



The neural time course of evaluating self-initiated joint attention bids



Nathan Caruana^{a,b,c,*}, Peter de Lissa^{a,b,d}, Genevieve McArthur^{a,b}

^a Department of Cognitive Science, Macquarie University, Sydney, Australia

^b ARC Centre of Excellence in Cognition and its Disorders, Sydney, Australia

^c Perception in Action Research Centre, Sydney, Australia

^d Department of Psychology, Macquarie University, Sydney, Australia

ARTICLE INFO

Article history:

Accepted 1 June 2015

Keywords:

Joint attention
Eye gaze
ERPs
Eye tracking
Virtual reality
Social interaction

ABSTRACT

Background: During interactions with other people, we constantly evaluate the significance of our social partner's gaze shifts in order to coordinate our behaviour with their perspective. In this study, we used event-related potentials (ERPs) to investigate the neural time course of evaluating gaze shifts that signal the success of self-initiated joint attention bids. **Method:** Nineteen participants were allocated to a "social" condition, in which they played a cooperative game with an anthropomorphic virtual character whom they believed was controlled by a human partner in a nearby laboratory. Participants were required to initiate joint attention towards a target. In response, the virtual partner shifted his gaze congruently towards the target – thus achieving joint attention – or incongruently towards a different location. Another 19 participants completed the same task in a non-social "control" condition, in which arrows, believed to be controlled by a computer program, pointed at a location that was either congruent or incongruent with the participant's target fixation. **Results:** In the social condition, ERPs to the virtual partner's incongruent gaze shifts evoked significantly larger P350 and P500 peaks compared to congruent gaze shifts. This P350 and P500 morphology was absent in both the congruent and incongruent control conditions. **Discussion:** These findings are consistent with previous claims that gaze shifts differing in their social significance modulate central-parietal ERPs 350 ms following the onset of the gaze shift. Our control data highlights the social specificity of the observed P350 effect, ruling out explanations pertaining to attention modulation or error detection.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Joint attention – the ability to achieve a common focus of attention with a social partner – supports language development, social communication, and learning (Charman, 2003; Mundy & Newell, 2007). Joint attention is achieved when one individual initiates a bid for joint attention – usually by gazing towards the intended focus of attention – and another individual responds by following their partner's line of regard (Bruner, 1974). Coordinating joint attention with others during dynamic interactions relies on the ability to evaluate the social significance of another's shift in gaze. This involves using the spatial properties of the gaze shift to represent a social partner's attentional, visual and mental perspective (Shepherd, 2010). For instance, to successfully respond to a joint attention bid, one must discriminate gaze shifts that signal

intentional bids for communication (Cary, 1978). Similarly, when initiating joint attention, we must evaluate our partner's responsive gaze to determine whether our bid for joint attention is successful. Despite the importance of this cognitive ability in our daily interactions, there is currently no empirical data elucidating the time course of neural processes involved in evaluating the social significance of gaze shifts in the context of joint attention interactions.

1.1. The second person neuroscience approach

The main challenge associated with investigating the neural processes supporting joint attention is that it can only be experienced during face-to-face interactions. Given that interactions are spontaneous and dynamic, scientific studies of joint attention have had to rise to the challenge of creating ecologically valid paradigms that also provide critical experimental control. In response to this challenge, the field of social neuroscience has seen the emergence of a *Second Person Neuroscience* approach (Schilbach et al., 2013). For example, several functional magnetic resonance

* Corresponding author at: Level 3, 16 University Avenue, Macquarie University, NSW 2109, Australia. Fax: +61 9850 6059.

E-mail addresses: nathan.caruana@mq.edu.au (N. Caruana), peter.delissa@mq.edu.au (P. de Lissa), genevieve.mcarthur@mq.edu.au (G. McArthur).

imaging (fMRI) studies have used interactive virtual reality paradigms to simulate experimentally controlled joint attention interactions (Caruana, Brock, & Woolgar, 2015; Schilbach et al., 2010; Wilms et al., 2010). These studies have pioneered gaze-contingent algorithms that allow participants to engage in a gaze-based interaction with a virtual character whom they believe is being controlled by another person in a nearby laboratory via live infrared eye tracking. This belief is important given the increasing evidence that perceiving agency and intentionality in social stimuli influences subjective experiences and eye movement patterns (Caruana et al., 2015), neural activation (Pfeiffer et al., 2014) and gaze-related ERPs (Pönkänen, Alhoniemi, Leppänen, & Hietanen, 2010; Wykowska, Wiese, Prosser, & Müller, 2014).

In an fMRI study, Schilbach et al. (2010) used a virtual reality paradigm that requested participants to interact with an anthropomorphic character whom – unbeknownst to them – was controlled by a computer. The virtual character was presented in the centre of the screen, surrounded by three squares that were positioned to the left, right, and directly above the animated face. On some trials, participants initiated joint attention by fixating on one of the three squares. A gaze-contingent algorithm was employed so that the virtual character would either respond congruently to achieve joint attention, or incongruently to avoid joint attention. Congruent responses were associated with greater activation in the ventral striatum. This supported claims that evaluating gaze shifts that signal the achievement of joint attention recruits reward-related neurocircuitry. Other gaze-congruency fMRI studies using pre-recorded actors have also found that congruent gaze and head orientation responses that result in joint attention correlate with differential activation in brain regions associated with perspective taking (mPFC, ACC, TPJ; Gordon, Eilbott, Feldman, Pelphrey, & Vander Wyk, 2013; Williams, Waiter, Perra, Perrett, & Whiten, 2005) and social reward (striatum; Gordon et al., 2013).

1.2. Event-related potential (ERP) studies

While fMRI studies have been useful in elucidating the brain structures involved in evaluating self-initiated joint attention bids, they have been unable to reveal the time course of the associated neural processes due to the sluggish nature of the BOLD signal (see Menon & Kim, 1999 for review). In contrast, event-related potentials (ERPs), which represent the average pattern of electrical activity related to a particular stimulus event, can be used to track the time course of neural processes in (practically) real time because electrical activity travels at the speed of light (Luck, 2005). Another advantage of ERPs is that they can be measured “passively” without a person’s overt attention. In the context of joint attention, this would allow participants to remain focused on their social partner without disrupting the continuity of the interaction and thus maintaining ecological validity.

Despite the advantages offered by passive ERPs, no study to date has used ERPs to investigate the time course of neural processes involved in evaluating joint attention bids. However, ERPs have been used to investigate the time course of neural processing associated with the evaluation of (1) the perceptual properties of gaze, (2) attention cues from gaze, and (3) the social significance of gaze.

1.2.1. Evaluating the perceptual properties of gaze

A number of studies have used the N170 ERP (a negative peak that occurs around 170 ms after the onset of a stimulus) to explore the timing of neural processes associated with evaluating the perceptual properties of gaze. Most of these studies have measured the N170 during the passive viewing of direct and averted gaze (see Itier & Batty, 2009 for review). The findings have been mixed. Studies employing static stimuli report either no modulation of

the N170 for direct and averted gaze (Grice et al., 2005; Schweinberger, Kloth, & Jenkins, 2007) or small effects in which averted gaze evokes larger (more negative) N170 amplitudes (Watanabe, Miki, & Kakigi, 2002). Larger N170 effects of gaze direction have been found using dynamic gaze stimuli. For example, Puce, Smith, and Allison (2000) reported that ‘direct – averted’ gaze shifts evoked larger N170 amplitudes than ‘averted – direct’ gaze shifts. In contrast, others report that ‘slightly averted – direct’ gaze shifts evoked larger N170 amplitudes than ‘slightly averted – extremely averted’ gaze shifts (Conty, N’Diaye, Tijus, & George, 2007).

