

A frontotemporoparietal network common to initiating and responding to joint attention bids



Nathan Caruana*, Jon Brock, Alexandra Woolgar

Department of Cognitive Science, Macquarie University, Sydney, Australia
ARC Centre of Excellence in Cognition and Its Disorders, Australia

ARTICLE INFO

Article history:

Accepted 9 December 2014

Available online 19 December 2014

Keywords:

Joint attention

fMRI

Eye tracking

Virtual reality

Social cognition

Interaction

ABSTRACT

Joint attention is a fundamental cognitive ability that supports daily interpersonal relationships and communication. The Parallel Distributed Processing model (PDPM) postulates that responding to (RJA) and initiating (IJA) joint attention are predominantly supported by posterior-parietal and frontal regions respectively. It also argues that these neural networks integrate during development, supporting the parallel processes of self- and other-attention representation during interactions. However, direct evidence for the PDPM is limited due to a lack of ecologically valid experimental paradigms that can capture both RJA and IJA. Building on existing interactive approaches, we developed a virtual reality paradigm where participants engaged in an online interaction to complete a cooperative task. By including tightly controlled baseline conditions to remove activity associated with non-social task demands, we were able to directly contrast the neural correlates of RJA and IJA to determine whether these processes are supported by common brain regions. Both RJA and IJA activated broad frontotemporoparietal networks. Critically, a conjunction analysis identified that a subset of these regions were common to both RJA and IJA. This right-lateralised network included the dorsal portion of the middle frontal gyrus (MFG), inferior frontal gyrus (IFG), middle temporal gyrus (MTG), precentral gyrus, posterior superior temporal sulcus (pSTS), temporoparietal junction (TPJ) and precuneus. Additional activation was observed in this network for IJA relative to RJA at MFG, IFG, TPJ and precuneus. This is the first imaging study to directly investigate the neural correlates common to RJA and IJA engagement, and thus support the assumption that a broad integrated network underlies the parallel aspects of both initiating and responding to joint attention.

Crown Copyright © 2014 Published by Elsevier Inc. All rights reserved.

Introduction

Joint attention – the ability to follow and direct another person's attention – is a critical aspect of interpersonal relationships and communication (Bruner, 1974; Mundy et al., 1990). Theoretical models suggest two functionally and developmentally distinct joint attention processes, which are each likely to be differentially represented in the brain; responding to joint attention (RJA) and initiating joint attention (IJA; Bruinsma et al., 2004). When an individual interprets the eye gaze of a social partner to determine their focus of attention, and then attends to the same thing, they are said to have *responded* to their partner's joint attention bid, achieving RJA. Individuals engage in IJA when they use their eye gaze to intentionally guide the attention of their social partner, thus *initiating* a bid for joint attention. In typical development, RJA emerges at around six months of age (Bakeman and Adamson, 1984; D. M. Bates, 2005; Scaife and Bruner, 1975), while IJA develops later, at approximately 12 months of age (Bates et al., 1979). Further

evidence for a dissociation comes from studies of autism. Autistic children typically exhibit RJA once their cognitive development is equivalent to approximately 30–36 months of age (Mundy et al., 1990). Contrastingly, IJA impairments often persist well into adolescence and adulthood (Dawson et al., 2004; Lord et al., 2000; MacDonald et al., 2006; Mundy and Jarrold, 2010; Mundy et al., 1990; Sigman & Ruskin, 1999).

According to the Parallel and Distributed-Processing model (PDPM) of joint attention (Mundy and Jarrold, 2010; Mundy and Newell, 2007; Mundy et al., 2009), RJA and IJA are executed within two partially independent yet parallel networks. Building on the work of Posner and colleagues (e.g., Posner & Rothbart, 2007), the model suggests that RJA depends on posterior and parietal regions which execute a range of attention-related functions. These functions include the processing of direction cues resulting in rapid and involuntary shifts of attention (supported by the precuneus, posterior parietal cortex, and occipital association cortex), eye gaze perception for attention modulation (intraparietal sulcus), and discrimination of gaze and head orientation (posterior superior temporal sulcus; pSTS). The PDPM purports that IJA exploits this posterior-parietal resource in addition to an anterior network involved in the suppression of automatic eye movements (frontal eye fields; superior colliculus pathway), and the execution of

* Corresponding author at: Level 3, 16 University Avenue, Macquarie University, NSW 2109, Australia.

E-mail addresses: nathan.caruana@mq.edu.au (N. Caruana), nathan.caruana@mq.edu.au (J. Brock), alexandra.woolgar@mq.edu.au (A. Woolgar).

goal-directed attention towards stimuli which signal rewarding experiences (anterior cingulate cortex; Mundy and Newell, 2007). This frontal network is also thought to include the medial prefrontal cortex (mPFC; Amodio and Frith, 2006; Williams et al., 2005), orbitofrontal cortex (OFC; Sabbagh, 2004) and anterior cingulate cortex (ACC; Redcay et al., 2010), which are believed to govern mental state attribution, self-referential thinking, and action monitoring processes. Importantly, the PDPM suggests that, although initially distinct, the neural substrates that support RJA and IJA become increasingly integrated throughout development (Mundy and Newell, 2007). The cognitive product of this neural integration is argued to be the emergence of complex joint attention behaviours, serviced by the cognitive ability to maintain parallel representation of self- and other-oriented attentional perspectives. Representing these perspectives simultaneously allows individuals to coordinate their interactive behaviour, enabling the dyad to align their attentional perspectives, and thus achieve joint attention.

Investigating the claims of the PDPM is challenging because joint attention is exclusively experienced during face-to-face social exchanges (Schilbach et al., 2013). However, two recent functional magnetic resonance imaging (fMRI) studies have risen to this challenge, using interactive joint attention tasks that can be performed during scanning. In the first of these, Schilbach et al. (2010) developed a virtual reality paradigm (see Wilms et al., 2010) in which participants interacted with an avatar who they believed was depicting the eye movements of a social partner outside the scanner. The avatar's responses were in fact controlled by a computer, using a gaze-contingent algorithm such that the avatar responded to the participant's gaze. On the screen participants were presented with three squares to the left, right and above the avatar's face. On RJA trials (referred to as OTHER_JA by Schilbach et al., 2010, p. 2702), participants were instructed either to look where the avatar looked or, in the control condition, to look at a different location. The contrast between these two conditions revealed differential activity in the ventral mPFC. This is consistent with previous gaze following and gaze congruency studies, and the idea that processing social gaze places additional demands on mentalising capacities (Amodio and Frith, 2006; Williams et al., 2005). However, this activation could also reflect differential gaze inhibition processes between the test and control conditions since the control condition involved executing a response that was incongruent to the gaze cue (Ishikawa and Raine, 2003; Simpson et al., 2001). These incongruent responses also resulted in a mismatch in the attentional perspective of the participant and their virtual partner. Accordingly, differential activation here may reflect the evaluation of self-other discrepancies in behaviour and attention. This is consistent with findings from studies employing action imitation-inhibition tasks (e.g., Brass et al., 2005).

On IJA trials (referred to as SELF_JA by Schilbach et al., 2010, p. 2702) one square would change from grey to blue. The participant had to fixate the square, and the avatar responded either by gazing in the congruent location or, in the control condition, by gazing at a different location. Congruent gaze was associated with increased activation of the anterior ventral striatum, argued to reflect reward neurocircuitry which reinforces IJA engagement. However, because, the initiating component was identical in the IJA and control conditions, this differential activity relates to the neural substrates involved when evaluating whether joint attention had been achieved, rather than the mechanisms involved in executing IJA itself. Additionally, the task employed was not intuitive or goal-driven, departing from ecological interactions where our current goals drive the need to coordinate our attention with others to share information.

In another pioneering study, Redcay et al. (2010, 2012) adopted a live video interaction paradigm in which the participant and experimenter (outside the scanner) each viewed a live video feed of each other's faces, whilst playing a cooperative game. Together they attempted to catch a mouse hidden behind one of four cheeses placed in the corners of the screen. On IJA trials, the participant saw a cue (a tail protruding behind one of the cheeses), saccaded towards the

location, and the experimenter followed their gaze to achieve joint attention. Compared to a "Solo Attention" condition in which the participant's task was identical but the experimenter's eyes remained closed, typically developed participants displayed activation of frontal/insular regions, including; inferior frontal gyrus (IFG), bilateral anterior operculum, medial superior frontal gyrus, left middle frontal gyrus, right precentral gyrus, and inferior parietal lobe. This could relate to the initiating component of the IJA task, however, as in the Schilbach et al. (2010) paradigm, it could equally reflect the activity associated with evaluating whether joint attention had been achieved.

On RJA trials, the roles were reversed. The experimenter saw the cue (mouse tail) and the participant followed their gaze. Relative to the Solo Attention condition, RJA was associated with pSTS, dorsal mPFC, and posterior cingulate activation. Again, it is unclear exactly what aspects of RJA this contrast reveals as the RJA condition involved gaze following whilst Solo Attention was a non-social visual search task. Furthermore, in this paradigm, as in that employed by (Schilbach et al., 2010), participants were overtly instructed as to their social role (initiator or responder). This made the interaction predictable and reduced the requirement for participants to monitor the attention of their social partner in order to interpret gaze cues as intentional bids for communication (Cary, 1978). Thus, these paradigms do not capture this 'attention monitoring' process, which is vital in achieving joint attention in ecological interactions.

