and will have to start using strategies to help deal with it, like tell them that "Hey, I'm full" and then go away and sit with my eyes closed or have some time-out. . . . it becomes really overwhelming.

The third patient explained how social gatherings can be particularly stressful when there is a need to track gaze information from multiple people:

> Normally I get overwhelmed with eyes. I was pretty good at Christmas. It didn't happen to me this year. When I used to smoke my mum would tell me to go out for a smoke and it was an excuse to get away. I could focus on something else. Calm down.

Future work is needed to investigate whether this social discomfort when interacting with others in real life can affect behavioural measures of social responsivity on our task—independent of task demands. To clarify, this can be achieved by comparing performance on the task when patients believe their virtual partner is being controlled by another person or a computer. We have demonstrated in previous work using the same task that response time measures (but not accuracy) differ depending on whether neurotypical participants believe the avatar is controlled by another human (i.e., Alan) or a computer programme (cf. Caruana, Spirou, & Brock, 2017). Similar manipulations of human agency beliefs in patients could be implemented to investigate whether differences in performance (of most interest, poorer accuracy) are particularly observed when patients believe they are interacting with an intentional agent capable of negatively evaluating them.

Subjective ratings

Performance and partner cooperativeness. Patients rated all conditions and versions of the task as being more difficult and less intuitive than controls—indicating that it required ongoing, deliberate thought—to maintain performance on the task.

This highlights that—in addition to any potential social-cognitive difficulties or advantages—patients were likely attempting to overcome cognitive challenges during the completion of the task that may have not been restricted to the social domain, but possibly resulting in the patients' poorer accuracy, which was most marked for the generally more difficult social trials. These challenges may have involved task switching (between the more difficult social and less difficult non-social conditions), attention orienting (to both gaze and arrow cues), action inhibition, and oculomotor control.

Interestingly, patients also rated Alan as being significantly less cooperative than controls. One explanation for this is that they may have attributed the negative feedback on incorrect trials as being due to Alan's performance rather than their own. However, both patients and controls generally rated Alan as being more cooperative than not.

Preference for virtual reality. When asked to elaborate or explain their preference for interacting with strangers through a virtual interface—rather than face-to-face—many patients commented on their persistent difficulty in navigating social interactions. In many cases, patients specifically reflected on the challenges they experience when confronted with eye gaze (as highlighted above). However, many patients also reported that the virtual interaction with Alan was "*less confronting*" and "*not as personal*" and several patients agreed that, consequently, this made the task "*easier than real life*" and more engaging, despite still finding the task challenging.

Recently, there has been growing recognition that social cognition and interaction skills should be the focus of targeted cognitive remediation programmes for patients with schizophrenia (e.g., Marsh et al., 2012; Marsh, Langdon, McGuire, Polito, & Coltheart, 2013). The current study suggests that virtual reality may provide a useful tool for implementing these interventions—at least in the first instance—as they provide a less intimidating and safer environment for social cognition and information processing training to be delivered without compromising ecological validity. Supporting this idea, one patient explained:

> I'd prefer virtual reality if it's a stranger. Don't have to deal with them. Don't have to deal with awkward moments. Easier. Less Stress.

Another described how virtual reality made eye contact less confronting:

> I think [my wife] tries to give eye contact to me to give me directions and I often go half-cocked rather than look at her. I think looking at the avatar was less-confronting than me looking at you.

Finally, the third patient explained how they wish there was a tool that would enable virtual interactions with others in their daily lives:

> I just know myself. I don't mix very well. I can't really say why. I dunno. If I had things like that [i.e., access to a virtual interaction interface] I'd probably use them all the time. Because I don't socialise that much. That's all I can say. But if I had a computer, or whatever, then I could do better at that, than talking to a human being . . . I was quite happy [using the virtual interface].

Conclusion

This was the first study to investigate gaze-based joint attention in patients with schizophrenia. Clinical diagnostic assessments revealed that the symptoms and social difficulties experienced by participants in our psychiatric sample were representative of those typically observed in patients with schizophrenia. Our behavioural data also reflected the daily social difficulties many patients experience in their interactions with others, particularly when required to deliberately infer and evaluate intentions. This was reflected in the subjective accounts of our participants and their performance on offline measures of ToM.