These studies have suggested the N170 may be sensitive to the different social signals conveyed by direct and averted gaze. Specifically, direct gaze may signal a readiness to communicate, whilst averted gaze may signal the initiation or response to a joint attention bid (Cary, 1978; Conty, N’Diaye, Tijus, & George, 2007; Kleinke, 1986). However, the outcomes of these studies provide limited insight into the time course of neural processes associated with evaluating self-initiated joint attention bids for three reasons. First, the direction of N170 effects across studies are inconsistent. Second, these studies typically employed passive viewing tasks that did not provide an interactive context. Third, these studies demonstrated that the N170 was affected by manipulating the perceptual properties of gaze (direct versus averted). There currently is no evidence to suggest that the N170 is affected by the evaluation of the social significance of gaze.

1.2.2. Evaluating attention cues from gaze

Studies examining the time course of neural processes associated with the evaluation of attention cues from gaze have often measured the EDAN (early attention direction negativity) and ADAN (anterior directing attention negativity) ERPs (e.g. Feng & Zhang, 2014; Hietanen, Nummenmaa, Nyman, Parkkola, & Hämäläinen, 2006; Holmes, Mogg, Garcia, & Bradley, 2010; Lassalle & Itier, 2013; van Velzen & Eimer, 2003). The EDAN is typically measured over posterior sites (e.g. P7 and P8) 200–300 ms post cue presentation. The ADAN is measured over anterior sites (e.g. C3 and C4) 300–500 ms post gaze cue presentation. These peaks are believed to represent reflexive attention orienting and attentional control mechanisms respectively, and produce maximal responses at electrode sites contralateral to cued locations in gaze-cueing tasks (c.f. Friesen & Kingstone, 1998). In these studies, participants are typically asked to detect a target on the left or right side of the screen following the presentation of a valid or invalid gaze cue (see Frischen, Bayliss, & Tipper, 2007 for review). Using impoverished schematic face stimuli, one study found that the EDAN and ADAN were modulated by arrow cues but not gaze (Hietanen, Leppänen, Nummenmaa, & Astikainen, 2008). A second study using photographic faces reported that gaze cues did not produce a significant EDAN effect, but did result in a significant gaze-congruent ADAN effect (Holmes et al., 2010). A third study using photographic gaze cues reported significant gaze-congruent EDAN and ADAN effects (Feng & Zhang, 2014).

This inconsistent evidence for gaze-congruent EDAN and ADAN effects suggests that these peaks may not be reliable neural markers of gaze processing. Furthermore, whilst studies investigating the EDAN and ADAN attempt to elucidate the effects of gaze on attention orienting – analogous to responding to joint attention bids – they do not inform the time course of neural processing when we evaluate a gaze shift after we have already fixated a peripheral target (analogous to initiating joint attention). As such, it is difficult to determine whether the EDAN or ADAN ERPs represent processes pertaining to the evaluation of attention cues from gaze or the social significance of gaze. Further, gaze cueing tasks used to elicit the EDAN and ADAN ERPs often lack ecological validity since participants are often instructed that the gaze cues do not predict the target’s location.

1.2.3. Evaluating the social significance of gaze

In contrast to the N170, EDAN, and ADAN studies outlined above, a number of researchers have begun to investigate whether later-developing ERPs are modulated by different evaluations of the social significance of gaze. For example, Sabbagh, Moulson, and Harkness (2004) presented participants with trials that comprised a written label relating to gender (e.g. female) or an emotion (e.g. happy) followed by a static picture of a pair of eyes. The eyes varied in gender, emotion, and gaze orientation. Participants were asked to indicate whether or not the label was congruent with the gender or emotion depicted in the gaze stimulus. Eyes judged for emotion generated large N270–N400 responses over right inferior frontal and anterior temporal regions, and larger P300–P500 responses over posterior parietal regions, than eyes judged for gender.

From these findings, Sabbagh et al. (2004) concluded that these ERPs reflected the decoding of another person's emotional mental state. However, the validity of the outcomes are questioned by four methodological limitations of this study. First, the stimuli were ecologically unrealistic, and were not presented in the context of a face. Second, emotion was expressed by the configuration of the eye region (e.g. frowning or raised eye brows). Thus, it is not clear if the effects in Sabbagh et al. relate to the evaluation of gaze or the evaluation of facial configuration. Third, there was no indication of what the eyes were gazing at. Gaze is generally processed in conjunction with the environmental context. For example, if a social partner averts their gaze, the gaze shift is typically evaluated with respect to the object that the agent is gazing towards. If these ERP effects truly associate with the evaluation of another person's perspective, it remains unclear as to whether the time course of these ERPs would be the same when individuals evaluate another person's perspective with respect to the environmental context (e.g. visual perspective taking). Finally, this study used static gaze stimuli, in which either direct or averted gaze was used. Thus it is unclear whether the time course of these effects would differ in real life, where gaze is dynamic.

Carrick, Thompson, Epling, and Puce (2007) addressed many of these limitations by presenting participants with trials that comprised three horizontally aligned faces (a central face and two flanker faces). The gaze of both flanker faces were directed either to the left or right. The gaze of the central face, which was initially directed towards the participant, was updated to either match the flanker faces (the “group” condition), to face towards one flanker face (and thus away from the other face; the “mutual” condition), or to gaze upwards away from both flanker faces (the “avoid” condition). The onset of the updated central faces in the group and mutual conditions generated earlier P350 and smaller P500 ERP peaks, relative to the avoid condition. Carrick et al. concluded that the P350 and P500 peaks reflected the integration of the spatial properties of gaze in order to evaluate its social significance. N170 responses measured at occipitotemporal sites were reliably elicited by each gaze shift, but were not modulated by the social significance of the gaze-shift. This is consistent with claims that the N170 involves activity reflecting the perceptual processing of gaze (e.g. gaze orientation; Itier & Batty, 2009).

Carrick et al.'s (2007) findings suggest that the brain begins to process or “recognise” the social significance of gaze 350–500 ms after the onset of relevant facial stimuli. However, the strength of this suggestion is mitigated somewhat by four methodological limitations of this study. First, the ecological validity of the stimuli was questionable. People rarely have to evaluate the social significance of gaze based on three faces presented in a row that look left, right, or upwards. Rather, gaze is usually evaluated during face-to-face interactions. Second, it was presumed that the “mutual condition” reflected the processing of mutual gaze since the central face and one flanker face were looking at each other.

However, the same stimuli could have been interpreted as a “social exclusion” condition since the central face was simultaneously looking away from the other flanker face. Third, the stimuli in the mutual condition (i.e. three faces gazing in the same direction) might be considered less complex than the group and avoid conditions, which comprised faces looking in different directions. Fourth, the avoid condition used faces that gazed in a different direction (upwards) than the mutual (left and right) and group (all left or all right) conditions. This meant that the faces in the avoid condition differed from the mutual and group condition in both gaze orientation (upwards rather than left or right) and social significance (avoid rather than group or mutual conditions). In sum, multi-face displays provide an innovative yet noisy method for manipulating the social significance of gaze shifts. Thus, we cannot be absolutely sure that the P350 and P500 effects identified by Carrick et al. (2007) specifically reflect differences in evaluating the social significance of gaze.