The current study built on these innovative joint attention paradigms (Redcay et al., 2012; Schilbach et al., 2010) with a view to identifying the neural substrates that are common and distinct to RJA and IJA. Participants completed a virtual reality task in which they interacted with an avatar to catch a burglar that was hiding inside one of six houses displayed on the screen (Fig. 1). Whoever found the burglar had to guide the other to that location by first establishing mutual gaze and then moving their eyes to guide their partner in the appropriate direction. Thus, the role of the participant (initiator or responder) only became apparent throughout the course of each trial. Our paradigm thereby created a social context that (1) elicited intentional, goal-driven joint attention (2) naturally informed participant of their social role without overt instruction, and (3) required participants to monitor the attention of their social partner throughout the interaction in order to correctly interpret gaze cues. Neural activity in the RJA and IJA conditions were each contrasted with a corresponding non-social control condition matched on task complexity, number of eye movements elicited and attentional demands, so that RJA and IJA could be directly contrasted. By examining the conjunction of RJA and IJA effects, we were able to identify the neural correlates common to these joint attention functions. In accordance with the PDPM, we anticipated that RJA and IJA engagement would both result in the recruitment of frontotemporoparietal areas, but that a subset of this network would be common to both joint attention functions (Mundy and Newell, 2007; Redcay et al., 2010, 2012; Schilbach et al., 2010). Given that the cognitive processes of representing self- and other-attentional perspectives are common to both RJA and IJA social interactions, it was hypothesised that the neural correlates identified in this common network would include areas previously implicated in tasks where participants represent another's attentional perspective, such as TPJ, IFG and mPFC (Halko et al., 2009; Ramsey et al., 2013; Williams et al., 2005).

Method

Participants

Seventeen right-handed adults with normal vision and no history of neurological impairment participated in this study. Due to technical challenges, eye tracking calibration was successful for only 14 participants. Additionally, the fMRI data for one participant could not be normalised, resulting in a final sample of 13 participants (9 male, $M_{age} = 24.85$, $SD = 5.65$). We selectively recruited participants with dark

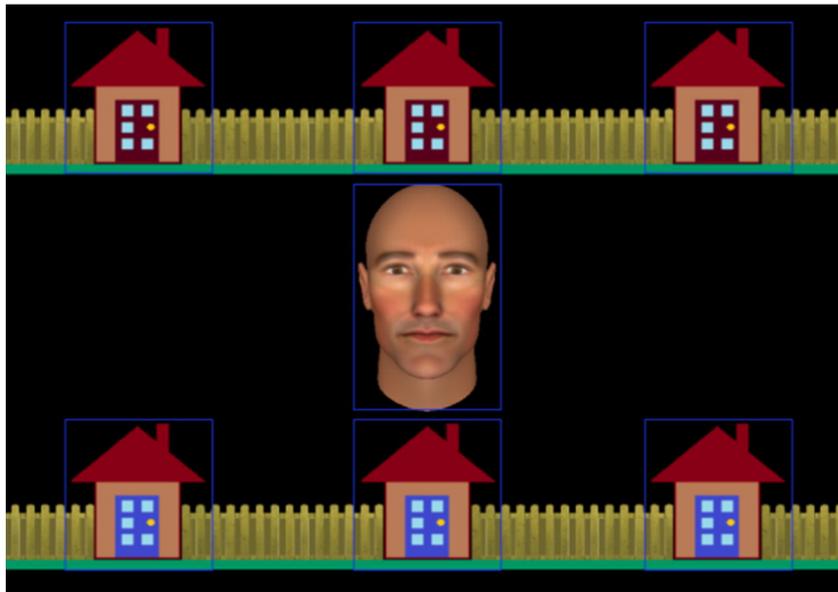


Fig. 1. Gaze areas of interest (GAOs) overlaid on participants' view of stimuli, represented as blue rectangles.

coloured irises as eyes with light irises tend to be difficult to calibrate in scanner environments (Gordon et al., 2013). Participants received payment for their time and provided consent before participating. The study was approved by the Macquarie University Human Research Ethics Committee.

Stimuli

An anthropomorphic avatar was generated using *FaceGen* (Singular Inversions, 2008). The avatar depicted a white Caucasian male, and subtended seven degrees of visual angle in the centre of the screen (Fig. 1). The avatar's gaze was manipulated to create eight images. The avatar's eyes were either directed at the participant, towards the top left of the screen, top-right, bottom-left, bottom-right, vertically upward or downward, or with eyes closed. Six houses, each subtending four degrees of visual angle, were arranged in two horizontal rows above and below the avatar. In each row, the houses were connected by fences which subtended two degrees of visual angle in height. Fences were included to make the task more concrete by providing a means by which the burglar could move between the houses without being seen. House and fence stimuli were created using *GIMP-2* (Kimball and Mattis, 1995).

The experiment was programmed using *Experiment Builder 1.10.165* (SR Research, 2004). Stimuli were presented on a projector and viewed through a mirror mounted on the head-coil.

Social task

We used a virtual reality paradigm which simulated live social interactions. Participants interacted with an on-screen avatar, whom they believed was being controlled by a second unseen person, named Alan, in a nearby eye tracking laboratory, via live infrared eye-tracking. However, the avatar was in fact programmed to respond contingently to the online recordings of participants' eye gaze using a novel gaze-contingent algorithm (see Fig. 2). Participants interacted with the avatar in a cooperative game, called 'Catch the Burglar'. The aim was to jointly locate and catch a burglar that was hiding behind one of six houses. To ensure that participants engaged in the task as naturally as possible, specific instructions about how participants should use their gaze were avoided. Instructions were presented on the stimulus screen at the beginning of the experiment in both the training and scanning sessions (see Supplementary resource 1 for full task instructions).

These were also read aloud by the experimenter at the beginning of the training session.

Search phase

The beginning of each trial consisted of a search phase, in which the participant and avatar would search their designated houses. The participant was always responsible for searching the houses with blue doors (e.g., the bottom row in Fig. 1), while the avatar was always responsible for searching the houses with red doors (e.g., the top row in Fig. 1). The blue doors appeared in the top or bottom row of houses, counterbalanced within participants across acquisition runs to prevent confounds driven by saccade trajectory, since downward saccades have previously been found to differentially recruit frontal regions (Tzelepi et al., 2010).

Participants conducted their search by looking at each house in any order they chose. When the participant fixated a house, the door opened to reveal that it was either empty or concealing the burglar (Fig. 3, first row). At the beginning of each trial, 0–2 of the participant's search houses (i.e., the houses with blue doors) were programmed to be already opened and empty. The number and location of already-opened houses was counterbalanced within each acquisition run. The purpose of this was to prevent participants from searching the houses in a systematic, left-to-right manner and so that the avatar could be programmed to search his houses in a random order without this behaviour appearing unusual. This was important because we wanted to prevent the avatar from appearing robotic or predictable. It enhanced the ecological appearance of the avatar's behaviour, and thus supported the deception regarding the interaction being with another person outside the scanner. This also reduced the likelihood of participants systematising their interaction with the avatar.

The avatar's search behaviour was fixed so that he only completed his search after the participant completed their search and fixated back on the avatar. This meant that participants were required to monitor the avatar's attention during their interaction, before responding or initiating. In this paradigm – as in ecological interactions – establishing mutual gaze was essential in determining whether the avatar was ready to guide the participant, or respond to the participant's initiation of joint attention. Our post-experimental inquiry revealed that participants did not detect this systematic delay in the avatar completing his search.

The onset latencies of the avatar's gaze behaviour (i.e., alternating between search houses, establishing mutual gaze, and executing responding or initiating saccades) were jittered with a uniform

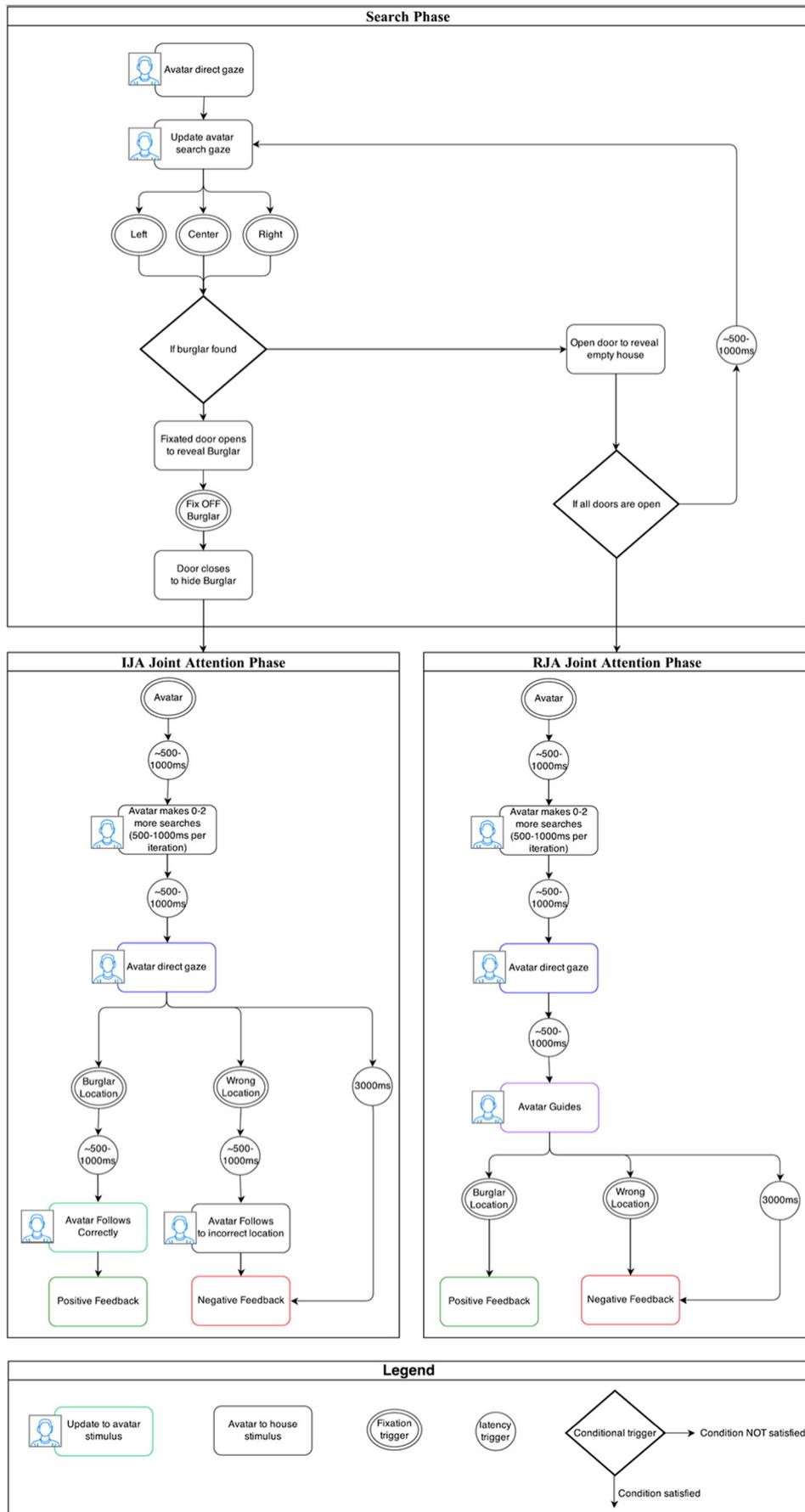


Fig. 2. Schematic summary of interactive algorithm. This algorithm was the same for the social and non-social conditions, apart from the central stimulus that is used (e.g., avatar direct gaze vs green fixation point). This diagram has been labelled to reflect the stimuli in the social conditions.