Patients in the current study experienced generalised difficulties in our interactive task, with poorer overall accuracy and generally slower response times compared with controls. Patients also demonstrated a selectively social difficulty in accurately evaluating and responding to gaze cues in a fast-paced and dynamic social interaction which required the disambiguation of multiple communicative and non-communicative gaze cues. However, when patients were able to accurately disregard non-communicative gaze shifts, they were relatively faster than controls to respond and achieve joint attention. We interpret these findings as reflecting a hypersensitivity to signals of communicative intent and a possible self-referential bias in schizophrenia. Our findings also demonstrate the importance of considering task context (e.g., online vs. offline; self-referential vs. third-person observation; and cooperative vs. non-cooperative interactions) when evaluating whether a cognitive difference observed in psychiatric populations should be categorised as an "impairment" and highlight the potential utility of virtual interaction paradigms for future social-cognitive interventions.

Acknowledgements

All authors contributed to the study's design, analysis plan, and revision of the manuscript. N.C. independently collected the data, carried out all analyses, and wrote the initial draft of the manuscript.

Declaration of conflicting interests

This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was supported by the Australian Research Council (ARC) and the ARC Centre of Excellence in Cognition and its Disorders (CCD, www.ccd.edu.au; CE110001021). Drs Caruana and Seymour were recipients of two ARC CCD Cross-Programme Support Scheme grants which directly supported this work. Dr Caruana also received financial support from a Macquarie University Research Fellowship (71990140). Dr Seymour also received financial support from the Society for Mental Health Research, Australia. The funding sources for this project were not involved in any decision-making, data collection, analysis, or dissemination of this study. The scientific process of this study was independent of any relationship with the funding sources that could be construed as a potential conflict of interest.

Open practices

The data and materials from the present experiment are publicly available at the Open Science Framework website (osf.io/fhmyb).

ORCID iD

Nathan Caruana https://orcid.org/0000-0002-9676-814X

References

Abu-Akel, A., & Bailey, A. L. (2000). The possibility of different forms of theory of mind impairment in psychiatric and developmental disorders. *Psychological Medicine*, *30*, 735–738.

Akiyama, T., Kato, M., Muramatsu, T., Maeda, T., Hara, T., & Kashima, H. (2008). Gaze-triggered orienting is reduced in chronic schizophrenia. *Psychiatry Research*, *158*, 287–296. doi:10.1016/j.psychres.2006.12.004

Andreasen, N. C. (1983). *The Scale for the Assessment of Negative Symptoms* (SANS). Iowa City, IA: University of Iowa.

Andreasen, N. C. (1984). *Scale for the Assessment of Positive Symptoms* (SAPS). Iowa City, IA: University of Iowa.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology*, *39*, 1563–1571. doi:10.1037/a0032186

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge: The MIT Press.

Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 1–41. doi:10.1016/j.jml.2012.11.001

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, *5*, 27–30.

Billeke, P., & Aboitiz, F. (2013). Social cognition in schizophrenia: From social stimuli processing to social engagement. *Frontiers in Psychiatry, 4*, Article 4.

Blakemore, S. J., Sarfati, Y., Bazin, N., & Decety, J. (2003). The detection of intentional contingencies in simple animations in patients with delusions of persecution. *Psychological Medicine*, *33*, 1433–1441.

Böckler, A., Timmermans, B., Sebanz, N., Vogeley, K., & Schilbach, L. (2014). Effects of observing eye contact on gaze following in high-functioning autism. *Journal of Autism and Developmental Disorders*, *44*, 1651–1658.

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*, 169–175. doi:10.1002/mrdd.20036

Brüne, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophrenia Bulletin*, *31*, 21–42.

Carpenter, R. H. S. (1988). *Movements of the eyes* (2nd ed.). London, England: Pion.

Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *NeuroImage*, *108*, 34–46. doi:10.1016/j.neuroimage.2014.12.041

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017a). Detecting communicative intent in a computerised test of joint attention. *PeerJ*, *5*, e2899. doi:10.7717/peerj.2899

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017b). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews*, *74*, 115–125.

Caruana, N., Spirou, D., & Brock, J. (2017). Human agency beliefs influence behaviour during virtual social interactions. *PeerJ*, *5*, e3819. doi:10.7717/peerj.3819

Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Palermo, R., Kloth, N., & McArthur, G. (2017). Joint attention difficulties in adults with Autism. *Autism*, *22*, 502–512. doi:10.1177/1362361316676204

Castle, D. J., Jablensky, A. J., McGrath, J. J., Carr, V. C., Morg, V., Waterreu, A., & Farmers, A. (2006). The diagnostic interview for psychoses (DIP): development, reliability and applications. *Psychological Medicine*, *36*, 69–80. doi:10.1017/S0033291705005969

Compton, M. T., Chien, V. H., & Bollini, A. M. (2007). Psychometric properties of the Brief Version of the Schizotypal Personality Questionnaire in relatives of patients with schizophrenia-spectrum disorders and non-psychiatric controls. *Schizophrenia Research*, *91*, 122–131.

Couture, S. M., Penn, D. L., & Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: A review. *Schizophrenia Bulletin, 32*(Suppl. 1), S44–S63.

Dalmaso, M., Galfano, G., Tarqui, L., Forti, B., & Castelli, L. (2013). Is social attention impaired in schizophrenia? Gaze, but not pointing gestures, is associated with spatial attention deficits. *Neuropsychology*, *27*, 608–613.

First, M. B., Spitzer, R. L, Gibbon, M., & Williams, J. B.W. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. New York: Biometrics Research, New York State Psychiatric Institute, November 2002.

Fiszdon, J. M., Fanning, J. R., Johannesen, J. K., & Bell, M. D. (2013). Social cognitive deficits in schizophrenia and their relationship to clinical and functional status. *Psychiatry Research*, *205*, 25–29.

Franck, N., Daprati, E., Michel, F., Saoud, M., Dalery, J., Marie-Cardine, M., & Georgieff, N. (1998). Gaze discrimination is unimpaired in schizophrenia. *Psychiatry Research*, *81*, 67–75.

Franck, N., Montoute, T., Labruyere, N., Tiberghien, G., Marie-Cardine, M., Dalery, J., . . . Georgieff, N. (2002). Gaze direction determination in schizophrenia. *Schizophrenia Research*, *56*, 225–234.

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

Frith, C. D. (2004). Schizophrenia and theory of mind. *Psychological Medicine*, *34*, 385–389.

Harggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M., & Franke, N. (2003). Awareness of action in schizophrenia. *NeuroReport*, *14*, 1081–1085.

Hooker, C., & Park, S. (2005). You must be looking at me: The nature of gaze perception in schizophrenia patients. *Cognitive Neuropsychiatry*, *10*, 327–345.

Kim, J., Park, S., & Blake, R. (2011). Perception of biological motion in schizophrenia and healthy individuals: A behavioral and FMRI study. *PLos ONE*, *6*, e19971. doi:10.1371/journal.pone.0019971

Kuznetsova, A., Brockhoff, P. B., & Christensen, H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package), v. 2.0 -25. Retrieved from http://cran. r-project. org/package= lmerTest

Langdon, R., Corner, T., McLaren, J., Coltheart, M., & Ward, P. B. (2006). Attentional orienting triggered by gaze in schizophrenia. *Neuropsychologia*, *44*, 417–429.

Langdon, R., Ward, P. B., & Coltheart, M. (2008). Reasoning anomalies associated with delusions in schizophrenia. *Schizophrenia Bulletin*, *36*(2), 321–330.