Given the promising, yet somewhat limited, outcomes of Carrick et al. (2007), the aim of the current study investigated whether the identified P350 and P500 effects are observed when the social significance of a gaze shift is evaluated during face-to-face interactions. To this end, we measured the P350 and P500 ERPs of 19 adults while they interacted with a virtual character believed to display the eye movements of another human via live infrared eye-tracking. In reality, the virtual agent was controlled by a gaze-contingent algorithm (c.f. Caruana et al., 2015; Wilms et al., 2010). This is important given that beliefs about the human agency of gaze stimuli have been found to specifically influence gaze-related ERPs (Pönkänen et al., 2010; Wykowska et al., 2014). To discount the possibility of effects being driven by non-social phenomena (e.g. stimulus congruity) we also employed a non-social control condition. Here another 19 adults completed the same task, except arrows replaced the interactive gaze stimuli and participants were instructed that they were completing the task alone.

Based on Carrick et al.'s (2007) seminal findings, we predicted that evaluating “incongruent” gaze shifts following a self-initiated bid for joint attention (i.e., “my partner is not attending to the same thing as I am”) would trigger larger and later ERPs (P350 and P500) than “congruent” gaze shifts (“my partner is attending to the same thing as I am”). We anticipated that these later ERP effects would be absent, or reduced, in the non-social control condition. It was also expected that all conditions would elicit clear occipitotemporal N170 peaks, but that these would not be modulated by congruency, given that the perceptual properties of the stimulus remained constant (Itier & Batty, 2009).

2. Method

The method of this study was approved by the Macquarie University Human Research Ethics Committee.

2.1. Participants

This study used an independent-groups design that included two condition groups (“social” versus “control”) that each responded to two conditions of stimuli (“congruent” versus “incongruent”). Participants volunteered or received course credit for their time and provided consent before participating.

2.1.1. Social condition

Twenty-four individuals were recruited into the social condition. Due to technical challenges, eye-tracking calibration was successful for 22 participants. Additionally, two participants reported that they were not completely convinced that the virtual agent was

representing the eye movements of another person. The behavioural data of another participant indicated that they had not appropriately engaged with the task (see Behavioural data in Results). These participants were excluded, resulting in a final sample of 19 participants (3 male, $M_{age} = 20.95$, $SD = 5.78$) for the social condition.

2.1.2. Control condition

Another 19 individuals participated in the control condition (7 male, $M_{age} = 29.12$, $SD = 9.24$). All participants were included in the final sample given that the eye tracking calibration was successful for all participants, and all participants were appropriately engaged in the task (see Behavioural data in Results). The difference between the mean age of the social and control groups was not statistically significant ($t(36) = -1.136$, $p = .264$).

2.2. Stimuli

2.2.1. Social stimuli

An anthropomorphic virtual character (whom we called “Alan”) was generated using *FaceGen* (Singular Inversions, 2008). The character depicted a white Caucasian male, and subtended $8 \times 12^\circ$ of visual angle in the centre of a computer screen (a Samsung SynchMaster SA950 HD LED monitor [60×34 cm] with a refresh rate of 120 Hz, at a distance of 65 cm from the participant). The virtual character’s gaze was manipulated to create five images so that the eyes were either directed at the participant or towards four cartoon buildings that were presented at each corner of the computer screen. The building stimuli, which were created using *GIMP-2* (Kimball & Mattis, 1995), each subtended 11° of visual angle, with 15° of visual angle between each building and the virtual character’s eyes. The stimuli were presented via *Experiment Builder* 1.10.165 (SR Research, 2004).

2.2.2. Control stimuli

Only the central stimulus differed between the social and control conditions. Specifically, the animated face remained on the screen with eyes closed to provide a close match for the presence of facial stimuli (c.f. Caruana et al., 2015). A fixation point subtending 1.4° of visual angle was positioned over the nose. Green arrow stimuli, which protruded from this fixation point replaced the gaze stimuli, and subtended 4° of visual angle (see Fig. 1 for a comparison of social and control task stimuli).

2.3. Stimulus conditions

2.3.1. Social condition

Participants in the social condition were told that they would be playing a cooperative game with “Alan” called “Catch the Prisoner”. The aim was to jointly catch a prisoner who would attempt to escape from the prison compound on each trial. Participants were told that they would be the “watch person” while Alan would play the “guard”. The watch person’s task was to monitor the outside of the prison, while the guard’s task was to monitor inside the prison. Participants were told that the guard may sometimes be distracted by inmates fighting in different locations of the prison.

Each trial began with the presentation of a crosshair subtending 1.4° of visual angle. Once the participant fixated on the crosshair for 150 ms, the stimulus updated to display the four prison buildings and the animated face in the centre of the screen (see Fig. 2). The face was positioned so that the nasion was in the same location as the crosshair. This ensured that participants were attending to the gaze stimulus from the beginning of the trial.

A prisoner then attempted to escape from one of four buildings that were each located in a different corner of the display screen.

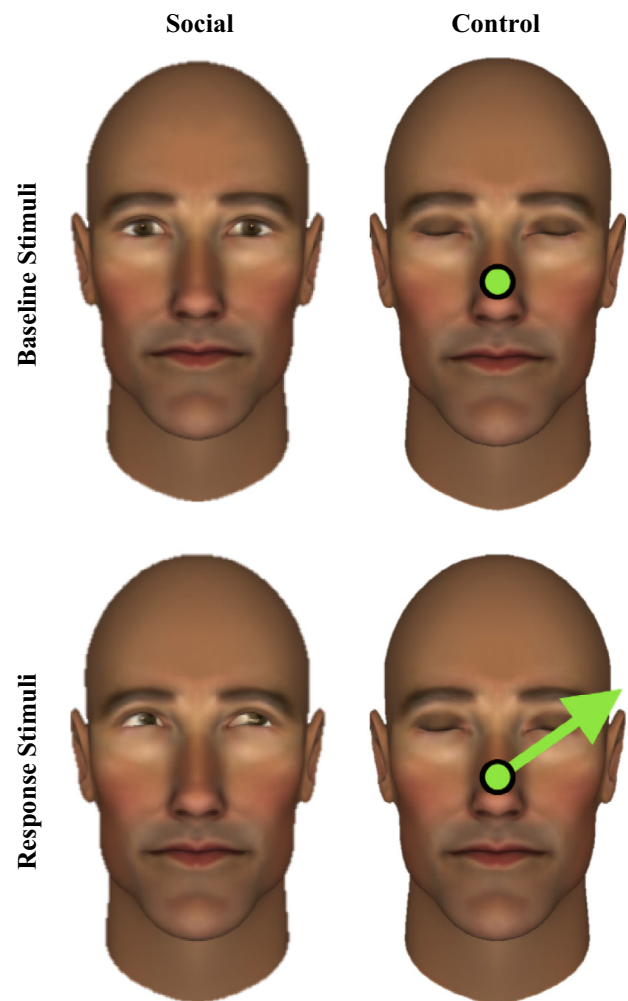


Fig. 1. Central stimuli used in the social and control versions of the gaze-contingent task.

Provided the participant remained fixated on the gaze stimulus, the location of this “breakout” was indicated by a yellow circular sensor light at one of the four prison exits after 200–1000 ms (i.e. only the watch person could see the light – not the guard). This exogenously cued the participant to the breached location. The participant was then required to initiate a joint attention bid (i.e., look at the location of the escaping prisoner). Once the participant had fixated upon the spotlight, a cartoon prisoner appeared after 200–1000 ms, provided fixation had been maintained for 150 ms.