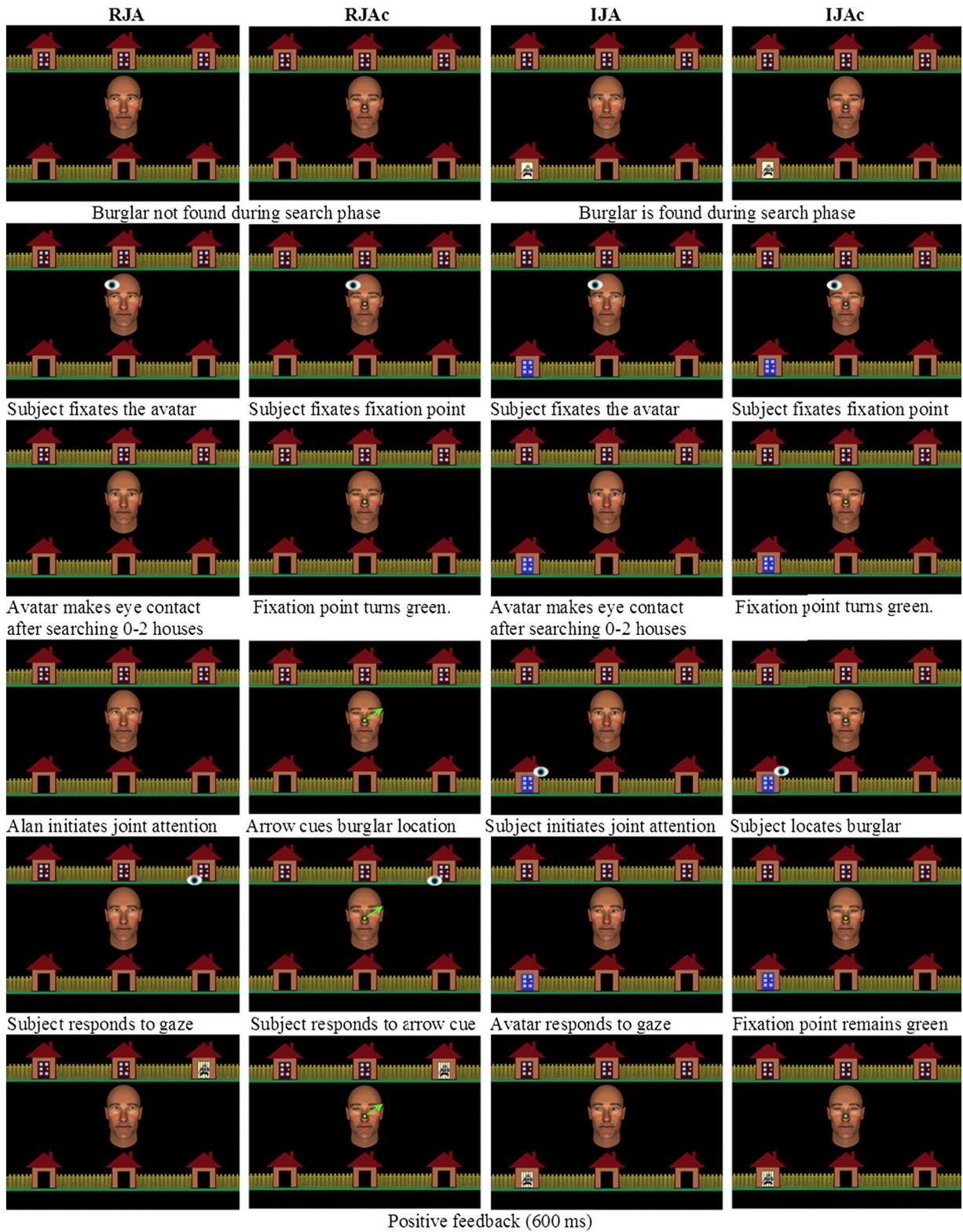


Fig. 3. Schematic representation of trial sequence by condition. Eye symbol represents the location of the participant's gaze and was not visible to the participant.

distribution between 500 and 1000 ms. This served to enhance the avatar's ecological appearance.

For RJA trials, where the burglar was “found” by the avatar, we pseudorandomised the location that the avatar searched last before directing gaze at the participant. This was implemented to ensure that the location searched last was not predictive of the location of the burglar, so that participants could not predict the burglar location before

the avatar made a guiding saccade. This was important because a premature cue to the burglar location could potentially reduce the time required by participants to process the avatar's guiding gaze. Whilst one might expect a social partner to terminate their search immediately upon finding the burglar, we found that participants rarely did this, and so this aspect of the avatar's behaviour again mimicked that of real participants.

RJA

On RJA trials, the search phase ended when the participant opened all of their designated houses, and found them to be empty (Fig. 3, first row, first column). This meant that the burglar was located in one of the avatar's search houses. The participant would then wait for the avatar to complete his search and establish mutual gaze. Once the participant fixated the avatar's face, the avatar searched 0–2 more houses and then directed his gaze towards the participant, establishing mutual gaze. Then, provided the participant was still fixating the avatar, the avatar would guide the participant to the correct location by directing his gaze there. The participant was then required to make an RJA saccade to fixate the appropriate house.

IJA

On IJA trials, the search phase ended when the participant found the burglar in one of their allocated houses (Fig. 3, first row, third column). Once the participant fixated away from the burglar, the door would close again to conceal it. This was to ensure that 'guiding' IJA saccades and 'following' RJA saccades were always towards the same visual stimulus (i.e., a closed door). After finding the burglar, the participant was then required to fixate the avatar in order to establish mutual gaze. As in the RJA condition, the avatar face was updated so that his gaze was averted between 0 and 2 more times, depicting a continuation of his search, and then updated again so that his gaze was directed at the participant. When mutual gaze was established, the participant was required to make an IJA saccade from the avatar to the burglar's location. Once the participant fixated one of their designated houses, the avatar responded by gazing toward the same house to achieve joint attention. The avatar was programmed to follow the participant's gaze to whichever house the participant fixated. This meant that the avatar would also follow the participant's gaze towards an empty house if the participant guided him there.

Feedback

Positive feedback (the burglar depicted behind bars, Fig. 3, last row) was given when participants succeeded in achieving joint attention at the burglar location. Negative feedback was displayed if participants failed to achieve joint attention at the correct location when responding to the avatar's guiding gaze (RJA) or when guiding the avatar to the burglar's location (IJA). Here, the burglar appeared in red at its true location. This also occurred if participants failed to (1) establish mutual gaze within three seconds of completing their search, or (2) fixate the burglar location within three seconds of establishing mutual gaze on IJA trials, or after being guided on RJA trials. Additionally, if participants did not begin searching their allocated houses within three seconds of the trial commencing, red text, reading "Failed Search" appeared on the screen. The two words were displayed to the left and right side of the avatar face so as to not occlude the visual stimulus.

Control task

To control for non-social aspects of each joint attention task, we developed responding (RJAc; Fig. 3, second column) and initiating (IJAc; Fig. 3, fourth column) control conditions. These conditions were designed to control for task complexity, number of eye movements required, and attentional demands of the RJA and IJA tasks. On control trials, participants were instructed to catch the burglar "on their own". Participants were told that during these trials, Alan was also completing the task alone. As in the social trials, participants were only required to search their designated houses. The control conditions proceeded identically to their counterparts in the social task, with the following differences to the task stimuli: (1) the avatar's eyes remained closed for the duration of the trial, (2) a small grey fixation point, subtending one degree of visual angle, was overlaid on the avatar's face and was visible until the participant completed their search and fixated it, (3) the grey

fixation point turned green when fixated, (instead of the avatar establishing mutual gaze), to signal the end of the search phase, and (4) in RJAc, the presentation of a green arrow, subtending three degrees of visual angle, cued the burglar's location. Example trials from each of the four conditions are depicted in Video 1.

Procedure

Training

Participants attended a training session before they were tested in the scanner. A deception induction was conducted, where participants were told that they would be interacting with 'Alan', the experimenter's colleague. The experimenter explained that Alan would be interacting with them from Lab 1 while they completed the experiment in Lab 2 (during training) as well as when they were in the scanner, via a high-speed network connection. To reinforce the deception, the participant was given a tour of the two adjacent eye tracking laboratories, which were identical in their set-up.

The training session consisted of 29 trials per condition (116 in total). The session began with a block of social (RJA, IJA) or control (RJAc, IJAc) trials, counterbalanced across participants to prevent any order confounds which may have affected the deception manipulation. At the beginning of each block of trials, the experimenter would ask the participant if they were ready to begin, then a screen appeared that read, "Initialising interface... Both participants ready!... Start!" with a three second lag between each string of text to simulate the interface 'loading' and waiting for Alan.