Langdon, R., Seymour, K., Williams, T., & Ward, P. (2017). Automatic attentional orienting to other people's gaze in schizophrenia. *The Quarterly Journal of Experimental Psychology*, *70*, 1549–1558. doi:10.1080/17470218.2016.1192658

Langdon, R., Still, M., Connors, M. H., Ward, P. B., & Catts, S. V. (2014). Theory of mind in early psychosis. *Early Intervention in Psychiatry*, *8*, 286–290. doi:10.1111/eip.12072

Loughland, C., Draganic, D., McCabe, K., Richards, J., Nasir, A., Allen, J., et al. (2010). Australian and New Zealand *Journal of Psychiatry*, *44*, 1029–1035. doi: https://doi.org/10.3109/00048674.2010.501758

Magnee, M. J., Kahn, R. S., Cahn, W., & Kemner, C. (2011). More prolonged brain activity related to gaze cueing in schizophrenia. *Clinical Neurophysiology*, *122*, 506–511. doi:10.1016/j.clinph.2010.07.014

Marsh, P. J., Langdon, R., McGuire, J., Harris, A. W. F., Polito, V., Anandakumar, T., ... Coltheart, M. (2012). SOCOG: A novel social cognitive training program for people with schizophrenia. *Schizophrenia Research*, *136*, S38. doi:10.1016/S0920-9964(12)70138-X

Marsh, P. J., Langdon, R., McGuire, J., Polito, V., & Coltheart, M. (2013). An open clinical trial assessing a novel training program for social cognitive impairment in schizophrenia. *Australasian Psychiatry*, *21*, 122–126.

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

Nelson, H., & Willison, J. (1991). The Revised National Adult Reading Test (NART) – Test Manual (2nd ed.). NFERNelson: Windsor.

Pallanti, S., Quercioli, L., & Hollander, E. (2004). Social anxiety in outpatients with schizophrenia: A relevant cause of disability. *American Journal of Psychiatry*, *161*, 53–58.

Palmer, C. J., Caruana, N., Clifford, C. W. G., & Seymour, K. (2018a). Adaptive sensory coding of gaze direction in schizophrenia. *Royal Society Open Science*, *5*, 180886. doi:10.1098/rsos.180886

Palmer, C. J., Caruana, N., Clifford, C. W. G., & Seymour, K. (2018b). Perceptual integration of head and eye cues to gaze direction in schizophrenia. *Royal Society Open Science*, *5*, 180885. doi:10.1098/rsos.180885

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Quene, H., & van den Bergh, H. (2004). On multi-level modelling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*, 103–121.

Quene, H., & van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425.

Raine, A., & Benishay, D. (1995). The SPQ-B: A brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders*, *9*, 346–355. Retrieved from https://doi.org/10.1521/pedi.1995.9.4.346

Rosse, R. B., Kendrick, K., Wyatt, R. J., Isaac, A., & Deutsch, S. I. (1994). Gaze discrimination in patients with schizophrenia: Preliminary report. *American Journal of Psychiatry*, *151*, 919–921.

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

Seymour, K. J., Rhodes, G., McGuire, J., Williams, N., Jeffery, L., & Langdon, R. (2017). Assessing early processing of eye gaze in schizophrenia: Measuring the cone of direct gaze and reflexive orienting of attention. *Cognitive Neuropsychiatry*, *22*, 122–136.

Seymour, K. J., Rhodes, G., Stein, T., & Langdon, R. (2016). Intact unconscious processing of eye contact in schizophrenia. *Schizophrenia Research*, *3*, 15–19. doi:10.1016/j.scog.2015.11.001

Sprong, M., Schothorst, P., Vose, E., Hox, J., & van Engeland, H. (2007). Theory of mind in schizophrenia: Meta-analysis. *British Journal of Psychiatry*, *191*, 5–13.

SR Research. (2004). Experiment builder (Version 1.10.165). Ottawa, Ontario, Canada: SR Research Ltd.

Tso, I. F., Mui, M. L., Taylor, S. F., & Deldin, P. J. (2012). Eye-contact perception in schizophrenia: Relationship with symptoms and socioemotional functioning. *Journal of Abnormal Psychology*, *121*, 616–627.

Watson, T. L. (2013). Implications of holistic face processing in autism and schizophrenia. *Frontiers in Psychology*, *4*, 1–11.

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