The participant was then required to fixate upon the animated face in order to evaluate their partner’s response. The virtual character’s gaze averted after 350–650 ms post-fixation. This ensured that (1) the gaze behaviour appeared realistic, and (2) the participant was fixating the virtual character when the gaze shift occurred. Epochs were time-locked to the onset of this gaze shift. In 50% of the trials, the virtual character’s gaze shifted congruently to the correct location of the escaping prisoner (joint attention), presumably “locking-down” the exit. In the remaining 50% of trials, gaze shifted to an incongruent location (failed joint attention), presumably due to distractions by fighting inmates within the prison. Congruent and incongruent trials were presented in random order across four blocks of 60 trials (i.e. 120 trials in each condition). The direction of congruent and incongruent gaze shifts were counter-balanced across all acquisition blocks, and thus the virtual character was equally likely to gaze towards one of the three houses not

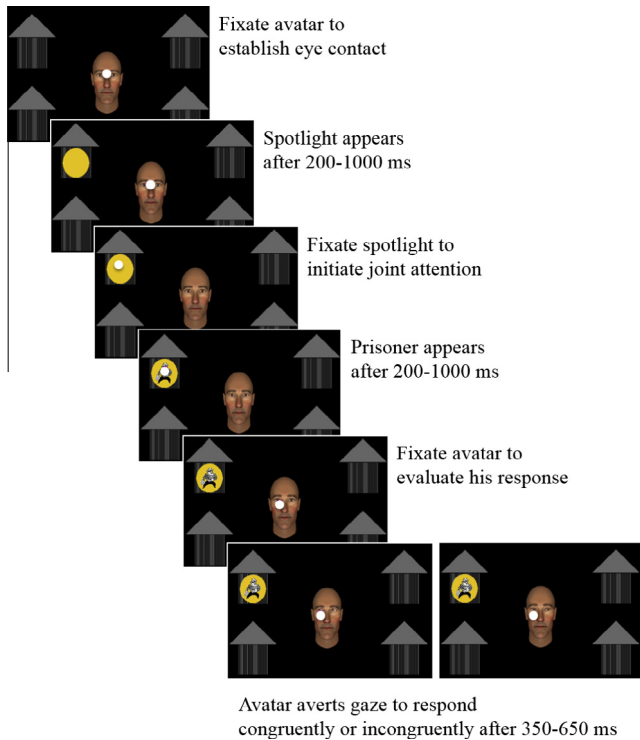


Fig. 2. Schematic representation of trial sequence. **White circle represents the location of the participant's gaze and was not part of the stimuli visible to the participant.**

fixated by the participant on incongruent trials. At the end of each block participants were asked to estimate the proportion of trials they thought they were successful in catching the prisoner, based on stimulus congruity. This provided a measure of task engagement.

2.3.2. Control condition

The task completed by individuals in the control condition was the same as the social condition except that they were told that they would be completing this task on the computer (i.e., not in conjunction with a fictitious partner). Once participants had revealed the burglar and fixated the central fixation point, it would turn green (analogous to establishing mutual gaze) and the arrow would point either (1) congruently towards the location previously fixated by the participant, or (2) incongruently to one of the three remaining locations. Again, participants were told that this signalled whether the computer had detected their response to “catch” the prisoner.

In both the social and control condition, negative feedback was provided for trials where the participant (1) failed to fixate the location where the spotlight appeared, (2) fixated away from the spotlight before the prisoner appeared, (3) took longer than 3000 ms to fixate back on the central stimuli (i.e. gaze stimulus/fixation point) after the prisoner appeared, or (4) fixated away from the central stimuli within 1000 ms of fixating the central stimuli. Thus, a key point of difference between the gaze-contingent algorithm developed in this study and previous studies (Schilbach et al., 2010; Wilms et al., 2010) is that whilst earlier algorithms updated the agent's gaze after the participant fixated one of the target locations, our algorithm also required participants to fixate back on their partner's eyes (or the fixation point in the control condition) before the stimulus was updated. This ensured that participants were fixating the gaze and arrow stimuli when ERPs were being measured. Our algorithm also employed temporal jitter for

the onset latencies of the gaze shift and arrow presentations to mitigate the influence of anticipation on the resulting ERPs.

2.4. Eye movement and electroencephalogram (EEG) recording

Each participant's eye-movements and EEG were recorded while they completed the stimulus conditions. Eye-movements were tracked using an EyeLink 1000 monocular tower-mounted eye tracker (right eye only) at a sampling rate of 1000 Hz. A chin rest was used to stabilise participants' heads, and standardise viewing distance. Participants' EEGs were measured using a montage of 29 electrodes positioned according to the 10–20 system (EasyCap; FP1, FP2, F7, F3, FZ, F4, F8, FT7, FC3, FC4, FT8, T7, C3, CZ, CPZ, C4, T8, TP7, CP3, CP4, TP8, P7, P3, Pz, P4, P8, O1, OZ, O2). Online and offline reference electrodes were attached to the left and right earlobes respectively, and the ground electrode was positioned between the FP1, FP2 and FZ electrodes. The impedances for all electrodes were maintained below 5 k Ω . Bipolar electrodes were used to measure horizontal electro-ocular activity (HEOG; positioned at the outer canthi) and vertical ocular activity (VEOG; positioned above and below the left eye). A Synamps II amplifier was used to record the online EEG with a sampling rate 1000 Hz, an online band pass filter of .05–100 Hz, and a notch filter at 50 Hz.

2.5. Creating ERPs

The EEG data was processed offline using Neuroscan 4.5 software (Neuroscan, El Paso, Texas, USA). VEOG activity was removed using a standard ocular reduction algorithm (Neuroscan, El Paso, Texas, USA). The EEG was then band-pass filtered (0.1–30 Hz) with a 12 dB octave roll-off. Epochs were time-locked to the onset of the virtual character's averted gaze, starting 100 ms before a gaze shift (i.e., a prestimulus baseline of –100 to 0 ms) and ending 700 ms later. Epochs exceeding ± 100 mV were deleted from the analysis. Remaining epochs were baseline corrected according to pre-stimulus electrical activity. Then each participant's accepted epochs for congruent and incongruent trials were averaged to produce congruent and incongruent ERPs respectively. Grand average ERPs were then formed from the congruent and incongruent ERPs.

2.6. Measuring ERPs

To ascertain the length of the intervals used to measure each ERP in this study (P350 and P500 at CZ and PZ, and N170 at P7 and P8), we visually inspected the relevant ERPs of each individual participant. This revealed that a P350 peak was the most reliably elicited ERP measured at centro-parietal electrodes across individuals in the congruent and incongruent social conditions. A clear P350 peak could only be identified in four individuals in the control condition. Visual inspection of the data suggested that the social and control conditions elicited similar voltages up until approximately 250 ms. After this time, evoked potentials in the social condition increased in voltage and peaked at approximately 350 ms, whereas those in the control condition did not manifest this strong positivity and drifted back to baseline (see Fig. 3).