Scanning

In total, participants completed four scanning runs, each consisting of 108 trials. Each run consisted of 27 trials of each condition; RJA, RJAc, IJA, and IJAc. A random permutation was used to pseudorandomise condition order within runs. Specifically, social and control trials were organised into alternating blocks of six trials, with responding and initiating trial types randomised within blocks. The randomisation within blocks was constrained to ensure that each block contained three responding and three initiating trials. Each block began with a 1000 ms cue, in which white text on a black panel appeared over the avatar's eyes, reading "Together" to indicate the onset of a social block, or "Alone" for a control block.

We used short blocks of six trials each to separate social and control events. This provided a compromise between a fully blocked design which would have enhanced the continuity of the interaction, and a design in which events were fully intermixed which would have reduced the temporal separation between social and control events (Henson, 2006). Prior to data acquisition, we employed Henson's (2012) algorithm to confirm that our design did not compromise the efficiency of our GLM when compared to a fully intermixed design.

Within each run, each condition (RJA, RJAc, IJA, IJAc) was matched on the (1) burglar location, (2) number of houses to-be-searched at the beginning of each trial, (3) location of search houses, and (4) the number of eye movements made by the avatar before returning the participant's bid for mutual gaze. Trial order was counterbalanced across scanning runs, and run order was counterbalanced across participants.

Post-experimental debrief

Following acquisition, participants rated the social and control tasks for difficulty, naturalness, intuitiveness and pleasantness on a 5-point Likert scale. For the social conditions only, participants rated how cooperative they thought Alan was. Participants were debriefed about the true nature of the social interaction. They were told that they were not interacting with 'Alan' but a computer programmed avatar. Participants then rated how convinced they had been that Alan was a real person. Participants also provided ratings on their perception of the accuracy of the virtual interface and eye-tracking set-up, and the difficulty

in switching between the social and control tasks. Finally, they indicated whether they preferred working with Alan, or on their own.

Interactive eye tracking

Eye-movements from the right eye were tracked with a sampling rate of 1000 Hz using an Eyelink 1000 Remote Eye-Tracking System (SR Research Ltd., Ontario, Canada). A desktop-mounted tracker and chinrest were used during training. For the scanning session, we used an MRI compatible tracker mounted behind the head-coil, and reflected into a head-coil-mounted mirror. A standard 9-point camera calibration and validation was conducted at the beginning of each acquisition run.

The stimulus screen was divided into seven gaze-related areas of interest (GAOI), one for each of the six houses, and the avatar (Fig. 1). These GAOIs were used to monitor participants' gaze online, so that the avatar's behaviour could be adapted accordingly by our gaze-contingent algorithm.

Acquisition

fMRI data was collected on a Siemens 3 T Verio scanner with a 32-channel head-coil (Siemens Medical Solutions) located at Macquarie Medical Imaging, Macquarie University Hospital. Whole-brain functional images were acquired using sequential descending T2*-weighted echo-planar imaging (EPI) with the following parameters: TR = 2000 ms; TE = 30 ms; flip angle = 78°; FOV = 191 mm; image matrix = 64²; voxel size = 3.0 × 3.0 × 3.75 mm; 32 oblique axial slices. Given the self-paced nature of the task, the number of EPIs acquired per run varied between 272 and 370. T1-weighted MPRAGE structural images were also acquired for each participant at the beginning of the scanning session (FOV = 256 mm; voxel size = 1.0 × 1.0 × 1.5 mm, 160 slices).

Preprocessing

SPM5 (Wellcome Department of Cognitive Neurology, London, U.K.; <http://imaging.mrc-cbu.cam.ac.uk/imaging/>) was used for all fMRI data preprocessing. After spatial realignment and slice-time correction, each participant's structural image was coregistered to the mean of their functional volumes (EPIs). Structural images were segmented and normalised to an MNI template (Montreal Neurological Institute). EPI images were normalised using the parameters derived from normalising the structural, and smoothed using an 8 mm full width at half maximum (FWHM) Gaussian kernel. All data were high-pass filtered (128 s). We fitted a general linear model (GLM) to the data for each run, with four regressors (RJA, IJA, RJAc, IJAc) plus an additional six movement regressors and the mean activation for each acquisition run. Our analyses focused on the 'joint attention phase' of each trial (see Fig. 2). Accordingly, event onset times were defined as the time at which the participant opened the last empty house (RJA and RJAc) or found the burglar (IJA and IJAc). Events were modelled as box cars lasting until the time at which joint attention was achieved and the burglar caught. This assisted in accounting for variation in reaction times between trials (Grinband et al., 2008; Henson, 2006; Vogeley et al., 2004; Woolgar et al., 2013).

Analyses

Some trials were excluded from the analysis of BOLD data due to calibration failure, participant error, or eye movement anomalies. This resulted in an imbalance in trial count per condition, which we redressed by removing the subsequent contra-condition trials. The BOLD data was analysed both with and without matching the number of trials in each condition after trial rejection. We have reported the results on the balanced data as this is the most conservative approach, although the pattern of results was the same for both analyses. For each participant, we performed four contrasts: (1) RJA minus RJAc, to isolate activity associated with the social RJA task; (2) IJA minus IJAc

to isolate activity associated with the social IJA task, (3) IJA minus RJAc after first removing activity due to the respective control conditions (i.e., (IJA – IJAc) – (RJA – RJAc)), to isolate any additional activity present for IJA over and above RJA and (4) RJA minus IJA after first removing activity due to the respective control conditions (i.e., (RJA – RJAc) – (IJA – IJAc)), to isolate any additional activity present for RJA over and above IJA. For each participant, the resulting contrasts were then entered into second level random effects analyses. We then performed a conjunction analysis of contrasts (1) and (2) to examine whether any voxels were active during both RJA and IJA engagement. All second level t-images were corrected for multiple comparisons using a false discovery rate (FDR) of $p < 0.05$. Given that the threshold assigned by an FDR correction varies across tested contrasts, all second level t-images were thresholded at $t > 3.70$. This was the threshold applied in the FDR correction for contrast (1) – RJA minus RJAc – and was the most conservative threshold applied to any of the tested contrasts. The use of a single threshold for visualisation allowed the results to be more easily compared across contrasts. The SPM5 anatomy toolbox (Eickhoff et al., 2007) was used to assign anatomical labels to the functional results of the second level analysis. We additionally examined this by comparing anatomical labels to Brodmann and AAL templates in MRICroN.

Results

Post-experimental inquiry

Table 1 details participants' ratings of experience during the task. Participants rated both the social and control tasks as easy, natural, intuitive and pleasant. There were no significant differences in subjective experiences between the social and control tasks in terms of difficulty, naturalness, intuitiveness, or pleasantness (paired t-tests, all $ps > 0.39$).

Participants reported that they did not find it difficult to switch between the social and control versions of the task ($M = 1.46$, $SD = 0.88$) and that the eye tracking and virtual interface was an accurate tool for capturing their eye movements, allowing them to effectively interact with Alan ($M = 4.62$, $SD = 0.87$).

Prior to debrief, Alan was rated as being highly cooperative ($M = 4.77$, $SD = 0.44$), and no participant claimed that they did not believe that Alan was a real person. Once the participants were debriefed, the majority (8/13) provided ratings which demonstrated that they were completely convinced that Alan was a real person ($M = 3.85$, $SD = 1.46$). However, three individuals did provide ratings lower than three. When probed in a follow-up question; "What aspects of the interaction made you more/less convinced?" these participants explained that they momentarily questioned or entertained the possibility that they might not be interacting with another human, but that they did not dwell on this thought. These ratings may be subject to report biases associated with the desire to not appear gullible. Two of these individuals reported that they questioned Alan's existence because they had been previously deceived in similar psychology experiments. The other individual explained that he thought it "unnecessary to get someone else to do the task when you could get a computer to do it."

Accuracy

Participants could fail a trial of the burglar task if they took more than three seconds to (1) begin searching their houses, (2) guide Alan once mutual gaze had been established on IJA trials or, (3) respond to Alan's guiding gaze on RJA trials. Participants could also fail the trial by initiating or responding to the incorrect location. All participants had over 90% accuracy across all trials. Performance was well matched on IJA ($M = 99.43\%$, $SD = 1.24$) and IJAc trials ($M = 99.80\%$, $SD = 0.31$; $t = -1.27$, $p = 0.229$), however participants made significantly more errors on RJA trials ($M = 96.33\%$, $SD = 3.26$) compared to RJAc ($M =$

98.30%, $SD = 1.35$; $t = -2.332$, $p = 0.038$). For neural analyses trial numbers were equated across conditions (see Analyses).

Target-bound saccade onset latency

The target-bound saccade onset latency was measured to investigate the effect of social context on the time taken to perform communicative eye movements. This was the time it took for participants to execute a saccade towards the burglar location, resulting in joint attention. On responding trials this was defined as the first saccade after the avatar (RJA) or the arrow (RJAc) indicated the burglar location. For initiating trials, it was the first saccade towards the burglar location after mutual gaze had been established (IJA) or the fixation point turned green (IJAc).

A two-way repeated measures ANOVA revealed a significant interaction effect of social context (social vs. control) and social role (responding vs initiating), $F(12) = 14.03$, $p = 0.003$, indicating that the effect of social role was different for initiating and responding trials. Post-hoc paired t -tests were therefore conducted to explore the interaction. There was no significant difference in saccadic reaction time between social and control initiating trials (IJA: $M = 460.96$, $SD = 117.36$; IJAc: $M = 439.53$, $SD = 108.38$; $t(12) = 0.70$, $p = 0.50$). However, participants were significantly slower to execute a saccade in response to Alan's guiding gaze (RJA; $M = 533.87$, $SD = 156.28$) compared to the arrow cue (RJAc; $M = 312.94$, $SD = 58.47$; $t(12) = 5.86$, $p < 0.005$). One possibility is that processing times were increased in the social context in responding but not initiating trials due to the ambiguity of gaze cues. On social responding (RJA) trials, the avatar's gaze is updated during the search phase, thus presenting gaze information that does *not* indicate the burglar's location. Participants must integrate the ostensive information of direct gaze to disambiguate whether the avatar's averted gaze is intentionally guiding their attention. Contrastingly, the arrow cue provided on control responding (RJAc) trials is unambiguous, perhaps demanding less processing time. The neural correlates observed for responding to joint attention (see [Neural correlates](#)), may in part also reflect this disambiguating process which is central to monitoring the attention of a social partner.