Given that only four individuals in the control condition displayed clear P350 peaks, it was only possible to measure the peak latency of the P350 in the social condition. We were also unable to detect reliable P500 peaks across individuals in any of the conditions. Thus, this study focussed on analysing mean amplitude (P350 and P500) and peak amplitude (N170) data. However, it is noteworthy that in line with Carrick et al. (2007), we found a significant latency effect in the social condition whereby the P350 following incongruent gaze shifts (CZ $M = 373.05$, $SD = 33.85$; PZ $M = 383.47$, $SD = 27.04$) was significantly slower to peak relative to those following congruent gaze shifts (CZ $M = 343.68$,

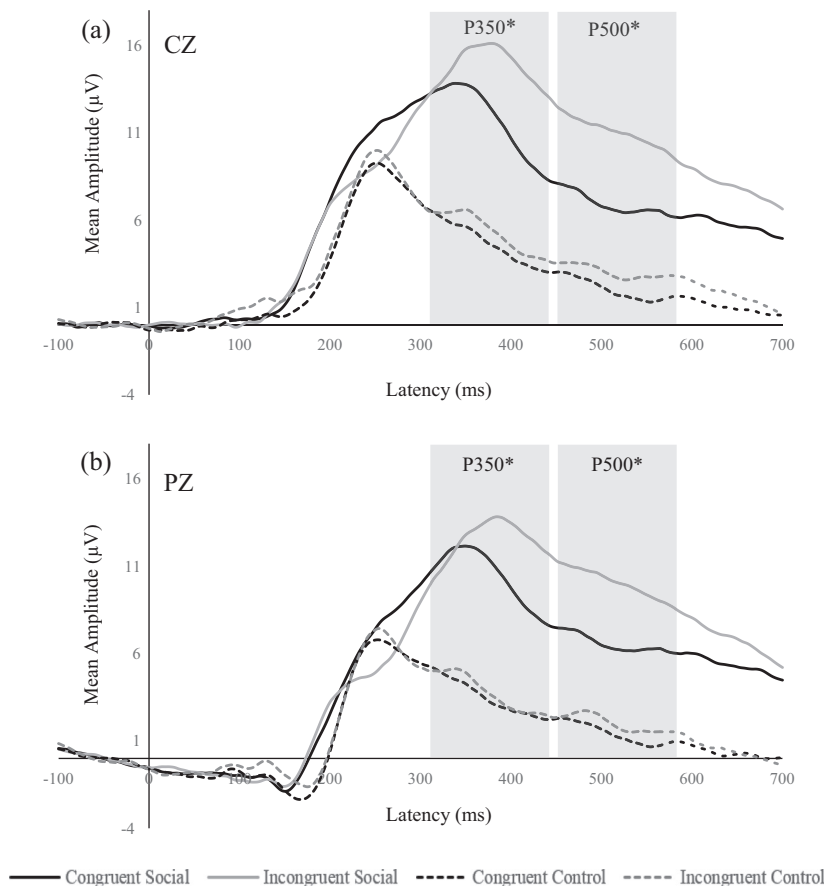


Fig. 3. Group average waveforms comprising the P350 and P500 at (a) Cz and (b) Pz electrodes. Epochs were time-locked to the onset of the virtual character's gaze shift. * Indicates a significant Group * Condition interaction effect.

$SD = 20.95$; PZ $M = 351.79$, $SD = 22.83$) at both CZ [$t(18) = 4.50$, $p < .001$] and PZ [$t(18) = 5.24$, $p < .001$].

A 130 ms interval (310–400 ms) captured each individual's P350 peak in both the congruent and incongruent conditions. Thus, we used 130 ms intervals to measure all the peaks of interest, including the P350 (310–400 ms) and P500 (from 450 to 580 ms), both at CZ and PZ, and the N170 response recorded at P7 and P8 (107–237 ms). The intervals used to measure the consecutive positive responses (P350 and P500) were separated by a gap of 10 ms to ensure that each participant's positive peaks were measured in the correct interval (e.g. to ensure that an individual's P350 did not fall in the P500 interval).

2.7. Analysis

For each individual and condition, the size of the ERP peaks were measured using mean amplitudes (for P350 and P500 at CZ and PZ) and peak amplitudes (for N170 at P7 and P8). A two-way ANOVA was used to assess the within-subjects factor of stimulus condition (congruent, incongruent) and the between-subjects factor of group condition (social, control) on each of the above measures (Statistical Package for the Social Sciences v19).

2.8. Ecological validity questionnaire

At the end of the experiment, participants rated their experience of the task and interaction with Alan on a five-point Likert scale (1 = not at all to 5 = extremely). All participants rated how difficult, intuitive, and pleasant the task was. Those in the social

condition also rated how natural they found the interaction with Alan, and how effective he was at responding correctly to their guiding gaze. Then these participants were asked how convinced they were that they had been interacting with another living person. Additionally, individuals in the control group were asked to rate how interactive they found the task.

3. Results

3.1. Ecological validity questionnaire

Participants in the social and control condition rated the interactive task as easy, intuitive and pleasant (see Table 1 for descriptive statistics). There were no significant differences between the social and control conditions on any of these dimensions (independent t -tests, all $ps > 0.29$).

3.1.1. Social condition

Participants in the social condition rated the interaction as feeling “moderately” to “very” natural on average ($M = 2.58$,

Table 1
Post experimental inquiry ratings of subjective experience.

Task aspect	Social $M(SD)$	Control $M(SD)$
Difficulty	1.68(0.75)	1.47(0.70)
Intuitiveness	4.32(0.82)	4.57(0.69)
Pleasantness	3.11(1.66)	3.10(0.66)

Note: Ratings provided on a 5-point scale (1 = low, 5 = high).

$SD = 1.26$). Consistent with the gaze congruency manipulation, these participants also rated their partner as performing poorly on the task ($M = 2.21$, $SD = 0.92$). All participants in the final sample reported that they were convinced that the virtual character was being controlled by a human interlocutor, and rated the degree of their belief on the same five-point scale ($M = 4.89$, $SD = 0.32$). Those who provided a 4/5 rating (rather than 5/5) claimed that they momentarily considered the possibility that the virtual character may have been controlled by a computer, but saw no reason not to accept the interaction as genuine.

3.1.2. Control condition

Participants in the control condition on average provided ratings suggesting that they did not find the arrow stimulus interactive at all ($M = 1.31$, $SD = 0.67$). This indicated that although the arrow stimulus was responding contingently to the participants gaze behaviour, this did not result in the anthropomorphisation of the arrow stimulus, confirming that our social manipulation was effective.

3.2. Attention to gaze shifts

After each block, participants estimated the percentage of trials that Alan ($M = 48.33\%$, $SD = 11.05$) or the computerised arrow ($M = 48.61\%$, $SD = 4.14$) responded congruently. One participant from the social group obtained an average congruency estimate that was two standard deviations above the group mean ($M = 87.00$), and they were consequently excluded from all analyses (see Method, Participants, Social group).

3.3. ERPs

Summary statistics for the amplitude measures are shown in Table 2. Group average waveforms comprising the P250, P350 and P500 at CZ and PZ are shown in Fig. 3, and for the N170 at P7 and P8 are shown in Fig. 4. Topographies demonstrating the effect of stimulus condition (congruent–incongruent) are depicted separately for the social and control conditions in Fig. 5.

3.3.1. P350 mean amplitude

A significant group * condition interaction effect was measured at CZ ($F(1,36) = 9.21$, $p = .004$), and PZ ($F(1,36) = 5.42$, $p = .026$). In the social condition, the mean amplitude of the P350 following incongruent gaze shifts was significantly larger than congruent gaze shifts, when measured at both CZ ($t(18) = -4.80$, $p < .001$) and PZ ($t(18) = 3.43$, $p = .003$). However a significant P350 effect was not observed for arrows in the control condition when

measured at either CZ ($t(18) = 1.45$, $p = .165$) or PZ ($t(18) = 0.64$, $p = .532$).

3.3.2. P500 mean amplitude

Like the P350 response, a significant group * condition interaction effect was measured at CZ ($F(1,36) = 9.21$, $p = .004$), and PZ ($F(1,36) = 5.42$, $p = .026$) for the P500. The mean amplitude of the P500 following incongruent gaze shifts was significantly larger than congruent gaze shifts, when measured at both CZ ($t(18) = 7.34$, $p < .001$) and PZ ($t(18) = 7.49$, $p < .001$) in the social condition. However a significant P500 effect was not observed for arrows in the control condition when measured at either CZ ($t(18) = 1.58$, $p = .131$) and PZ ($t(18) = 0.87$, $p = .395$). Unlike the P350, this P500 effect was not represented by clear peaks in the individual data from either of the group conditions. Thus, it seems likely that the larger mean amplitude for incongruent gaze responses at this latency may merely represent an artefact of the earlier diverging P350 response.

3.3.3. N170 peak amplitude

A group * condition interaction effect did not reach statistical significance when measured at P7 ($F(1,36) = 0.25$, $p = .621$), or P8 ($F(1,36) = 0.21$, $p = .885$). However, a main effect of group was identified at P8 ($F(1,36) = 8.65$, $p = .006$) in which the N170 was significantly larger in the social condition than the control condition. This was not surprising given the perceptual differences between gaze and arrows. However, this main effect did not reach statistical significance when measured at P7 ($F(1,36) = 0.878$, $p = .355$). This is consistent with claims that the face-related N170 is most sensitively measured from the right hemisphere (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Joyce & Rossion, 2005).

4. Discussion

The current study investigated the time course of neural processes involved in evaluating whether a gaze shift signals the success or failure of a self-initiated joint attention bid. Participants allocated to the social condition interacted with an anthropomorphic character whom they believed represented the gaze behaviour of another human participant. The task required participants to initiate goal-oriented bids for joint attention. The virtual partner either responded congruently or incongruently with equal probability, thus indicating a success or failure to achieve joint attention respectively. The ERPs following congruent or incongruent gaze shifts were measured to assess the time point at which they differed, indicating a divergence of their neural processing. To determine whether potential differences between congruent and incongruent gaze shifts were specific to social cognitive processes, these ERPs were compared with those measured in a second group of individuals who completed a non-social analogue of the same task. This control condition only differed in that arrow stimuli replaced the virtual character's gaze shift, and participants did not believe that they were engaged in an interaction with another human. As expected, congruent and incongruent gaze and arrow stimuli did not result in the modulation of the occipitotemporal N170. However a significant group * condition interaction effect was observed for P350 and P500 peaks measured at centro-parietal electrodes. As depicted in Fig. 3, these interaction effects are characterised by (1) large differences in the mean amplitude of congruent and incongruent ERPs in the social condition, and (2) little discrimination between congruent and incongruent ERPs in the control condition.

Consistent with Carrick et al.'s (2007) findings, a modulation of the P350 at centro-parietal sites was observed when participants'

Table 2
Summary statistics for amplitude and latency measures by electrode.

	CZ		PZ	
	Congruent	Incongruent	Congruent	Incongruent
<i>P350 mean amplitude</i>				
Social	11.85(4.60)	14.90(5.45)	10.52(4.15)	12.53(4.53)
Control	4.71(4.00)	5.38(4.94)	3.52(4.06)	3.81(4.73)
<i>P500 mean amplitude</i>				
Social	6.91(4.49)	11.07(5.35)	6.57(4.32)	10.08(4.65)
Control	2.02(2.80)	3.03(4.26)	1.41(2.69)	2.01(4.22)
	P7		P8	
	Congruent	Incongruent	Congruent	Incongruent
<i>N170 peak amplitude</i>				
Social	-7.42(4.59)	-7.24(4.27)	-9.79(5.48)	-9.89(5.05)
Control	-6.19(2.96)	-6.25(2.76)	-5.24(4.45)	-5.28(4.30)

Note: Summary statistics are provided in the format of $M(SD)$.

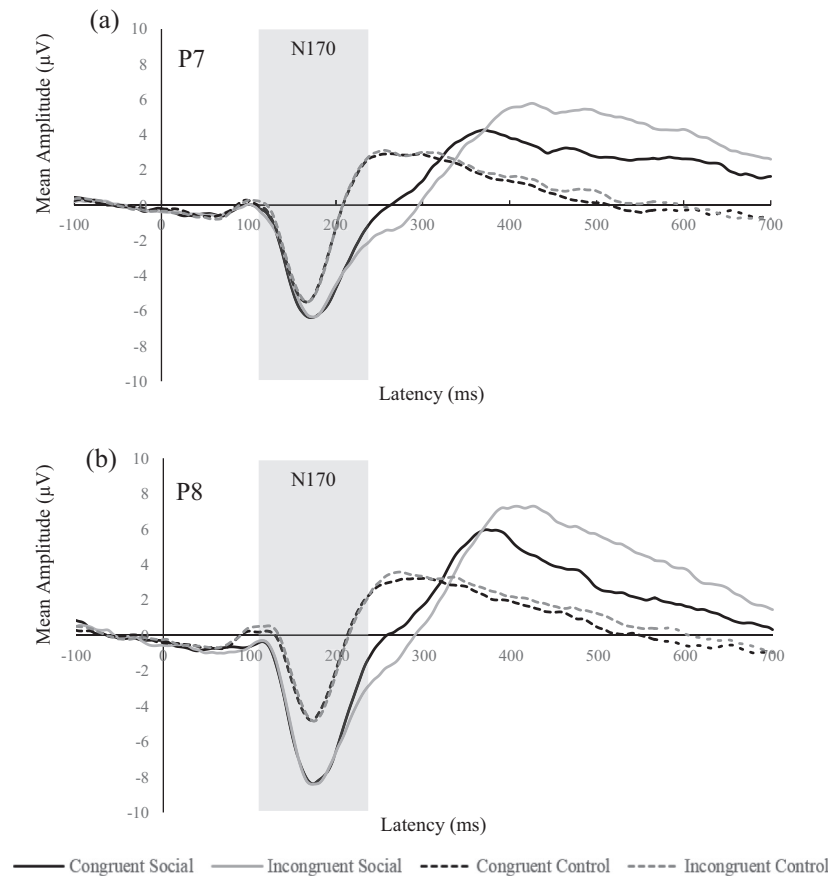


Fig. 4. Group average waveforms comprising the N170 at (a) P7 and (b) P8 electrodes. Epochs were time-locked to the onset of the virtual character's gaze shift.

viewed perceptually identical gaze shifts that differed only in whether they signalled the failure or success of a self-initiate joint attention bid. Larger and later P350 peaks were elicited by incongruent gaze shifts signalling a failed joint attention bid compared to congruent (joint attention) gaze shifts. The same effect was also measured during the P500 interval. However, the absence of any clear peaks during this interval suggests that the P500 effect is unlikely to reflect an independent effect or cognitive process. The present study is the first to investigate the morphology of ERPs when actively processing a social partner's gaze in the context of an interaction, involving genuine and goal-oriented joint attention bids.