Saccade count

To ensure that differences in neural activation could not be driven by differences in eye movements between conditions, we measured the number of saccades participants made between GAOs on each trial ([Fig. 1](#)). There was no significant difference between RJA ($M = 4.27$, $SD = 0.22$) and RJAc trials ($M = 4.31$, $SD = 0.31$), $t(12) = -0.68$, $p = 0.51$. However, the saccade count was significantly higher for IJA ($M = 6.25$, $SD = 0.76$) than IJAc ($M = 5.45$, $SD = 0.32$), $t(12) = 5.10$, $p < 0.05$.

Further analysis of the eye tracking record revealed that on some IJA trials participants guided the avatar prematurely, before establishing mutual gaze. This was followed by a "double-take" saccade back to the avatar for a second initiation attempt. To identify these trials, a narrow temporal interest period was defined between the time that the participant fixated the avatar after finding the burglar, and the time that mutual gaze was established. If the participant fixated the burglar location within this interest period, the trial was identified as including a premature saccade, and was excluded from further analysis.

There were significantly more premature trials for IJA ($M = 30.00$, $SD = 19.18$) than IJAc ($M = 2.92$, $SD = 3.55$, $t(12) = 5.57$, $p < 0.05$). To balance the amount of trials in the contrasted conditions, an algorithm was employed to remove the next-occurring correct trial from the contrasted condition. This included the removal of additional trials from the RJA ($M = 24.23$, $SD = 16.39$), RJAc ($M = 25.93$, $SD = 18.02$) and IJAc ($M = 25.92$, $SD = 16.51$) conditions. The algorithm accounted for the number of trials that had already been excluded in each condition due to errors. This included trials where participants took longer than four seconds to begin their search ($M = 1.90$, $SD = 1.60$), took

longer than three seconds to execute an initiating or responding saccade ($M = 4.57$, $SD = 5.13$), or fixated an incorrect location after making an initiating or responding saccade ($M = 7.25$, $SD = 4.11$). On average a total of 117.77 trials ($SD = 72.18$) were excluded across all four conditions.

After the number of trials were equated across conditions for each participant, there were no significant differences in the number of eye movements made during RJA ($M = 4.26$, $SD = 0.23$) and RJAc ($M = 4.30$, $SD = 0.31$), $t(12) = -0.41$, $p = 0.69$ or between IJA ($M = 5.34$, $SD = 0.26$) and IJAc ($M = 5.36$, $SD = 0.24$), $t(12) = -0.53$, $p = 0.61$.

Neural correlates

The central aim of this study was to investigate the neural correlates of responding to and initiating joint attention, over and above activation associated with non-social processes involved in typical joint attention tasks. We used experimental control conditions that were specifically matched to each of the social responding and initiating tasks. This allowed us to examine activation relating to responding to and initiating joint attention bids separately, and to investigate the neural substrates that are common and different across these distinct joint attention functions. The neural correlates for each contrast are reported in [Table 2a–d](#), with corresponding contrast maps displayed in [Figs. 4a–d](#).

First we investigated the neural correlates of responding to a joint attention bid, controlling for non-social task-related activity (RJA – RJAc). Responding to joint attention in a social context recruited lateral portions of the middle frontal gyrus (MFG) extending to the right dmPFC as well as the left insula, right middle temporal gyrus (MTG), bilateral pSTS, bilateral supramarginal gyrus (temporoparietal junction; TPJ), right precuneus and bilateral amygdala (see [Table 2a](#), [Fig. 4a](#)).

Next we examined the neural correlates of intentionally initiating joint attention after controlling for non-social task-related processing (IJA – IJAc). Here, initiating joint attention resulted in bilateral recruitment of anterior portions of MFG, extending to dmPFC, as well as right IFG, bilateral ACC and MCC, right inferior temporal gyrus, left MTG, bilateral pSTS, PCC, cuneus, calcarine gyrus, left lingual gyrus, right thalamus and left cerebellum (see [Table 2b](#), [Fig. 4b](#)).

Next, we asked whether any of the areas activated for RJA and IJA were common across the two tasks. Accordingly we performed a conjunction analysis of the (RJA – RJAc) and (IJA – IJAc). This revealed that RJA and IJA recruited common substrates within a right lateralised frontotemporoparietal network. This included MFG, IFG, MTG, pSTS, TPJ, precentral gyrus and bilateral precuneus (see [Table 2c](#), [Fig. 4c](#)).

Finally, we examined differences in activation between initiating and responding to joint attention. When directly contrasted with RJA, IJA engagement resulted in increased activation across frontotemporoparietal regions after controlling for non-social task demands; (IJA – IJAc) – (RJA – RJAc). This included right MFG, IFG, superior frontal gyrus, bilateral SMA, left precentral gyrus, bilateral MCC, right inferior temporal gyrus, left MTG, rTPJ, left temporal pole, bilateral precuneus, calcarine gyrus, right thalamus and cerebellum (see [Table 2d](#), [Fig. 4d](#)). No voxels survived FDR correction when the inverse contrast was tested (RJA – RJAc) – (IJA – IJAc) indicating that responding to joint attention does not result in more activation than initiating joint attention, after controlling for non-social task-related activity. This contrast resulted in a small cluster of activation in the left hemisphere, including the precentral gyrus, when assessed with a more liberal threshold ($p < 0.005$) that was not corrected for multiple comparisons (included for completeness, see [Lieberman and Cunningham, 2009](#)). Also see Supplementary resource 2 for the full results of this analysis.

Discussion

Extending on previous interactive studies of joint attention, we developed a novel virtual reality paradigm which balanced the need

for a dynamic interactive environment, whilst maintaining full experimental control. Our task was intuitive, goal-directed, and established a context that naturally informed the participant of their social role during the interaction. We also developed closely matched control conditions to account for non-social task-related activity so that RJA- and IJA-related activity could be examined directly. This allowed us to examine whether common neural substrates underlie RJA and IJA engagement in the adult brain. Our data provides support for the PDPM claim that, in adulthood, the neural substrates supporting these developmentally distinct joint attention functions are integrated within a common neural network. The anatomical nature of this network is consistent with the idea that complex joint attention behaviours rely on the parallel processing of self- and other-oriented visual attention (Marchetti and Koster, 2014; Mundy et al., 2009).

The PDPM claims that RJA function emerges from a posterior-parietal network developing in the first six months of life, whilst IJA utilises these regions in conjunction with later-developing anterior areas including frontal eye fields, prefrontal association area, ACC, and the orbital prefrontal association cortex (Mundy and Newell, 2007). The model further claims that joint attention is defined by a “cognitive synthesis” in which there is an integrated processing of the visual attention of the individual themselves, and that of the person that they are interacting with (pp. 7; Mundy et al., 2009). The model also emphasises that these processes, although different by definition, and in their developmental onset, may depend on common cognitive and neural substrates.

To investigate this overlap, we observed the conjunction of our (RJA – RJAc) and (IJA – IJAc) contrasts so as to determine a network common to RJA and IJA engagement. We found evidence for a right lateralised frontotemporoparietal network activated for both initiating and responding to joint attention. This consisted of TPJ, precuneus, IFG, pSTS, MFG and MTG. These regions have been previously correlated with cognitive processes related to the achievement of self-other representations, although the social specificity of these regions remain uncertain. For instance, whilst TPJ has been implicated in tasks where participants must form representations of another's mental state (Samson et al., 2004), it has also been implicated during non-social stimulus-driven shifts of attention, particularly when the stimulus is relevant to the task at hand (Kincade et al., 2005). Our social task unavoidably loads on both of these cognitive processes, however our control tasks – which evokes similar stimulus-driven shifts of attention – does not require the representation of another's perspective. As such the increased involvement of TPJ in the social version of the task suggests that this region may be particularly engaged in social contexts when the task requires representing another's focus of attention, or one's own attentional state as it is relevant to others. This is corroborated by previous findings of temporoparietal modulation, when individuals evaluate their own visual perspective relative to an avatar's (Ramsey et al., 2013). Previous studies have presented inconsistent accounts of the role of TPJ in joint attention. Whilst Redcay et al. (2012) report TPJ activation when contrasting IJA with baseline task engagement, Schilbach et al. (2010) reported increased activation in rTPJ during trials when joint attention was not achieved (NO_IJA trials), relative to joint attention trials. In the latter study, understanding the role of TPJ is further complicated as this contrast was collapsed across RJA and IJA conditions,

making it difficult to determine whether this is an effect specific or common to RJA and IJA engagement. The absence of a non-social baseline condition also makes it difficult to determine whether TPJ involvement is sensitive to the social aspects of the task. Contrastingly, the current study is the first to specifically associate TPJ with the social aspects of both RJA and IJA.

Like TPJ, the precuneus, IFG and pSTS have also been implicated in tasks which involve various self- and other-oriented representations, from visual perspective taking to evaluating the intentionality of actions. Specifically, the precuneus has been recruited in tasks which involve representing the beliefs of others (Saxe et al., 2006). Vogeley et al. (2004) also found precuneus activation to be common to tasks involving the representation of first person (self) and third person (other) visual perspectives, with increased activation for self over other representations. Our data suggests that this involvement of the precuneus generalises to social interactions where the need to represent self and other attention perspectives is less explicit. In our social task, participants had to represent the attentional focus of their partner to determine when they could respond to or initiate joint attention. They also had to represent their own attentional focus so as to plan guiding saccades during IJA trials, and to shift their attentional focus when responding during RJA trials.