These findings are particularly compelling for a number of reasons. Firstly, our findings corroborate with those reported by Carrick et al. (2007) despite employing a different paradigm. What both studies have in common is the manipulation of a gaze-shift's social consequence. That is, whether a gaze shift signals a willingness or reluctance to interact (Carrick et al., 2007), or the achievement or failure-to-achieve joint attention. The corroborating findings suggest that the centro-parietal P350 response may be particularly sensitive to the interpretation of these social consequences.

Considering the social condition data alone, one could argue that the P350 effect observed in the current study could represent a non-social attention mechanism that is being modulated by stimulus congruity (e.g. odd-ball, error detection or attention orienting effects). However, explaining the P350 effect as a non-social phenomenon is mitigated by the absence of a congruity effect in the control condition (see Fig. 5). Participants in the social and control condition completed identical tasks. In both cases they oriented to the prisoner's location, then check to see whether their partner (signalled by gaze) or the computer (signalled by arrows)

registered their response to 'capture' the prisoner. If congruity were modulating attention, then such an effect would be expected to manifest – to some extent – in both the social and control conditions. The absence of a congruity effect in the control condition highlights the social specificity of the measured P350 effect.

Clear N170 responses were measured in all conditions at occipitotemporal sites (P7 and P8). However, these were not modulated by congruity in either the social or control condition. Observing the morphology of ERPs at these sites in the control condition also mitigates any concern that the congruity of the central stimuli has an influence on attention. These occipitotemporal sites have been used to measure EDAN effects in cueing studies using both gaze and arrow stimuli (e.g. Feng & Zhang, 2014). If the congruity of these stimuli were having any influence on the attention of participants – in a way that was independent of social interaction – then we would expect this to result in some modulation of ERPs measured at these sites between 200 and 300 ms post stimulus onset. However, we found no evidence of this in the control condition.

Interpreting the P350 as an attentional effect is further mitigated by the fact that it was also observed in Carrick et al.'s (2007) multi-face paradigm, which did not employ a congruity manipulation. It seems unlikely in their passive-viewing task that attention was modulated by task-irrelevant gaze shifts, made by non-agent photographic faces. Taken together, the P350 effect cannot be explained as an effect of attention modulation. Rather, this effect seems specific to the process of evaluating the social consequence of a gaze shift – in this case whether a social partner shares a common or different focus of attention.

This social interpretation of the P350 effect compliments fMRI data from a study employing a similar task (Schilbach et al., 2010). Schilbach et al. reported that increased activity was

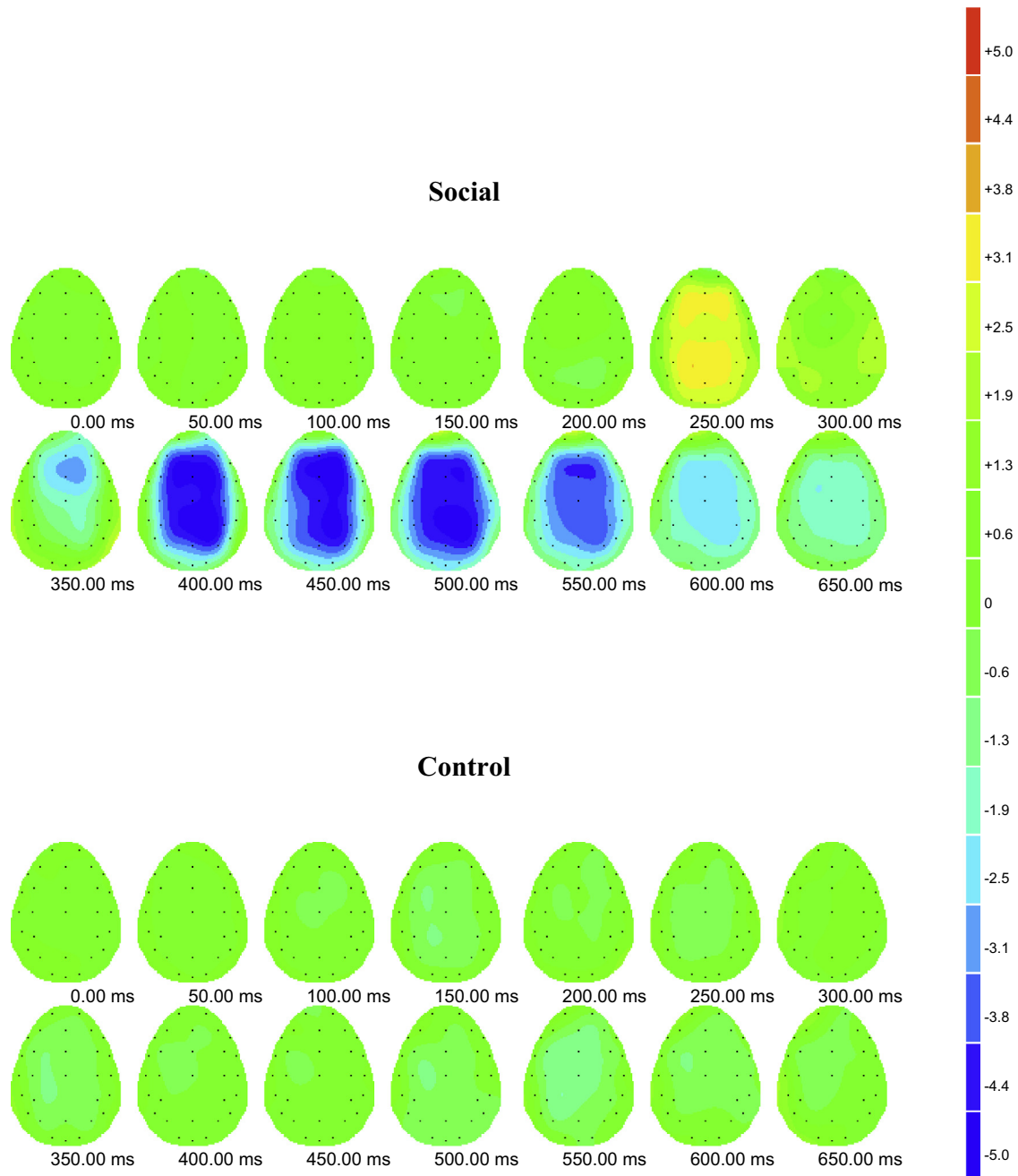


Fig. 5. Effect topographies (congruent–incongruent) by group condition.

observed in the ventral striatum when the virtual character responded congruently to a participant's joint attention bid, relative to trials where he responded incongruently. The authors argued that this reflected the hedonic response associated with achieving self-initiated joint attention, which may be the mechanism reinforcing future instances of interaction. Combining the results of Schilbach et al. with the current study suggests that neural processes that occur at around 350 ms may reflect the integration of gaze information from posterior-parietal regions to disambiguate their social relevance. This may be followed by concurrent or subsequent processing in subcortical structures (e.g. ventral striatum) where the hedonic valence of the gaze evaluation is assessed.

Given the goal-oriented task employed in the current study, it is possible that the observed P350 effect may have been partially

influenced by the different affective experiences elicited by congruent (i.e. hedonic) and incongruent (i.e. disappointed or frustrated) gaze shifts. However, this is unlikely given that in order to be 'happy' or 'frustrated' with their partner's response, participants first had to evaluate the social significance of their partner's gaze. Specifically, participants had to evaluate whether the gaze shift signalled the achievement of joint attention. As such, the P350 effect is likely to represent the neural processes responsible for discriminating gaze shifts that differ in their social outcome (e.g. success or failure in achieving joint attention). Secondly, the P350 effect observed in the social condition is unlikely to have been influenced by affective experience given that the congruent and incongruent stimuli in the control condition were also associated with task success and failure respectively and did not result in a P350 effect. To further separate the influence of social evaluation

and affect, future investigations may contrast ERP responses to gaze shifts which signal the success or failure to achieve joint attention in both collaborative and competitive contexts. Manipulating the interactive context in this way presents instances where a congruent gaze shift may be interpreted as either consistent (during collaboration) or inconsistent (during competition) with the participant's goals, which should in turn elicit different affective responses.