The involvement of IFG has been reported in tasks involving self- and other-oriented perspective representations, including the ultimatum game (Halko et al., 2009). These tasks involve a dyadic interaction where one individual proposes how a reward can be divided. If their partner accepts the reward is divided accordingly, otherwise neither player receives any reward. These profit-oriented decisions intrinsically involve representations of the potential gains of the decision for the self and for the other, in order for the outcome of the economic decision to be evaluated. Thus, the ultimatum game, like joint attention interactions, provides a context in which self and other perspectives must be considered simultaneously. Interestingly however, this is the first joint attention study to associate IFG activation with RJA engagement. Previously, IFG has only been correlated with IJA execution (Redcay et al., 2012). The identification of this region for both RJA and IJA in the current study is potentially the result of the superior ecological validity and experimental control afforded by our paradigm. Specifically, the fact that participants were not instructed on the social role that they or their partner would play on each trial, is likely to have presented an increased demand on perspective-taking processes.

The pSTS is well known for its involvement in gaze processing, however this has been found to be modulated in tasks where the participant must determine the intentionality of another's behaviour (Morris et al., 2008; Pelphrey et al., 2004; Saxe et al., 2004). This form of representing another's perspective is relevant in both our RJA and IJA tasks, as participants were required to integrate the ostensive direct gaze of the avatar in order to (1) differentiate between gaze that was intentionally guiding, as opposed to searching (RJA), and (2) to determine Alan's readiness to be guided (IJA). However, because the avatar's eyes remain closed in both control conditions, we cannot determine whether the involvement of pSTS in this contrast reflects processes beyond gaze discrimination.

Consistent with previous interactive joint attention studies, our social task also recruited the MFG (Redcay et al., 2012; Schilbach et al., 2010) and MTG (Redcay et al., 2012). These regions have been implicated in ocular tracking tasks where participants orient attention in the absence (MFG) and presence (MTG) of eye movements (Ohlendorf et al., 2007). Although our RJAc and IJAc tasks controlled for many aspects of cognition, they necessarily differed in that they did not involve changes to the avatar's gaze. This manipulation was crucial to the establishment of a social and non-social condition as gaze is unavoidably a social stimulus. However, it means that in our data we cannot be sure whether the activation of these regions reflects the social nature of the task, or the processing of the avatar's eye gaze per se, or both. As such, further work is needed to address the social specificity of these substrates.

Table 1
Post experimental inquiry ratings of subjective experience.

Task aspect	Social	Control
	M (SD)	M (SD)
Difficulty	1.31 (0.48)	1.31 (0.63)
Naturalness	3.92 (1.26)	3.92 (1.26)
Intuitiveness	4.77 (0.60)	4.54 (0.78)
Pleasantness	4.54 (0.66)	4.46 (0.97)

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

Table 2
Neural correlates.

	(a) Responding to joint attention (RJA – RJAc)							(b) Initiating joint attention (IJA – IJAc)							(c) Conjunction of initiating and responding to joint attention (IJA – IJAc) with (RJA – RJAc)							(d) Initiating joint attention minus responding to joint attention (IJA – IJAc) – (RJA – RJAc)						
	H	BA	x	y	z	k	T	H	BA	x	y	z	k	T	H	BA	x	y	z	k	T	H	BA	x	y	z	k	T
<i>Frontal</i>																												
Middle frontal gyrus	R	47	34	52	0	51	4.32	R	46	38	52	0	35	5.23	R	6	44	6	54	220	4.83	R	46	36	54	16	21	4.72
	R	46	24	52	26	3075	9.74	R	46	38	36	34	190	4.92	R	6	38	0	56	208	6.73	R	6	38	0	56	208	6.73
	L	46	-26	44	14	44	4.36	R	46	34	54	16	57	4.77	L	6	-28	2	60	301	5.12	L	6	-28	2	60	301	5.12
							L	6	-26	4	58	2387	10.56															
							L	10	-40	56	4	62	5.72															
Superior medial gyrus	R	46	3	42	36	3075	9.10	R		8	44	48	2387	8.49								R	6	6	36	50	12	3.96
Insula	L	47	-34	22	2	96	5.26	R	45	42	26	-6	455	5.05								R	45	48	36	8	10	4.63
Inferior frontal gyrus	R	44	48	22	18	3075	5.55	R	44	54	22	26	16	4.13	R	44	54	22	26	16	4.13	R	45	48	36	8	10	4.63
Superior frontal gyrus ^a	R	45	46	24	8	455	6.73	R	44	54	22	26	16	4.13	L	6	-22	6	60	2387	8.98	B	6	20	6	58	12	4.03
																						R	6	18	12	68	58	8.22
<i>SMA</i>																												
Precentral gyrus																												
Anterior cingulate cortex																												
								B	32	4	40	12	27	4.84								R		12	-32	46	11	4.06
Middle cingulate cortex																												
								R		10	18	32	27	4.45								L	23	-4	-8	32	24	5.69
								L		-6	-6	32	32	7.53								L	6	-36	-4	50	301	5.65
Precentral gyrus	R	44	42	8	44	3075	5.55	R	44	40	6	48	2387	6.76	R	44	38	4	46	220	4.70	L	6	-36	-4	50	301	5.65
								L	6	-30	-2	58	2387	8.07														
<i>Temporal</i>																												
Inferior temporal gyrus																												
Middle temporal gyrus ^a	R	20	52	-22	-14	52	5.04	R	37	40	-52	-14	27	4.50								R	37	56	-58	-4	1471	6.61
								R		48	-72	-2	2330	7.73	R	22	58	-46	10	932	6.06	L	22	-58	-50	20	36	4.68
								L	21	-46	-44	6	18	4.55								L	21	-46	-44	6	18	4.72
Superior temporal gyrus	R	42	52	-44	24	1288	8.08	R	42	56	-42	22	2330	8.98	R	42	54	-44	16	932	6.06							
	L	22	-60	-48	14	169	7.50	L	22	-52	-46	18	204	6.68														
Temporoparietal junction	R	42	52	-44	24	1288	8.08	R		56	-40	40	2330	6.56	R	40	48	-44	42	10	3.79	R	40	54	-34	40	16	4.34
	L	40	-52	-48	34	56	4.85															L	38	-30	6	-26	14	4.43
<i>Temporal pole</i>																												
Fusiform gyrus																												
								R	V4	38	-60	-12	27	4.45														
<i>Parietal</i>																												
Posterior cingulate cortex																												
Precuneus	B	7	4	-62	58	355	5.18	B	29	8	-40	14	36	5.61	B	5	2	-52	54	93	4.06	B		10	-54	46	80	5.48
								B		10	-62	51	5137	8.15														
<i>Occipital</i>																												
Cuneus																												
Calcarine gyrus	B	18	4	-84	24	5137	8.57																					
Lingual gyrus	B	17	0	-96	-2	13	4.41															B	18	-4	-80	14	6118	11.66
Superior occipital gyrus	L	19	-22	-66	-4	64	6.58																					
Middle occipital gyrus	R		18	-82	32	5137	7.36																					
								L	V5	-46	-74	0	5137	7.60								R		30	-74	30	6118	9.59
																						L	V5	-46	-74	0	6118	8.83
<i>Subcortical</i>																												
Thalamus																												
Amygdala ^a	B	34	16	-6	-10	21	4.67	R		12	-12	6	187	9.88								R		12	-12	6	21	5.13
Cerebellum	L							L		-30	-70	-28	449	7.74								R	18	16	-72	-16	58	7.21

Note. Regions were assigned using SPM5 Anatomy Toolbox. Final solutions based on $T = 3.70$. This was the FDR ($p < 0.05$) threshold assigned to the contrast for RJA – RJAc. This threshold was selected for final analyses as it was the most conservative of all four contrasts. Coordinates are in Montreal Neurologic Institute Space. All reported p-values are FDR-corrected for whole-brain comparisons. Extent threshold = 10. The threshold for $p < 0.05$ FDR correction would have been 2.87, 3.10, and 3.18 in b, c, and d respectively. No voxels survived FDR correction for Responding over and above Initiating contrast [(RJA – RJAc) – (IJA – IJAc)]. H = Hemisphere, BA = Brodmann Area.

^a Cluster labels based on nearest grey matter to which the significant cluster extended.

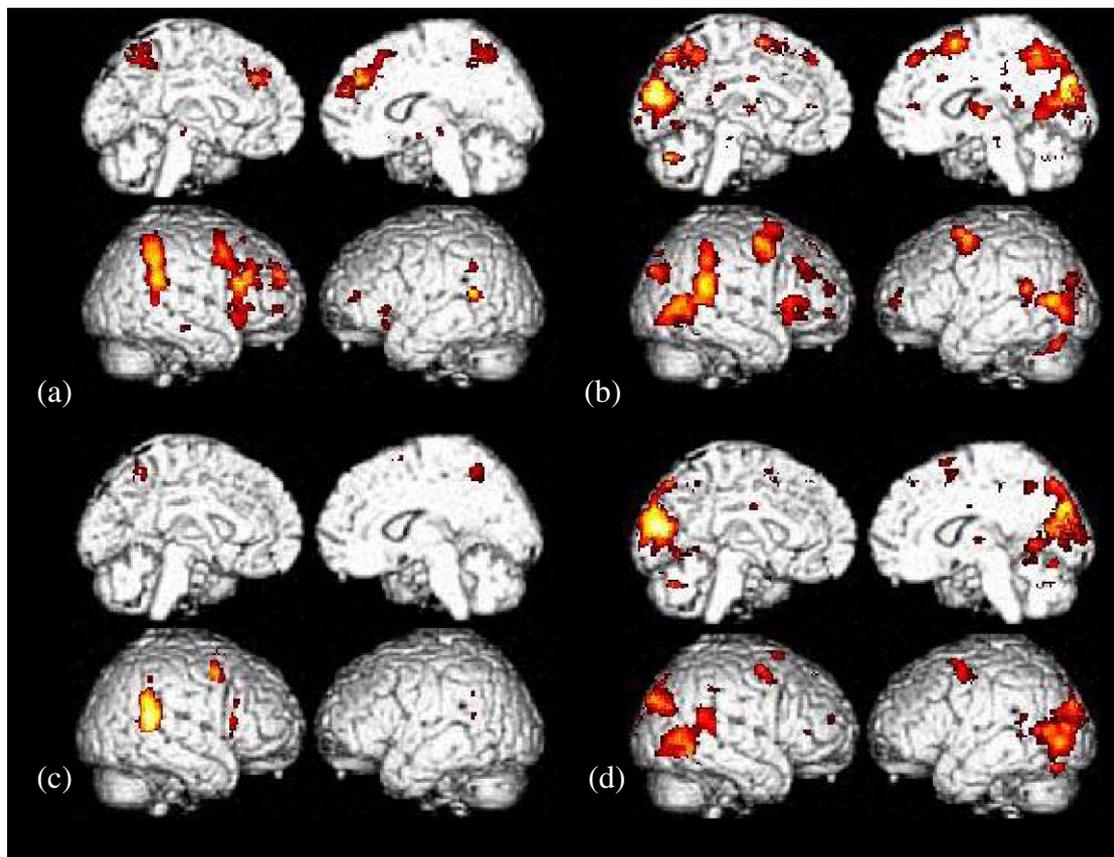


Fig. 4. Thresholded statistical parametric maps showing activity associated with (a) Responding to joint attention (RJA – RJAc) (b) Initiating joint attention (IJA – IJAac) (c) Activation common to Responding and Initiating (d) Initiating over and above Responding [(IJA – IJAac) – (RJA – RJAc)]. $t > 3.70$, equivalent to $p < 0.05$ FDR correction in (a), with extent threshold 10 voxels. The threshold for $p < 0.05$ FDR correction would have been 2.87, 3.10, and 3.18 in b, c, and d respectively. No voxels survived FDR correction for Responding over and above Initiating contrast [(RJA – RJAc) – (IJA – IJAac)].

In the present study, we also exploited our control conditions, to directly contrast activity associated with initiating and responding to joint attention bids, so as to elucidate differences in the networks employed by these different joint attention functions. Whilst there was no additional activation for RJA over and above IJA; (RJA – RJAc) – (IJA – IJAac), IJA resulted in significantly more activation in a bilateral frontotemporoparietal network than RJA; (IJA – IJAac) – (RJA – RJAc). This included a portion of the ‘parallel’ network common to RJA and IJA, including MFG, IFG, TPJ and precuneus, which was active for both tasks but more active during IJA. The involvement of the IFG, TPJ and precuneus may reflect the fact that participants are required to make an additional representation of their partner’s focus of attention on IJA trials, compared to RJA trials. This is consistent with the role of these regions in representing self- and other-oriented visual perspectives (Kincade et al., 2005; Samson et al., 2004; Saxe et al., 2006; Vogeley et al., 2004). Specifically, during IJA trials participants must represent two shifts in the perspective of their partner; (1) from searching to waiting to be guided, indicated by the avatar’s direct gaze and (2) from waiting to be guided to attending to the location indicated by the participant, indicated by the avatar’s congruent gaze. Contrastingly, on RJA trials, there is only one major shift in the social partner’s perspective; from searching for the burglar to guiding the participant to the burglar, indicated by the combination of the avatar’s direct and then averted gaze.

Additionally, we found activation for IJA over and above RJA that was not found in our conjunction analysis. This included bilateral SMA, right superior frontal gyrus, left temporal pole and cerebellum. The activation in and around the SMA may reflect the increased demand on visually guided motor responses and oculomotor control during IJA engagement (Mundy and Newell, 2007; Picard and Strick, 2003).

One unexpected finding was the large area of activation over occipital regions, for IJA over and above RJA. This may have been driven by the extra demands on visual processing presented in the IJA > IJAac contrast, compared to the RJA > RJAc contrast. This is because there was more central visual information in the social initiating (IJA) condition, where the avatar updated his gaze to follow the participant after they initiate joint attention, than in the control initiating (IJAac) condition, where the avatar’s eyes remained closed and the central stimulus did not change once the participant fixates the burglar’s location. One option to match the social and control initiating trials in this respect would have been to have an arrow appear to ‘follow’ the participant after they fixated the burglar by pointing in the congruent location. However, this would have anthropomorphised the cue, potentially disrupting the social manipulation.

In addition to examining the common and separate components of the networks supporting IJA and RJA, our design allowed us to visualise the neural correlates of RJA and IJA separately. This also allowed us to determine whether the increased ecological validity and experimental control achieved in our study resulted in a different representation of the neural correlates of joint attention compared to previous interactive studies. Whilst our data largely corroborated the existing literature, we found some additional sources of activation, including bilateral TPJ and left insula during RJA engagement and ACC during IJA engagement.

The ecologically valid design of our paradigm had the greatest implications for the measurement of RJA, which in the past has been operationalised in tasks involving gaze following, without the need to monitor or disambiguate the attentional focus of the social partner. Unlike previous studies of joint attention, our RJA contrast (RJA – RJAc) resulted in bilateral activation in TPJ, particularly in the right hemisphere (cf. Redcay et al., 2012; Schilbach et al., 2010). The function of the TPJ

remains unclear, with ongoing contention about whether it is directly involved in higher order mentalising computations (Saxe & Kanwisher, 2003), or whether it serves social cognition indirectly through lower level processes (Mitchell, 2008). One argument is that it plays a central role in orienting attention away from internally driven or invalidly cued locations to externally driven, important or task-relevant locations (Corbetta et al., 2008). This could serve in navigating attention during unpredictable social interactions. The need for such processing may come into play during RJA, as an individual adapts from a self-referenced focus of attention to one that is indicated by their social partner (Gallese, 2001). However the specificity of TPJ involvement for social coordination to date remains uncertain (Carter and Huettel, 2013). Here we found TPJ involvement after subtracting activation associated with our non-social task (RJAc), which also elicited comparable goal-directed external shifts of attention. This suggests that TPJ involvement is enhanced when attention is oriented in the context of a social interaction or by social cues; possibly because the attentional cue must be evaluated in conjunction with information about the mental state of the cue provider (Saxe & Kanwisher, 2003). Alternatively, the recruitment of TPJ in this condition could reflect the increased complexity in interpreting the external attentional cue on RJA trials. That is, in our RJA condition, but not in our control RJAc condition, participants had to differentiate between search-related gaze that was not indicative of the burglar's location and averted gaze that followed the avatar's ostensive mutual gaze (Cary, 1978; Senju and Johnson, 2009). Further investigation is needed to elucidate the precise role that TPJ plays in supporting joint attention.

Also inconsistent with previous joint attention studies, RJA resulted in activation of the left insula (cf. Redcay et al., 2012; Schilbach et al., 2010). Although the interpretation of this result is speculative, this region has been previously associated in the perception of emotion (Phan et al., 2002) and self-agency; that is perceiving an outcome as resulting from one's own actions, versus that of another (Farrer and Frith, 2002). On RJA trials there is a shift in agency between the participant and their partner, where first the avatar guides and then the participant uses this information to follow and thereby catch the burglar. This shift in agency does not occur during RJAc trials, because agency is not assigned to the arrow stimulus, and it is the participant who is perceived as the sole agent, responsible for catching the burglar.

The current study also found IJA (IJA – IJAc) to be associated with substrates not yet reported in previous joint attention studies (Redcay et al., 2012; Schilbach et al., 2010), including bilateral ACC. This region has been previously associated with executing goal-directed behaviours and action monitoring (Shackman et al., 2011). It has also been recruited in tasks where participants interact with a virtual agent, and are required to look in a location incongruent with the avatar's gaze (Schilbach et al., 2011). During our IJA task, the participant learns the correct location of the burglar, but before they can guide the avatar, they must observe him search in houses that they know do not conceal the burglar. As such, the ACC could be implicated as a result of integrating the incongruity of a social partner's gaze with a goal-directed action that needs to be executed (i.e., gazing at the burglar location).

In summary, our findings provide support for the PDPM's hypothesis of an integrated neural network supporting RJA and IJA engagement in neurotypical adults. Using a dynamic virtual reality task, we were able to capture RJA- and IJA-related activity from pre-engagement attention monitoring to the execution and evaluation of joint attention. Specifically, we found a right-lateralised and distributed network common to initiating and responding to joint attention bids, which may be supported by the capacity of this network to simultaneously represent self- and other-perspectives of attention. Additional activation for IJA compared to RJA was observed in regions associated with the coordination of these dual perspectives, as well as additional frontotemporoparietal regions that were specific to IJA engagement. This is likely to reflect the increased social complexity of IJA engagement. Our findings provide support for the PDPM by identifying the neural correlates common to

RJA and IJA (TPJ, precuneus, IFG, pSTS, MFG and MTG) and unique to IJA (SMA, right superior frontal gyrus, left temporal pole and cerebellum). These data inform a neural model of joint attention in adults, and may guide future clinical applications of our paradigm to investigate whether the developmental delay of joint attention in autism is associated with a differential organisation of this integrated network.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2014.12.041>.

Acknowledgments

This work was funded by a Macquarie University Department of Cognitive Science Postgraduate Grant to Mr. Caruana. Mr. Caruana is the recipient of an Australian Postgraduate Award.

Dr. Woolgar is the recipient of an Australian Research Council Discovery Early Career Researcher Award (DECRA, DE120100898), and

Dr. Brock is the recipient of an Australian Research Council Research Fellowship Grant (DP098466).