It is also important to recognise that whilst arrows are an obvious and well-established non-social substitute for gaze, these stimuli are ultimately perceptually different. Furthermore, they are not entirely equal in the effects they produce on spatial attention in cueing paradigms (see Frischen et al., 2007 for review). Future research may take even more conservative approaches to elucidate the social specificity of the P350 effect using the same task and stimuli, whilst reducing the social fidelity of the interactive context.

The current study is the first to investigate the time course of neural processes related to disambiguating another person's attentional perspective during joint attention interactions. Our findings suggest that a gaze shift is evaluated as signalling the success or failure to achieve a self-initiated joint attention bid around 350 ms after a gaze shift is observed. Importantly, the observed centro-parietal P350 effect was not replicated when a non-social analogue of the interactive paradigm was used in a second group of participants. This highlights the social specificity of the ERP effects identified in this study. Gaze scaffold our daily interactions by providing ongoing feedback about the attentional, emotional and mental perspective of those we interact with. Therefore, understanding how gaze is processed at the neural level is important as it guides research attempting to elucidate biomarkers of social impairment in autism populations, and provides an objective outcome measure for interventions targeting social communication.

References

- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8(6), 551–565. <http://dx.doi.org/10.1162/jocn.1996.8.6.551>.
- Bruner, J. S. (1974). From communication to language—A psychological perspective. *Cognition*, 3(3), 255–287. [http://dx.doi.org/10.1016/0010-0277\(74\)90012-2](http://dx.doi.org/10.1016/0010-0277(74)90012-2).
- Carrick, O. K., Thompson, J. C., Epling, J. A., & Puce, A. (2007). It's all in the eyes: Neural responses to socially significant gaze shifts. *Neuroreport*, 18(8), 763–766.
- Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *NeuroImage*, 108, 34–46. <http://dx.doi.org/10.1016/j.neuroimage.2014.12.041>.
- Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Social Psychology*, 41(3), 269–271. <http://dx.doi.org/10.2307/3033565>.
- Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions Royal Society London Biological Sciences*, 358, 315–324. <http://dx.doi.org/10.1098/rstb.2002.1199>.
- Conty, L., N'Diaye, K., Tijus, C., & George, N. (2007). When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing. *Neuropsychologia*, 45(13), 3024–3037. <http://dx.doi.org/10.1016/j.neuropsychologia.2007.05.017>.
- Feng, Q., & Zhang, X. (2014). Eye gaze triggers reflexive attention shifts: Evidence from lateralised ERPs. *Brain Research*, 1589, 37–44. <http://dx.doi.org/10.1016/j.brainres.2014.09.029>.
- Friesen, C., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490–495. <http://dx.doi.org/10.3758/bf03208827>.
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4), 694–724. <http://dx.doi.org/10.1037/0033-2909.133.4.694>.
- Gordon, I., Eilbott, J. A., Feldman, R., Pelphrey, K. A., & Vander Wyk, B. C. (2013). Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses. *Social Neuroscience*, 1–11. <http://dx.doi.org/10.1080/17470919.2013.832374>.
- Grice, S. J., Halit, H., Farroni, T., Baron-Cohen, S., Bolton, P., & Johnson, M. H. (2005). Neural correlates of eye-gaze detection in young children with autism. *Cortex*, 41(3), 342–353. [http://dx.doi.org/10.1016/S0010-9452\(08\)70271-5](http://dx.doi.org/10.1016/S0010-9452(08)70271-5).
- Hietanen, J. K., Leppänen, J. M., Nummenmaa, L., & Astikainen, P. (2008). Visuospatial attention shifts by gaze and arrow cues: An ERP study. *Brain Research*, 1215, 123–136. <http://dx.doi.org/10.1016/j.brainres.2008.03.091>.
- Hietanen, J. K., Nummenmaa, L., Nyman, M. J., Parkkola, R., & Hämäläinen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMRI study. *NeuroImage*, 33(1), 406–413. <http://dx.doi.org/10.1016/j.neuroimage.2006.06.048>.
- Holmes, A., Mogg, K., Garcia, L. M., & Bradley, B. P. (2010). Neural activity associated with attention orienting triggered by gaze cues: A study of lateralized ERPs. *Social Neuroscience*, 5(3), 285–295. <http://dx.doi.org/10.1080/17470910903422819>.
- Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews*, 33(6), 843–863. <http://dx.doi.org/10.1016/j.neubiorev.2009.02.004>.
- Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology*, 116(11), 2613–2631. <http://dx.doi.org/10.1016/j.clinph.2005.07.005>.
- Kimball, S., & Mattis, P. (1995). *GIMP image manipulation program (version 2.8.2)*. Berkeley.
- Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1), 78–100.
- Lassalle, A., & Itier, R. J. (2013). Fearful, surprised, happy, and angry facial expressions modulate gaze-oriented attention: Behavioral and ERP evidence. *Social Neuroscience*, 8(6), 583–600. <http://dx.doi.org/10.1080/17470919.2013.835750>.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Menon, R. S., & Kim, S. G. (1999). Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends in Cognitive Sciences*, 3(6), 207–216.
- Mundy, P., & Newell, L. (2007). Attention, joint attention and social cognition. *Current Directions in Psychological Science*, 16, 269–274.
- Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., et al. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, 101, 124–137. <http://dx.doi.org/10.1016/j.neuroimage.2014.06.061>.
- Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2010). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*. <http://dx.doi.org/10.1093/scan/nsq068>.
- Puce, A., Smith, A., & Allison, T. (2000). Erps evoked by viewing facial movements. *Cognitive Neuropsychology*, 17(1), 221–239.
- Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: An event-related potential study. *Journal of Cognitive Neuroscience*, 16(3), 415–426.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*.
- Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, 22(12), 2702–2715.
- Schweinberger, S. R., Kloth, N., & Jenkins, R. (2007). Are you looking at me? Neural correlates of gaze adaptation. *Neuroreport*, 18(7), 693–696. <http://dx.doi.org/10.1097/WNR.0b013e3280c1e2d2>.
- Shepherd, S. V. (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, 4(5).
- Singular Inversions (2008). *FaceGen modeller (version 3.3)* [computer software]. Toronto, ON: Singular Inversions.
- SR Research (2004). *Experiment builder (version 1.10.165)*. Ontario.
- van Velzen, J., & Eimer, M. (2003). Early posterior ERP components do not reflect the control of attentional shifts toward expected peripheral events. *Psychophysiology*, 40(5), 827–831.
- Watanabe, S., Miki, K., & Kakigi, R. (2002). Gaze direction affects face perception in humans. *Neuroscience Letters*, 325(3), 163–166.
- Williams, J. H. G., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *NeuroImage*, 25(1), 133–140. <http://dx.doi.org/10.1016/j.neuroimage.2004.10.047>.
- 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(1), 98–107. <http://dx.doi.org/10.1093/scan/nsq024>.
- Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS One*, 9(4), e94339. <http://dx.doi.org/10.1371/journal.pone.0094339>.