References

- Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7 (4), 268–277. <http://dx.doi.org/10.1038/nrn1884>.
- Bakeman, R., Adamson, L.B., 1984. Article. *Child Dev.* 55 (4), 1278. <http://dx.doi.org/10.1111/1467-8624.ep7302943>.
- Bates, D.M., 2005. Fitting linear mixed models in R. *R News* 5, 27–30.
- Bates, E., Benigni, L., Bretherton, I., Cismaioni, L., Volterra, V., 1979. *The Emergence of Symbols: Cognition and Communication in Infancy*. Academic Press, New York.
- Brass, M., Derrfuss, J., Von Cramon, D.Y., 2005. The inhibition of imitative and overlearned responses: a functional double dissociation. *Neuropsychologia* 43 (1), 89–98.
- Bruinsma, Y., Koegel, R.L., Koegel, L.K., 2004. Joint attention and children with autism: a review of the literature. [Article]. *Ment. Retard. Dev. D R* 10 (3), 169–175. <http://dx.doi.org/10.1002/mrdd.20036>.
- 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).
- Carter, R.M., Huettel, S.A., 2013. A nexus model of the temporal–parietal junction. *Trends Cogn. Sci.* 17 (7), 328–336. <http://dx.doi.org/10.1016/j.tics.2013.05.007>.
- Cary, M.S., 1978. The role of gaze in the initiation of conversation. *Soc. Psychol. Ger.* 41 (3), 269–271. <http://dx.doi.org/10.2307/3033565>.
- Corbetta, M., Patel, G., Shulman, G.L., 2008. The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58 (3), 306–324.
- Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J.A., Estes, A., Liaw, J., 2004. Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Dev. Psychol.* 40 (2), 271–283.
- Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A.C., Zilles, K., Amunts, K., 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage* 36 (3), 511–521. <http://dx.doi.org/10.1016/j.neuroimage.2007.03.060>.
- Farrer, C., Frith, C.D., 2002. Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage* 15 (3), 596–603. <http://dx.doi.org/10.1006/nimg.2001.1009>.
- Gallese, V., 2001. The 'shared manifold' hypothesis: from mirror neurons to empathy. *J. Conscious. Stud.* 8 (6–7), 33–50.
- 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. *Soc. Neurosci.* 1–11. <http://dx.doi.org/10.1080/17470919.2013.832374>.
- Grinband, J., Wager, T.D., Lindquist, M., Ferrera, V.P., Hirsch, J., 2008. Detection of time-varying signals in event-related fMRI designs. *NeuroImage* 43 (3), 509–520. <http://dx.doi.org/10.1016/j.neuroimage.2008.07.065>.
- Halko, M.-L., Hlushchuk, Y., Hari, R., Schürmann, M., 2009. Competing with peers: mentalizing-related brain activity reflects what is at stake. *NeuroImage* 46 (2), 542–548. <http://dx.doi.org/10.1016/j.neuroimage.2009.01.063>.
- Henson, R.N.A., 2006. Efficient experimental design for fMRI. In: Friston, K.J., Ashburner, A.T., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.), *Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images*. Elsevier, London.
- (Producer) Henson, R.N.A., 2012. Matlab function for calculating efficiency of fMRI GLMs (Retrieved from <http://www.mrc-cbu.cam.ac.uk/people/rik.henson/personal/analysis/>).
- Ishikawa, S.S., Raine, A., 2003. Prefrontal deficits and antisocial behaviour: a causal model. In: Lahey, B.B., Moffitt, T.E. (Eds.), *Causes of Conduct Disorder and Juvenile Delinquency*. Guilford, New York, pp. 277–304.
- Kimball, S., Mattis, P., 1995. *GIMP Image Manipulation Program (Version 2.8.2)*. Berkley.
- Kincaid, J.M., Abrams, R.A., Astafiev, S.V., Shulman, G.L., Corbetta, M., 2005. An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. *J. Neurosci.* 25 (18), 4593–4604.
- Lieberman, M.D., Cunningham, W.A., 2009. Type I and Type II error concerns in fMRI research: re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* 4 (4), 423–428.
- Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M., 2000. The autism diagnostic observation schedule—generic: a standard measure

- of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* 30 (3), 205–223.
- Marchetti, I., Koster, E.H.W., 2014. Brain and intersubjectivity: a Hegelian hypothesis on the self-other neurodynamics. [Opinion]. *Front. Hum. Neurosci.* 8. <http://dx.doi.org/10.3389/fnhum.2014.00011>.
- Mitchell, J.P., 2008. Activity in right temporo–parietal junction is not selective for theory–of–mind. *Cereb. Cortex* 18 (2), 262–271.
- Morris, J.P., Pelphrey, K.A., McCarthy, G., 2008. Perceived causality influences brain activity evoked by biological motion. *Soc. Neurosci. UK* 3 (1), 16–25. <http://dx.doi.org/10.1080/17470910701476686>.
- Mundy, P., Jarrold, W., 2010. Infant joint attention, neural networks and social cognition. *Neural Netw.* 23 (8), 985–997. <http://dx.doi.org/10.1016/j.neunet.2010.08.009>.
- Mundy, P., Newell, L., 2007. Attention, joint attention and social cognition. *Curr. Dir. Psychol. Sci.* 16, 269–274.
- Mundy, P., Sigman, M., Kasari, C., 1990. A longitudinal study of joint attention and language development in autistic children. *J. Autism Dev. Disord.* 20 (1), 115–128. <http://dx.doi.org/10.1007/bf02206861>.
- Mundy, P., Sullivan, L., Mastergeorge, A.M., 2009. A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Res.* 2 (1), 2–21. <http://dx.doi.org/10.1002/aur.61>.
- Ohlendorf, S., Kimmig, H., Glauche, V., Haller, S., 2007. Gaze pursuit, 'attention pursuit' and their effects on cortical activations. *Eur. J. Neurosci.* 26 (7), 2096–2108. <http://dx.doi.org/10.1111/j.1460-9568.2007.05824.x>.
- Pelphrey, K.A., Morris, J.P., McCarthy, G., 2004. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J. Cogn. Neurosci.* 16 (10), 1706–1716. <http://dx.doi.org/10.1162/0898929042947900>.
- Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I., 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage* 16 (2), 331–348. <http://dx.doi.org/10.1006/nimg.2002.1087>.
- Picard, N., Strick, P.L., 2003. Activation of the Supplementary Motor Area (SMA) during performance of visually guided movements. *Cereb. Cortex* 13 (9), 977–986. <http://dx.doi.org/10.1093/cercor/13.9.977>.
- Posner, M.I., Rothbart, M., 2007. *Educating the human brain*. American Psychological Association, Washington, DC.
- Ramsey, R., Hansen, P., Apperly, I., Samson, D., 2013. Seeing it my way or your way: frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *J. Cogn. Neurosci.* 25 (5), 670–684.
- Redcay, E., Dodell-Feder, D., Mavros, P.L., Klein, A.M., Pearrow, M.J., Triantafyllou, C., Saxe, R., 2012. Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Hum. Brain Mapp.* 34, 2511–2523.
- Redcay, E., Dodell-Feder, D., Pearrow, M.J., Mavros, P.L., Kleiner, M., Gabrieli, J.D.E., Saxe, R., 2010. Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *NeuroImage* 50 (4), 1639–1647. <http://dx.doi.org/10.1016/j.neuroimage.2010.01.052>.
- Sabbagh, M.A., 2004. Understanding orbitofrontal contributions to theory-of-mind reasoning: implications for autism. *Brain Cogn.* 55 (1), 209–219. <http://dx.doi.org/10.1016/j.bandc.2003.04.002>.
- Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporoparietal junction is necessary for representing someone else's belief. [10.1038/nn1223]. *Nat. Neurosci.* 7, 5, pp. 499–500 doi: http://www.nature.com/neuro/journal/v7/n5/supplinfo/nn1223_S1.html
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: The role of the temporo–parietal junction in “theory of mind”. *NeuroImage* 19 (4), 1835–1842. [http://dx.doi.org/10.1016/S1053-8119\(03\)00230-1](http://dx.doi.org/10.1016/S1053-8119(03)00230-1).
- Saxe, R., Schulz, L.E., Jiang, Y.V., 2006. Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Soc. Neurosci. UK* 1 (3–4), 284–298. <http://dx.doi.org/10.1080/17470910601000446>.
- Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42 (11), 1435–1446. <http://dx.doi.org/10.1016/j.neuropsychologia.2004.04.015>.
- Scaife, M., Bruner, J.S., 1975. The capacity for joint visual attention in the infant. [10.1038/253265a0]. *Nature* 253 (5489), 265–266.
- Schilbach, L., Eickhoff, S.B., Cieslik, E., Shah, N.J., Fink, G.R., Vogeley, K., 2011. Eyes on me: an fMRI study of the effects of social gaze on action control. *Soc. Cogn. Affect. Neurosci.* 6 (4), 393–403. <http://dx.doi.org/10.1093/scan/nsq067>.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Vogeley, K., 2013. Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–462.
- Schilbach, L., Wilms, M., Eickhoff, S.B., Romanzetti, S., Tepest, R., Bente, G., Vogeley, K., 2010. Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22 (12), 2702–2715.
- Senju, A., Johnson, M.H., 2009. The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13 (3), 127–134. <http://dx.doi.org/10.1016/j.tics.2008.11.009>.
- Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J., 2011. The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12 (3), 154–167. <http://dx.doi.org/10.1038/nrn2994>.
- Sigman, M., Ruskin, E., 1999. Continuity and change in the social competence of children with autism, Down syndrome and developmental delays. *Monogr. Soc. Res. Child Dev.* 64 (1), 1–114.
- Simpson, J.R., Snyder, A.Z., Gusnard, D.A., Raichle, M.E., 2001. Emotion-induced changes in human medial prefrontal cortex: I. During cognitive task performance. *Proc. Natl. Acad. Sci. U. S. A.* 98, 683–687.
- Inversions, Singular, 2008. *FaceGen Modeller (Version 3.3)* [Computer Software]. Singular Inversions, Toronto, ON.
- Research, S.R., 2004. *Experiment Builder (Version 1.10.165)* (Ontario).
- Tzelepi, A., Laskaris, N., Amditis, A., Kapoula, Z., 2010. Cortical activity preceding vertical saccades: a MEG study. *Brain Res.* 1321, 105–116. <http://dx.doi.org/10.1016/j.brainres.2010.01.002>.
- Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., Fink, G.R., 2004. Neural correlates of first-person perspective as one constituent of human self-consciousness. *J. Cogn. Neurosci.* 16 (5), 817–827. <http://dx.doi.org/10.1162/089892904970799>.
- 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. *Soc. Cogn. Affect. Neurosci.* 5 (1), 98–107. <http://dx.doi.org/10.1093/scan/nsq024>.
- Woolgar, A., Golland, P., Bode, S., 2013. Coping with confounds in multivoxel pattern analysis: what should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. *NeuroImage* <http://dx.doi.org/10.1016/j.neuroimage.2014.04.059>.